Laser and Optics

Masroor Ikram

and

Muhammad Yousaf Hamza

Pakistan Institute of Engineering and Applied Sciences

P.O. Nilore, Islamabad.

Department of Physics
Allama Iqbal Open University
Islamabad
1999

Reviewed by: Dr. Shahid Qamar
(**Pakistan Institute of Engineering and Applied Sciences)**

Edited by: Asloob Ahmed Mudassar
(**Pakistan Institute of Engineering and Applied Sciences)**

Preface

Lasers are widely used in most of the industrial processing, engineering, metrology, scientific research, communications, holography, medicine, and for defense purposes. Due to the importance of the subject, most of the western universities are offering courses in this discipline. In our country also, the importance of the subject has been felt, and some universities and establishments are offering courses and doctoral research programs. Allama Iqbal Open University (AIOU), although new in the field of natural sciences has also included courses of Lasers & Quantum Optics in their M.Sc. (Phys.) and M.Phill. (Phys.) programs. This book has been written for the students of AIOU taking M.Sc. courses in physics in their 2nd semester. It is equally good for the students of M.Sc. Physics, Applied Physics, Electronics, and B.E. in Electronics from other conventional universities.

The book contains nine units and is equivalent to half credit course of AIOU. An effort is made to include the material according to the approved contents of AIOU for the course entitled *Laser and Optics* for M.Sc. Physics. The first three units are devoted for *Optics*, which is very important to understand the lasers. Four units (Unit-4 to Unit-7) cover the basic principles of lasers, Unit-8 describes some of the laser systems, while Unit-9 gives the flavor of a few applications of lasers. Keeping in mind the distant learning approach of the AIOU, the detailed explanations have been given for most of the involved concepts. However, the book has been written in a very short span of time, therefore, there may be mistakes and errors. All the suggestions concerning the improvement of the text will be welcomed.

I highly appreciate the efforts of Muhammad Yousaf Hamza for writing the Laser Systems and Medical Applications of Lasers. Mr. Hamza also helped me in making the problems and positive criticism after reading the text very carefully.

In the end, I am grateful to the Pakistan Institute of Engineering and Applied Sciences (PIEAS) for allowing me to collaborate with AIOU, and to my family for their untiring support during the write-up. I also acknowledge my colleagues Dr. M. R. Najam and Dr. S. M. Farooqi for their useful suggestions.

Masroor Ikram

April 20, 1999.

Table of Contents

Unit-1

Light

Objective

Most of the information we receive from our surroundings passes through our eyes. This information is carried by light. Nature of light and its application was the subject of great interest among the scientists. During the magnificent period of Islam, Muslim scientists and philosophers contributed significantly in this field but unfortunately in most of the books their work is neither referred nor acknowledged. It is my intentional efforts to highlight the contribution of Muslim scientists in this field. This unit describes briefly the history of light and its nature. Furthermore electromagnetic wave, and superposition of these waves are discussed in some detailed to make the background for understanding the phenomena of polarization, interference, and diffraction in the next unit.

1.1 History

In the fifth century BC, the Greek philosophers determined the link between the eye and the object seen. The linked could be thought to be *'something'* which was emitted by the eye and traveled to the object, or *'something'*, which traveled towards the eye from the object, or the co-existence of both *'something'* traveling in the opposite directions.

The first idea, that of emission from the eye traveling to the object, was accepted by Pythagorean. The Atomists of the school of Democritus were in favor of the theory of an emission from the object traveling to the eye. Empedocles was among the first to support the idea of a combination of two fluxes. None of the theories was ever fully accepted but followers of one or the other of the theories were there for many centuries.

After the glory of Islam, the center of learning was shifted from Europe to the Middle East. Where in the ninth century at Baghdad very important work on visual sensation was carried out. A famous scientist, Abu Yosuf Yaqub Ibn Is-haq (known as Alkindi in the west) worked over a wide field and many branches of physics. In the book *De Aspectibus*, Alkindi dealt very explicitly with the problems of optics. He asserted that vision had to take place by means of rays capable of having a physical action upon the eye. Alkindi transformed and perfected the idea of a ray. He noted that the formation of shadows produced by bodies when illuminated by *lumina* which entered from a window, led without any doubt to the conclusion that the rays emanating from luminous bodies traveled along rectilinear paths.

In the tenth century two scientists working on the physiology of the eye, were Ali Ibn Isa (known as Jesus Hali) and Abu Bekr Muhammed Ibn Zakariya Al Razi (known as Rasis). Their work was latter-on translated in Latin, during the middles ages their works assumed the character of classic texts. After Alkindi, the prominent Muslim philosopher was Abu Ali Mohammed Ibn Al Hasan Ibn Al Haytham (known as Alhazan), who made the long lasting effect in all the sciences especially in the field of medicine and optics. The translation of one of his book, *Opticae Thesaurus libri septem, per Episcopios*, earned a great respect in the west and was taught for many centuries. Alhazan accepted the theory of Alkindi about light and elaborated it with fine details. Alhazan's results were supported by experiments; he proved that by looking at the sun or its reflection through a mirror produced pain in the eye. This clearly shows that if there were something, which travels from the eye to the object, then there would be no reason to feel pain. Similarly he proved the persistence of the image on the ratina. In his theory of light, Alhazan attributed to this light the property of reflection when it met a mirror, and of refraction when it traveled through transparent surfaces. He gave full reasoning of the reflection.

Much more famous than Alhazan was his contemporary Abu Ali Hosain Ibn Abudallah Ibn Sina (known as Avicenna). He became one of the best known of the mediaeval philosophers and his influence spread also to the west. He studied with great care and in detail the problem of the functioning of the sense of vision. He explicitly said that in the process of vision there was a very important intervention of the mind of the observer that played a very important part in the total process. Avicenna was against the idea of Alkindi and Alhazan that the rays capable of acting on the eye were of a material nature. He reintroduced the theory of emission of simulacra from bodies, but strips these simulacra of all material form. These ideas were well received in the west and received much more support than the ideas of the materialists, i.e., the ideas of Alkindi and Alhazan.

Before the death of an all time great Muslim philosopher, Avicenna, another influential Muslim scholar was born in Spain, his name was Abul Walid Mohammed Ibn Ahmad Ibn Muhammad Ibn Rushd (known as Averroes). Regarding the theory of light, Averroes openly challenged the Avicenna's ideas concerning the mechanism of vision but he kept the argument purely in the philosophical field. His fundamental philosophical ideas were felt so deeply in the official quarters that his writing were burnt by royal decree.

From the sixteenth century the center of learning were shifted to Europe. Newton and Huygens were among the early prominent workers in this field. Newton's corpuscular theory got dominance for a century and was replaced by wave theory proposed by Huygens and supported by Fresnel and others. Latter Maxwell polished it. The twentieth century again modified the earlier theories of light.

1.2 Nature of Light

Light is an elephant.

An old story tells of three blind men who were asked to describe an elephant. One blind man touched the elephant's tail and said the elephant was long and thin like a rope. The second blind man touched the elephant's leg and described the elephant as round and hard like a tree trunk. The third man felt an ear and said that the elephant was thin and flat, like a huge leaf. Each man's description was correct, but didn't give the complete picture.

Scientists who study the nature of light are like the blind men in the story. They try to describe light, but their descriptions depend very much on which aspects of light they study. Each description of light is merely an approximation to the reality that is light. During the last four centuries, light is sometimes considered as particle and sometimes as wave, now scientists agreed on its dual nature. In the following discussion a brief chronological evaluation of the theories of light is described.

Corpuscular theory of Newton (1643-1727) described that light rays consisted of streams of tiny particles, which are emitted by the light sources and propagate through space in straight lines.

The theory satisfied the phenomena of reflection and refraction and received a general acceptance in the seventeenth century. But it failed to explain the phenomena like diffraction and interference.

Wave theory of Huygens (1629-1695) described the light as waves like water or sound waves. The theory not only satisfied the diffraction, interference and polarization but also explained the reflection and refraction of light. According to the known principles of wave propagation at that time, the theory assumed some medium (ether) in space to propagate light waves from the sun.

In the nineteenth century, Maxwell (1831-1879) combined the laws of electricity and magnetism and proposed the electromagnetic nature of light, which do not require a medium to travel. The theory satisfied all the problems faced till the end of nineteenth century and received such a universal acceptance that scientists believed that problem regarding the nature of light had been solved and the corpuscular theory was comprehensively rejected.

In the beginning of the twentieth century, experimental phenomenon of photoelectric effect and the blackbody radiation introduced the quantum theory of light and gave rebirth to the particle nature. According to this theory, light is an electromagnetic radiation and consists of photons of energy E, thus light exhibits in both ways. On one hand, wave theory explains the interference and diffraction phenomena, while on the other hand, photoelectric effect, Compton effect etc., are best explained by considering the light as photons having energy and momentum.

In the dual nature, light consists of photons (of energy E) and each photon has a wavelength $\lambda = \frac{hc}{\mathsf{E}}$, where h is Plank's constant and c is velocity of light in vacuum. Furthermore, this wave nature satisfies the Maxwell's equations, thus is an electromagnetic wave. Since some important phenomena e.g. diffraction, interference and polarization require the wave nature and are frequently used in optics and lasers, therefore, we will discuss the electromagnetic wave and the superposition of two or more waves in the following discussions.

1.3 Electromagnetic waves

In studying electricity and magnetism, one knows the fact that a time-varying electric field generates a magnetic field that is everywhere perpendicular to the direction in which **E** changes.

In the same way, a time-varying **B**-field generates an **E**-field that is every where perpendicular to the direction in which B changes. This behavior shows the transverse nature of the **E**- and **B**-fields in an electromagnetic disturbance.

Consider a charge that is forced to accelerate from rest. When the charge is motionless, it has associated with it a radial **E**-field extending to all direction to infinity. At the instant the charge begins to move, the **E**-field is altered in the vicinity of the charge. This alteration propagates out into space at some finite speed. The time-varying electric field induces a magnetic field (**B**-field), which is also time dependent. The time varying **B**-field generates an **E**-field, and the process continues, with **E** and **B** coupled in the form of a pulse. As one field changes, it generates a new field that extend a bit further, and the pulse moves out from one point to the next through space.

The **E**- and **B**- fields can more appropriately be considered as two aspects of a single physical phenomenon, the electromagnetic field, whose source is a moving charge. The disturbance, once it has been generated in the electromagnetic field, is an independent wave beyond its source. Bound together as a single entity, the time varying electric and magnetic fields regenerate each other in an endless cycle.

Maxwell showed that both the electric and magnetic fields satisfy the same partial differential equations (see Appendix. 1A).

$$\nabla^2\mathbf{E} = \varepsilon_o\mu_o.\frac{\partial^2\mathbf{E}}{\partial t^2} \tag{1.1a}$$

and

$$\nabla^2\mathbf{B} = \varepsilon_o\mu_o.\frac{\partial^2\mathbf{B}}{\partial t^2} \tag{1.1b}$$

where ε_o and μ_o are the relative permittivity and permeability of free space, respectively. This is called the wave equation and the speed of propagation of these waves in free space is given by

$$c = \sqrt{\frac{1}{\varepsilon_o\mu_o}}\,. \tag{1.2}$$

The implication of equation (1.1a & 1.1b) is that changes in the fields propagate through space with a speed c, the speed of light. The frequency of oscillation of the fields, ν, and their wavelength in vacuum, λ_o, are related by

$$c = \nu\lambda_o. \tag{1.3}$$

In any other medium the speed of propagation is given by

$$v = c/n = \nu\lambda \tag{1.4}$$

where n is the refractive index of the medium and λ is the wavelength in the medium. Here n is given by

$$n = \sqrt{\varepsilon_r\mu_r} \tag{1.5}$$

where ε_r and μ_r are the relative permittivity and permeability of the medium, respectively.

Example 1.1: *The thickness of a human hair is about 40 μm. Compare its dimension to the wavelength of green light.*

Solution*: The wavelength of green light is around 500 nm, therefore, the thickness a human hair is about 80 times the wavelength of the green light.*

Example 1.2*:Calculate the speed of light from the values of μ_o and ε_o. Also find the speed of light in water and in glass.*

Solution*: Values of the constants are:*

$\mu_o = 4\pi x10^{-7}$ J s^2 C^{-2} m^{-1} *and* $\varepsilon_o = 8.854188 \times 10^{-12}$ J^{-1}C^2 m^{-1}

Therefore, $c = \sqrt{\frac{1}{\varepsilon_o\mu_o}}$ = 2.9979 m s^{-1}.

The refractive index of water, $n_w = 1.333$

Therefore, velocity of light in water is c/n_w is 2.249 m/sec.

The refractive index of ordinary glass, $n_g = 1.5$

Thus velocity of light in glass is 1.9986 m/sec.

The electric and magnetic field vibrate perpendicularly to one another and perpendicular to the direction of propagation as illustrated in Figure 1.1; that is, light waves are transverse waves. In describing optical phenomena it is common to omit the magnetic field vector. This simplifies diagrams and mathematical descriptions but we should always remember that there is a magnetic field component, which also behave in a similar way to the electric field component.

The simplest waves are sinusoidal waves, which can be expressed mathematically by the

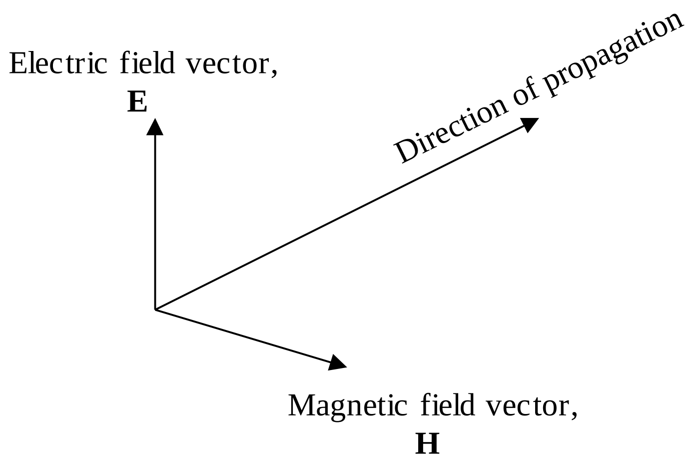


Figure 1.1 An electromagnetic wave; the electric field vector (**E**) and the magnetic field vector (**H**) vibrate in orthogonal planes and perpendicularly to the direction of propagation.

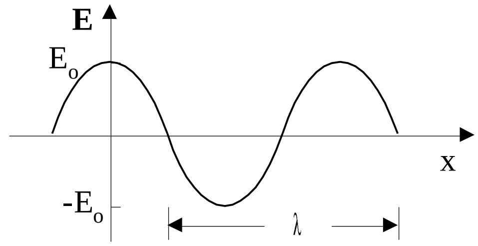
equation:

$$E(x,t) = E_o . \cos(\omega t - kx + \phi) \qquad (1.6)$$

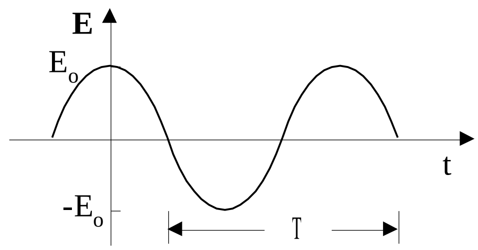
where E(x,t) is the value of the electric field at the point x at time t, E_o the amplitude of the wave, ω the angular frequency ($\omega = 2\pi\nu$), k the wavenumber ($k = 2\pi/\lambda$) and ϕ is the phase constant. The term ($\omega t - kx + \phi$) is the phase of the wave. Equation (1.6), which describes a plane and monochromatic wave propagating in the positive x-direction, is a solution of the wave equation (1.1a).

We can represent equation (1.1a) diagrammatically by plotting E as a function of either x or t as shown in Figure 1.2, where we have taken $E = E_o$ at x and t equal to zero so that $\phi = 0$. Figure (1.2a) shows the variation of the electric field with distance at a given instant of time. If we take time t equal to zero, then the spatial variation of the electric field is given by

$$E = E_o.\cos(kx) \tag{1.7a}$$



(a)



(b)

Figure 1.2 The electric field (E) of an electromagnetic wave plotted as function of (a) the spatial coordinate x and (b) the time t.

Similarly Figure (1.2b) shows the variation of electric field as a function of time at some specific location in space. If we take x equal to zero then the temporal variation of electric field is given by

$$E = E_o.\cos(\omega t) \tag{1.7b}$$

Equation (1.6) can be written in a variety of equivalent forms using relationship between ν, ω, λ, k and c. The time for one cycle is period T ($T = 1/\nu$) as shown in Figure (1.2b). If the value of E at x=0 and t=0 is not E_o then we must include arbitrary phase constant ϕ. Equation (1.6) can also be expressed using sine rather than a cosine function or alternatively using complex exponential.

Equation (1.6) represent plane waves moving along the x-axis; we can generalize our mathematical description to include plane waves moving in arbitrary directions. Such a wave can be characterized by a wavevector **k** where $|\mathbf{k}| = 2\pi/\lambda$ and equation (1.6) becomes

$$E(x,y,z,t) = E_o.\cos(\omega t - \mathbf{k.r} + \phi) \qquad (1.8)$$

where r is a vector from the origin to the point (x,y,z). Thus, for example, if we have a plane wave propagating in a direction θ to the x-axis with its wavefronts normal to the (x,y) plane as shown in Figure 1.3, we can write

$$\mathbf{k} = \mathbf{i}k_x + \mathbf{j}k_y \qquad (1.9)$$

and

$$\mathbf{r} = \mathbf{i}x + \mathbf{j}y \qquad (1.10)$$

where **i** and **j** are unit vectors in the x and y directions, respectively. Combining equations (1.9) and (1.10) we have

$$\mathbf{k} \,.\, \mathbf{r} = x\,k_x + y\,k_y = x\,k\cos\theta + y\,k\sin\theta. \qquad (1.11)$$

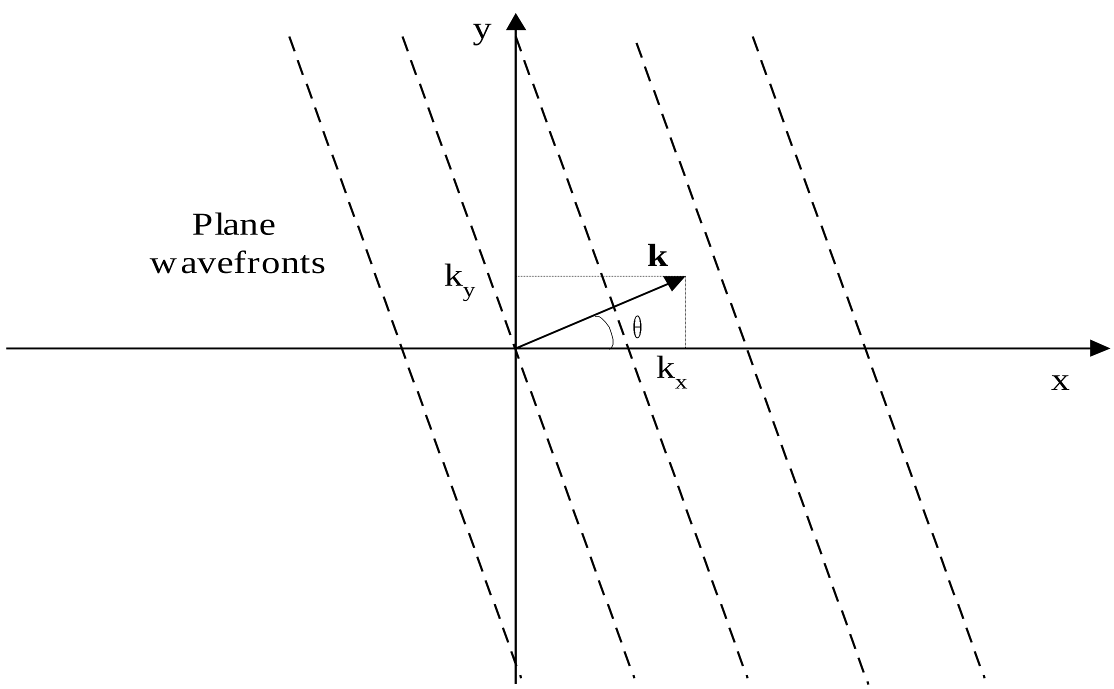


Figure 1.3 Plane wave with its propagation vector **k** in the (x,y) plane. The components of the propagation vector are $k_x = k.\cos\theta$ and $k_y = k.\sin\theta$.

Hence we can rewrite equation (1.8) as

$$E(x,y,t) = E_o.\cos(\omega t - x\, k \cos\theta - y\, k \sin\theta + \phi) \qquad (1.12)$$

An equally important concept is that of spherical waves which we can imagine, are generated by a point source of light. If such a source is located in an isotropic medium it will radiate uniformly in all directions; the wavefronts are thus a series of concentric spherical shells. We can describe this situation by

$$E = \frac{\mathscr{A}}{r}\cos(\omega t - \mathbf{k.r}) \qquad (1.13)$$

where the constant $\mathscr{A}$ is known as source strength. The factor 1/r in the amplitude term accounts for the decrease in amplitude of the wave as it propagates further and further from the source. As the irradiance is proportional to the square of the amplitude, there is an inverse square law decrease in irradiance. If the medium in which the source is located is an isotropic then the wave surfaces are no longer spheres; their shapes depend on the speed of propagation in different directions.

1.4 The Superposition of Waves

When two sets of waves are made to cross each other, e.g., the waves created by dropping two stones simultaneously in a quite pool, interesting and complicated effects are observed. In the region of crossing there are places where the disturbance is practically zero and others where it is greater than that given by either wave alone. A very simple law can be used to explain these effects, which states that the resultant displacement of any point is merely the sum of the displacements due to each wave separately. This is known as *principle of superposition* and was first clearly stated by Young in 1802. The truth of this principle is at once evident when we observe that after the waves have passed out of the region of crossing, they appear to have been entirely uninfluenced by the other set of waves. Amplitude, frequency, and all other characteristics are just as if they had crossed an undisturbed space. This could hold only provided the principle of superposition was true. Two different observers can see different objects through the same aperture with perfect clearness, whereas the light reaching the two observers crossed in going through the aperture. The principle is therefore applicable with great precision to light, and we can use it in investigating the disturbance in regions where two or more light waves are superimposed.

1.4.1 Addition of waves of the same frequency

1.4.1.1 The Algebraic Method

Consider the effect of superimposing two sine waves of the same frequency, the problem resolve itself into finding the resultant motion when a particle executes two simple harmonic motions of the same time. The displacements due to the two waves are taken to be along the same line, which we shall call the x-direction. The solution of the differential wave equation given by equation (1.1a) will have the form

$$E(x,t) = E_o.\ \sin[\omega t - (kx + \phi)]\,, \qquad (1.14)$$

where E_o is the amplitude of the harmonic disturbance propagating along the positive x-direction. Let

$$\alpha\ (x, \phi) = -\ (kx + \phi) \qquad (1.15)$$

so that

$$E(x,t) = E_o.\ \sin\ [\omega t + \alpha\ (x, \phi)], \qquad (1.16a)$$

Suppose that we have two such waves of amplitudes E_{o1} and E_{o2}, then we have

$$E_1(x,t) = E_{o1}.\sin\ [\omega t + \alpha_1], \qquad (1.16b)$$

and

$$E_2(x,t) = E_{o2}.\sin\ [\omega t + \alpha_2], \qquad (1.16c)$$

The two waves have same frequency, amplitudes and they are overlapping in space. The resultant disturbance is the linear superposition of these waves. Therefore,

$$E = E_1 + E_2 \qquad (1.17)$$

or on expanding equations (16b) and (16c),

$$E = E_{o1}.(\sin\omega t\ \cos\alpha_1 + \cos\omega t\ \sin\alpha_1) + E_{o2}.(\sin\omega t\ \cos\alpha_2 + \cos\omega t\ \sin\alpha_2). \qquad (1.18)$$

By separating the time dependant terms, this equation becomes

$$E = (E_{o1} \cos\alpha_1 + E_{o2} \cos\alpha_2) \sin\omega t + (E_{o1} \sin\alpha_1 + E_{o2} \sin\alpha_2) \cos\omega t. \tag{1.19}$$

Since the terms in the brackets are independent of time, so let

$$E_o \cos\alpha = E_{o1} \cos\alpha_1 + E_{o2} \cos\alpha_2 \tag{1.20}$$

and

$$E_o \sin\alpha = E_{o1} \sin\alpha_1 + E_{o2} \sin\alpha_2. \tag{1.21}$$

This is not the obvious solution, but will be legitimate as long as we can solve for E_o and α. By squaring and adding of equation (1.20) and (1.21) we get

$$E_o^2 = E_{o1}^2 + E_{o2}^2 + 2.E_{o1}E_{o2} \cos(\alpha_2 - \alpha_1) \tag{1.22}$$

and dividing equation (1.21) by (1.20) we get

$$\tan\alpha = \frac{E_{o1} \sin\alpha_1 + E_{o2} \sin\alpha_2}{E_{o1} \cos\alpha_1 + E_{o2} \cos\alpha_2}. \tag{1.23}$$

If these two expressions are satisfied for E_o and α, then equations (1.20) and (1.21) are valid. The total disturbance then becomes

$$E = E_o \cos\alpha \sin\omega t + E_o \sin\alpha \cos\omega t \tag{1.24a}$$

or

$$E = E_o \sin(\omega t + \alpha). \tag{1.24b}$$

Thus a single disturbance results from the superposition of the sinusoidal waves E_1 and E_2. *The composite wave (equation 1.24) is harmonic and of the same frequency as the constituents, although its amplitudes and phase are different.* The flux density of a light wave is proportional to the square of its amplitude. From equation (1.22) the resultant flux density is not simply the sum of the component flux densities, there is an additional contribution $2E_{o1}E_{o2} \cos(\alpha_2 - \alpha_1)$, known as interference term. The important factor is the difference in phase between the two interfering waves E_1 and E_2, i.e., $\delta = (\alpha_2 - \alpha_1)$. When d = 0, $\pm 2\pi$, $\pm 4\pi$, … the resultant amplitude is maximum, whereas $\delta = \pm\pi$, $\pm 3\pi$, … yields a minimum. In the former case, the waves are said to be in phase, i.e., crust overlaps crust. In the latter case, the waves are 180^o out of phase and trough overlaps crust, as shown in Figure 1.4. The phase difference may arise from a difference in path length traversed by the two waves, as well as a difference in the initial phase angle, that is,

$$\delta = (k.x_1 + \phi_1) - (k.x_2 + \phi_2) \qquad (1.25a)$$

or

$$\delta = 2\pi/\lambda.(x_1 - x_2) + (\phi_1 - \phi_2). \qquad (1.25b)$$

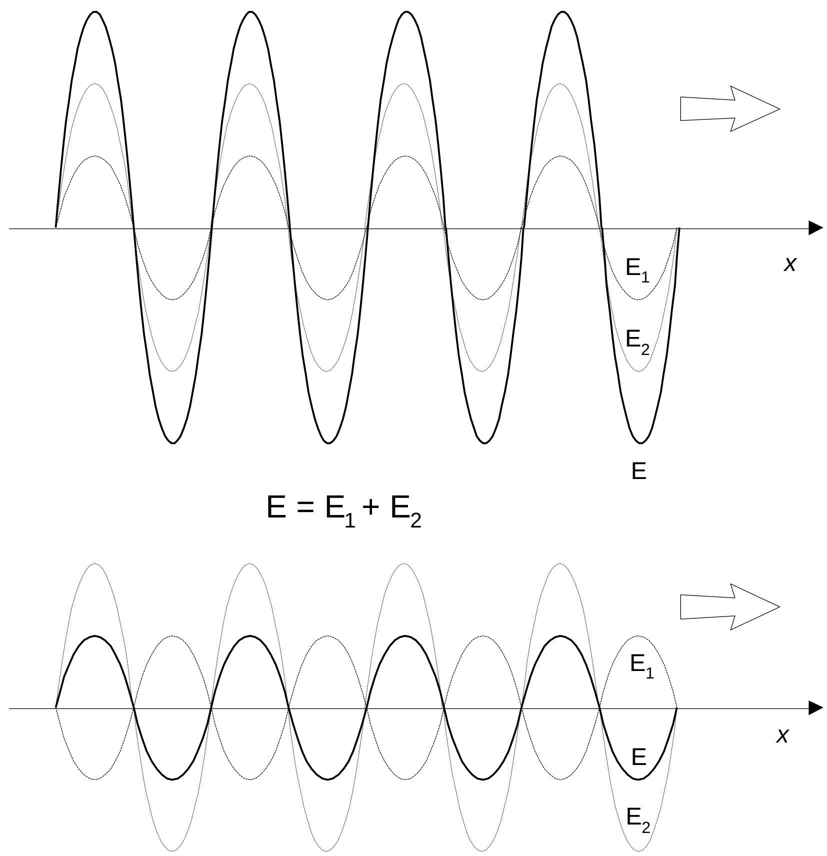


Figure 1.4 The superposition of two harmonic waves in and out of pha

Here x_1 and x_2 are the distances from the sources of the two waves to the point of observation, and λ is the wavelength in the medium. If the waves are initially in phase, then $\phi_1 = \phi_2$, and

$$\delta = 2\pi/\lambda.(x_1 - x_2). \qquad (1.26)$$

This would also apply to the case in which two disturbances from the same source traveled different paths before arriving at the point of observation. Since $n = c/v = \lambda_o/\lambda$,

$$\delta = \frac{2\pi}{\lambda_o} n.(x_1 - x_2) \tag{1.27}$$

The quantity $n(x_1 - x_2)$ is called *optical path difference* and is represented by the abbreviation OPD or by the symbol Λ. It is the difference of the two optical path lengths. Thus equation (1.27) can be written as

$$\delta = k_o.\Lambda, \tag{1.28}$$

where k_o is the wave number in vacuum and is equal to $2\pi/\lambda_o$.

Waves for which $(\phi_1 - \phi_2)$ is constant, regardless of its value, are said to be coherent. In the subsequent discussion we will assume that waves are coherent.

1.4.1.2 Sum of 'N' Harmonic Waves

When we have 'N' number of coherent harmonic waves having a given frequency and travelling in the same direction then their superposition will lead to a harmonic wave of the same frequency. So far we have chosen to represent the two waves above in terms of sin function, but the same result will be obtained if we used cosine function. In general the sum of N such waves

$$E = \sum_{i=1}^{N} E_{oi}.\cos(\alpha_i \pm \omega t)\ , \tag{1.29}$$

is given by

$$E = E_o \cos(\alpha + \omega t), \tag{1.30}$$

where

$$E_o^2 = \sum_{i=1}^{N} E_{oi}^2 + 2.\sum_{j>i}^{N}\sum_{i=1}^{N} E_{oi}E_{oj}.\cos(\alpha_i - \alpha_j) \tag{1.31}$$

and

$$\tan\alpha = \frac{\sum_{i=1}^{N} E_{oi}.\sin\alpha_i}{\sum_{i=1}^{N} E_{oi}.\cos\alpha_i}. \tag{1.32}$$

Consider N number of atomic emitters comprising an ordinary light source (e.g. a bulb, candle flame, or discharge lamp). Each atom is effectively an independent source of photon wave trains that varies in its phase rapidly and randomly. In any event, the phase of the light from one atom, $\alpha_i(t)$, will remain constant with respect to the phase from another atom $\alpha_j(t)$ for a very short time (i.e. atomic transition time, which is typically up to 10 nano-seconds) before it changes randomly. Thus atoms are coherent for up to about 10 ns. The flux density is proportional to the time average of E_o^2, generally taken over a comparatively long interval of time, therefore, the second summation in equation (1.31) will contribute terms proportional to $\langle\cos[\alpha_i(t) - \alpha_j(t)]\rangle$, each of which will average out to zero because of the random nature of the phase change. Only the first summation remains in the time average, and its terms are constants. If the atoms are each emitting wave-trains of the same amplitude E_{o1}, then

$$E_o^2 = N.E_{o1}^2. \tag{1.33}$$

The resultant flux density arising from N sources having random, rapidly varying phases is given by N times the flux density of any one source. In other words it is determined by the sum of individual flux densities. A flashlight bulb whose atoms are all emitting randomly, the superposition of these essentially “incoherent” wave-trains is also rapidly and randomly varying in phase. Therefore two or more bulbs will emit light that is essentially incoherent (for duration longer than atomic transition time ~ 10ns) give light whose combined irradiance will simply equal the sum of the irradiance contributed by each bulb. This is also true for candle flames, flashtubes, and all thermal (not laser) sources. We can not expect to see interference when the lightwaves from two reading lamps overlap.

At the other extreme, if the sources are coherent and in phase at the point of observation (i.e., $\alpha_i = \alpha_j$), equation (1.31) will become

$$E_o^2 = \sum_{i=1}^{N} E_{oi}^2 + 2.\sum_{j>i}^{N}\sum_{i=1}^{N} E_{oi}E_{oj} \tag{1.34a}$$

or, equivalently,

$$E_o^2 = \left(\sum_{i=1}^{N} E_{oi}\right)^2 . \tag{1.34b}$$

Again suppose that amplitude of each wave is E_{o1}, we get

$$E_o^2 = (N.E_{01})^2 = N^2.E_{o1}^2. \tag{1.35}$$

In this case of in-phase coherent sources, we have situation in which the amplitudes are added first and then squared to determine the resulting flux density. The superposition of coherent waves generally has the effect of changing the spatial distribution of the energy but the total amount remains constant. If there are regions where the flux density is greater than the sum of the individual flux densities, then there will be regions where it is less than that sum.

1.4.1.3 The Complex Method

It is sometimes convenient to make use of the complex representation of the trigonometric functions when dealing with the superposition of harmonic disturbances. The wave

$$E_1 = E_{o1}\cos(kx + \omega t + \phi) \tag{1.36a}$$

or

$$E_1 = E_{o1}\cos(\alpha_1 + \omega t) \tag{1.36b}$$

can be written (for real part only) as

$$E_1 = E_{o1}e^{i(\alpha_1+\omega t)}, \tag{1.37}$$

Suppose that there are N such overlapping waves having the same frequency and travelling in the positive x-direction. The resultant wave is given by

$$E = \left[\sum_{j=1}^{N} E_{oj} e^{i\alpha_j} \right] . e^{i\omega t} \tag{1.38}$$

The quantity

$$E_o e^{i\alpha} = \sum_{j=1}^{N} E_{oj} e^{i\alpha_j} \tag{1.39}$$

is called the complex amplitude of the composite wave and is simply the sum of the complex amplitudes of the constituents. Thus the resultant wave is

$$E = E_o e^{i(\alpha + \omega t)} . \tag{1.40}$$

Since

$$E_o^2 = \left(E_o e^{i\alpha} \right) . \left(E_o e^{i\alpha} \right)^* , \tag{1.41}$$

we can compute the resultant irradiance from equation (1.39) and (1.41). For a simple case taking N = 2, we have

$$E_o^2 = \left(E_{o1} e^{i\alpha_1} + E_{o2} e^{i\alpha_2} \right) . \left(E_{o1} e^{-i\alpha_1} + E_{o2} e^{-i\alpha_2} \right) \tag{1.42a}$$

thus

$$E_o^2 = E_{o1}^2 + E_{o2}^2 + E_{o1} E_{o2} \left(e^{i(\alpha_1 - \alpha_2)} + e^{-i(\alpha_1 - \alpha_2)} \right) \tag{1.42b}$$

or

$$E_o^2 = E_{o1}^2 + E_{o2}^2 + 2 E_{o1} E_{o2} \cos(\alpha_1 - \alpha_2) . \tag{1.42c}$$

This is an equation identical to equation (1.22), therefore, the two methods give the same conclusion.

1.4.2 The Addition of Waves of Different Frequencies

In the previous discussion we have been restricted to the superposition of waves of same frequency. In fact strictly monochromatic source does not exist. It will be more realistic to talk of quasi-monochromatic light, which is composed of a narrow range of frequencies. The study of such light led us to the important concepts of linewidth and coherence time.

1.4.2.1 Beats

Consider the disturbances arising from a combination of waves of different frequencies ω_1 and ω_2 with zero initial phase angle and have equal amplitudes. We have

$$E_1 = E_{o1} \cos(k_1x - \omega_1t) \qquad (1.43a)$$

and

$$E_2 = E_{o1} \cos(k_2x - \omega_2t) \qquad (1.43b)$$

The net wave

$$E = E_{o1}[\cos(k_1x - \omega_1t) + \cos(k_2x - \omega_2t)] \qquad (1.44)$$

Using the trigonometric identity

$$\cos A + \cos B = 2.\cos \tfrac{1}{2} (A+B).\cos \tfrac{1}{2} (A-B) \qquad (1.45)$$

We can rearrange the equation (1.44) as

$$E = 2.E_{o1}[\cos \tfrac{1}{2} [(k_1 + k_2).x - (\omega_1+\omega_2)t] \times [\cos \tfrac{1}{2} [(k_1 - k_2)x - (\omega_1-\omega_2)t] \qquad (1.46)$$

Now define the quantities ω_a and k_a, which are the average values of the angular frequency and wavenumber, respectively. Similarly the quantities ω_m and k_m designate the modulation frequency and modulation wavenumber, respectively. Mathematically,

$$\omega_a = \tfrac{1}{2} (\omega_1+\omega_2), \qquad k_a = \tfrac{1}{2} (k_1 + k_2) \qquad (1.47a)$$

and

$$\omega_m = \tfrac{1}{2} (\omega_1-\omega_2) \qquad k_m = \tfrac{1}{2} (k_1 - k_2) \qquad (1.47b)$$

thus

$$E = 2.E_{o1} \cos(k_mx - \omega_mt).\cos(k_ax - \omega_at) \qquad (1.48)$$

The total disturbance may be regarded as a travelling wave of frequency ω_a having a modulated amplitude $E_o(x,t)$ such that

$$E(x,t) = E_o(x,t) \cos(k_ax - \omega_at) \qquad (1.49)$$

where

$$E_o(x,t) = 2.E_{o1}\cos(k_m x - \omega_m t). \quad (1.50)$$

In general ω_1 and ω_2 are very large. If they are comparable to each other, i.e., $\omega_1 \sim \omega_2$, then $\omega_a >> \omega_m$ and $E_o(x,t)$ will change slowly, whereas $E(x,t)$ will vary quite rapidly (Figure 1.5). The irradiance is proportional to

$$E_o^2(x,t) = 4.E_{o1}^2\cos^2(k_m x - \omega_m t) \quad (1.51a)$$

or

$$E_o^2(x,t) = 2.E_{o1}^2(1+\cos(2.k_m x - 2.\omega_m t)). \quad (1.51b)$$

It is important that $E_o^2(x,t)$ oscillates about a value of $2.E_{o1}^2(x,t)$ with an angular frequency of $2\omega_m$ or simply $(\omega_1 - \omega_2)$, which is known as the beat frequency. In other words, E_o varies at the modulation frequency, whereas $E_o^2(x,t)$ varies at twice of this value.

Beats were first observed by the use of light in 1955. The advent of laser made the observation of beats using light very easy. Even a beat frequency of a few Hz out of 10^{14} Hz can be seen as a variation in phototube current. The observation of beats now represents a particularly sensitive and fairly simple means of detecting small frequency differences. For example, a modern version of Michelson interferometer that beats two light beams can gives the information of the frequency or linewidth of the source. The ring laser functioning as a gyroscope utilizes beats to measure frequency differences induced as a result of the rotation of the system. The Doppler effect, which accounts for the frequency shift when light is reflected off a moving surface, provides another series of applications of beats. By scattering light off a target and then beating the original and reflected waves, we get a precise measure of the target speed.

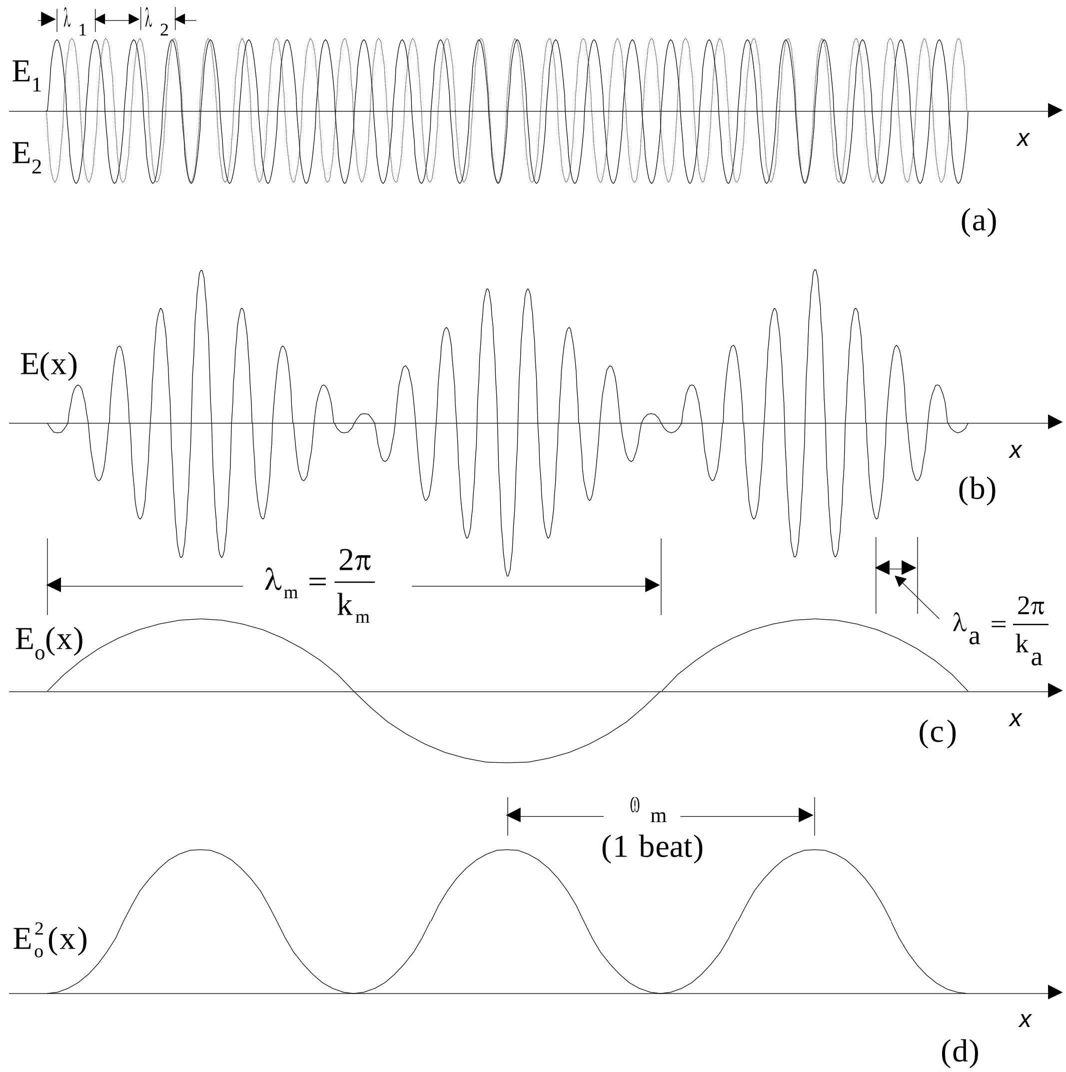


Figure 1.5 The superposition of two harmonic waves of different frequency.

Problems

1.1 Write expressions for the E- and B- fields that constitute a plane harmonic wave traveling in the + z-direction and having zero amplitudes for t = z = ϕ =0.

1.2 If we write the wave function in complex notation

$$E = E_o e^{i\phi}$$

show that E is unchanged when its phase increases or decreases by 2π.

1.3 Determine the resultant of the superposition of the waves

$$E(x,t) = E_{o1}.\sin[\omega t + \phi_1],$$

and

$$E(x,t) = E_{o2}.\sin[\omega t + \phi_2],$$

when $\omega = 120\pi$, $E_{o1} = 12$, $E_{o2} = 16$, $\phi_1 = 0$, and $\phi_2 = \pi/2$. Plot each function and result.

1.4 Show that when the two waves are represented as

$$E(x,t) = E_{o1}.\cos[\omega t + \alpha_1],$$

and

$$E(x,t) = E_{o2}.\cos[\omega t + \alpha_2],$$

are in phase, the resulting amplitude squared is a maximum equal to $(E_{o1} + E_{o2})^2$, and when they are out of phase it is a minimum equal to $(E_{o1} - E_{o2})^2$.

1.5 Show that the resultant of the two waves

$$E_1 = E_{o1}.\sin[\omega t - k(x + \Delta x)],$$

and

$$E_2 = E_{o2}.\sin[\omega t - kx],$$

is

$$E = 2.E_{o1}.\cos\left(\frac{k.\Delta x}{2}\right).\sin\left[\omega t - k\left(x + \frac{\Delta x}{2}\right)\right].$$

1.6 Determine the optical path difference for the two waves (of He-Ne laser, i.e., $\lambda = 632.8$ nm) when one traveled one centimeter through vacuum and the other passed through water, ($n = 1.333$) for the same distance.

1.7 Use the complex representation to find the resultant $E = E_1 + E_2$, where

$$\mathbf{E}_1 = E_o \cos(kx + \omega t)$$

and

$$\mathbf{E}_2 = E_o.\cos(kx - \omega t).$$

Describe the composite wave.

1.8 Determine the resultant when the two functions $\mathbf{E}_1 = 2.E_o.\cos \omega t$ and $\mathbf{E}_2 = \frac{1}{2}.E_o.\sin 2\omega t$ are superimposed. Draw $\mathbf{E}_1$, $\mathbf{E}_2$, and the resultant $\mathbf{E}$.

Books for further reading

V. Ronchi, *The Nature of Light: An Historical Survey*, (Heinemann, London, 1970)

R. Ditteon, *Modern Geometrical Optics* (John-Wiley & Sons, NY, 1998).

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed. (McGraw-Hill, New York, 1985).

E. Hecht, *Optics*, 2nd ed. (Addison-Wesley, Reading, Mass., 1990).

R. Guenther, *Modern Optics*, (John Wiley & Sons, New York, 1990).

M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1986).

Appendix 1A

Maxwell's Equations

In mid-nineteen century, Maxwell showed that light was a form of electromagnetic radiation. He combined the laws of electricity and magnetism to show that electromagnetic wave existed. A simplified version of Maxwell's equations is given below. More rigorous derivation can be found in *Optics* by E. Hecht.

The Gauss's law for electricity can be written in the form

$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{q}{\varepsilon}, \tag{1A.1}$$

where q is the charge enclosed by the surface and ε is a property of the medium called permittivity. Equation (1A.1) leads to Coulomb's law and allows calculating electric field strength for simple charge distributions.

There is also a Gauss's law for magnetism:

$$\oint \mathbf{B} \cdot d\mathbf{A} = 0. \tag{1A.2}$$

The fact that the right-hand side of equation (1A.2) is observed to be zero implies that there are no magnetic mono-poles (i.e., single, isolated magnetic sources).

Faraday's law

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{d\phi_B}{dt}, \tag{1A.3}$$

shows that electric field can be created by changing magnetic field. In this equation, ϕ_B is the magnetic flux through the area defined by the integral.

Finally, the Ampere-Maxwell law is given by

$$\oint \mathbf{B} \cdot d\mathbf{s} = -\mu\varepsilon \frac{d\phi_E}{dt} + \mu I, \tag{1A.4}$$

where m is the magnetic permeability of the medium. This equation means that magnetic fields are created by currents or changing electric fields.

When discussing the propagation of light, we can simplify these equations because the material through which the light propagates is normally not a conductor. Therefore, there are no free charges (q = 0) or currents (I = 0) to contend with.

The equations given above are in the familiar integral form. For our purposes, the differential form is preferable. These equations can be converted from one form to another via Gauss's divergence theorem and Stokes's theorem from vector calculus. The above four equations are equivalent to the following four equations with q = 0 and I = 0.

$$\nabla \cdot \mathbf{E} = 0, \qquad (1A.5)$$

$$\nabla \cdot \mathbf{B} = 0, \qquad (1A.6)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad (1A.7)$$

$$\nabla \times \mathbf{B} = \mu\varepsilon \frac{\partial \mathbf{E}}{\partial t}. \qquad (1A.8)$$

These expressions can yields a wave equation. To derive expression for waveform, take curl of equation (1A.8)

$$\nabla \times (\nabla \times \mathbf{B}) = \mu\varepsilon \frac{\partial}{\partial t}(\nabla \times \mathbf{E}), \qquad (1A.9)$$

since space and time are completely independent, therefore, their derivative can be interchanged. Equation (1A.7) can be substituted in equation (1A.9) to obtained the second derivative

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{B}) &= \mu\varepsilon \frac{\partial}{\partial t}\left(-\frac{\partial \mathbf{B}}{\partial t}\right) \\ &= -\mu\varepsilon \frac{\partial^2 \mathbf{B}}{\partial t^2} \end{aligned} \qquad (1A.10)$$

The vector triple product can be simplified by making use of the operator identity

$$\nabla \times (\nabla \times \mathbf{B}) = \nabla(\nabla \cdot \mathbf{B}) - \nabla^2 \mathbf{B}. \qquad (1A.11)$$

Since the divergence of **B** is zero, thus equation (1A.10) becomes

$$\nabla^2 \mathbf{B} = \mu\varepsilon \frac{\partial^2 \mathbf{B}}{\partial t^2}. \tag{1A.12}$$

A similar equation is satisfied by the electric field intensity. Following essentially the same procedure as above, take the curl of equation (1A.7),

$$\nabla \times (\nabla \times \mathbf{E}) = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}). \tag{1A.13}$$

Eliminating B from this expression by substituting equation (1A.8)

$$\nabla \times (\nabla \times \mathbf{E}) = -\mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}, \tag{1A.14}$$

using the identity of equation (1A.11) we get

$$\nabla^2 \mathbf{E} = \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{1A.15}$$

The equations (1A.12 & 1A.15) are called differential wave equations. In free space the medium constant μ and ε are replaced by μ_o and ε_o and these equations simply becomes

$$\nabla^2 \mathbf{B} = \mu_o\varepsilon_o \frac{\partial^2 \mathbf{B}}{\partial t^2} \tag{1A.16}$$

and

$$\nabla^2 \mathbf{E} = \mu_o\varepsilon_o \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{1A.17}$$

Unit-2

Wave Optics

Objective

Light has a dual nature and there are some phenomena which can only be explained if the light is considered to have a wave nature, e.g. polarization, interference and diffraction. These phenomena of optics are generally grouped in the form of wave optics. In lasers wave optics has fundamental importance, therefore, this unit describes the basic principles and physical understanding of these fundamental aspect.

2.1 Polarization

In the previous unit we have learned that light is a transverse electromagnetic wave. The electric field vector, **E**, oscillates in magnitude as well as in direction but always remains perpendicular to the propagation direction. In any real beam, light comprises many individual waves and in general the planes of vibrations of their electric field will be randomly orientated. Such a beam of light is unpolarized and the resultant electric field vector changes orientation randomly in time. It is possible, however, to have light beams characterized by highly orientated electric fields and such light is referred to as being polarized. The simplest form of polarization is linearly polarized light in which electric field vector oscillates only along a straight line.

2.1.1 Linear Polarization

Let us consider two harmonics, linearly polarized light waves of the same frequency, and moving in the same direction. If the electric field vectors are collinear, the superimposing disturbances will simply combine to form a resultant linearly polarized wave. We can represent the two orthogonal optical disturbances in the form

$$\mathbf{E}_x(z,t) = \mathbf{i}\, E_{ox} \cos(kz - \omega t) \tag{2.1}$$

and

$$\mathbf{E}_y(z,t) = \mathbf{j}\, E_{oy} \cos(kz - \omega t + \phi) \tag{2.2}$$

where ϕ is the relative phase difference between the waves, both of which are travelling in the z-direction. Since the phase is in the form $(kz - \omega t)$, the addition of a positive ϕ means that the cosine function in equation (2.2) will not attain the same value as the cosine in Equation (2.1) until a latter time (ϕ/ω). Accordingly, E_y lags E_x by $\phi>0$. If ϕ is a negative quantity, E_y leads E_x by $\phi<0$. The resultant optical disturbance is the vector sum of these two perpendicular waves:

$$\mathbf{E}(z,t) = \mathbf{E}_x(z,t) + \mathbf{E}_y(z,t). \tag{2.3}$$

If ϕ is zero or an integral multiple of $\pm 2\pi$, the waves are said to be in phase. In that particular case Equation (2.3) becomes

$$\mathbf{E}(z,t) = (\mathbf{i}\, E_{ox} + \mathbf{j}\, E_{oy}).\cos(kz - \omega t) \tag{2.4}$$

The resultant wave therefore has a fixed amplitude equal to $(\mathbf{i}\, E_{ox} + \mathbf{j}\, E_{oy})$; in other words the resultant wave is also linearly polarized as shown in Figure 2.1. The waves advance toward a plane of observation where the fields are to be measured. There one sees a single resultant **E** oscillating, along a tilted line, co-sinusoidally in time. The E-field progresses through one complete oscillatory cycle as the wave advances along the z-axis through one wavelength. This process of addition can be carried out equally well in reverse; that is we can resolve any plane-polarized wave into two orthogonal components. If ϕ is an odd integer multiple of $\pm\pi$, the two waves are said to be 180^o out of phase, and

$$\mathbf{E}(z,t) = (\mathbf{i}\, E_{ox} - \mathbf{j}\, E_{oy}) \cos(kz - \omega t) \tag{2.5}$$

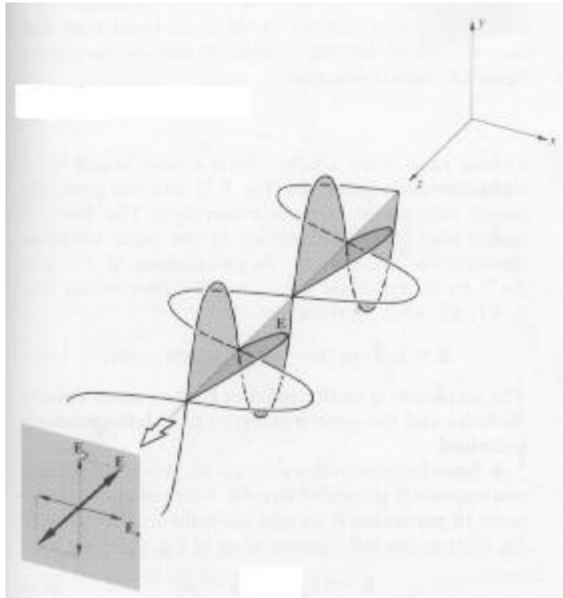


Figure 2.1 Linear Light

The wave is again linearly polarized, but the plane of vibration has been rotated from that of the previous condition, as indicated in Figure 2.2.

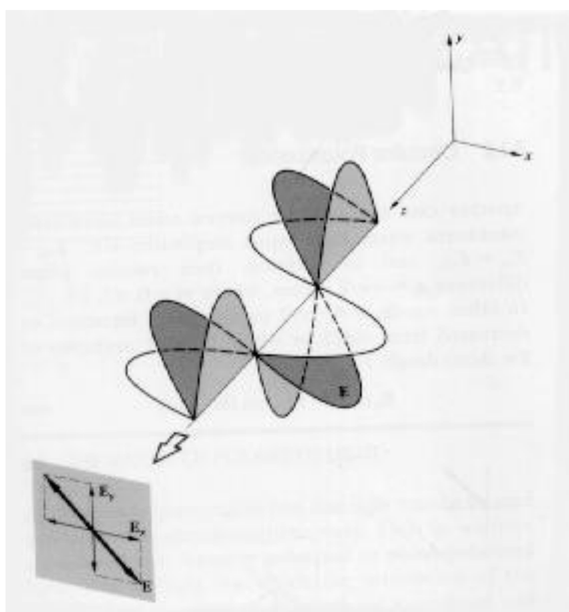


Figure 2.2 Linear Light

2.1.2 Circular Polarization

An interesting situation arises when the amplitudes of the two superimposing beam are equal (i.e., $E_{ox} = E_{oy} = E_o$), and in addition, their relative phase difference $\phi = -\pi/2 \pm 2m\pi$, where m is an integer. In other words, $\phi = -\pi/2$ or any value increased or decreased by from $-\pi/2$ by a whole number multiples of 2π. Accordingly

$$\mathbf{E}_x(z,t) = \mathbf{i}\, E_o \cos(kz - \omega t) \qquad (2.6a)$$

and

$$\mathbf{E}_y(z,t) = \mathbf{j}\, E_o \sin(kz - \omega t). \qquad (2.6b)$$

The consequent wave is given by

$$\mathbf{E}(z,t) = E_o\, [\mathbf{i} \cos(kz - \omega t) + \mathbf{j} \sin(kz - \omega t)] \qquad (2.7)$$

and is shown in Figure 2.3. In equation (2.7) the scalar amplitude of **E**, that is, $(\mathbf{E.E})^{1/2} = E_o$, is constant. The direction of **E** is time varying and it is not restricted to a single plane. Figure 2.4 shows what is happening at some arbitrary point z_o on the axis. At t = 0, **E** lies along the reference axis in Figure 2.4, and so

$$\mathbf{E}_x(z,t) = \mathbf{i}\, E_o \cos(kz_o) \qquad (2.8a)$$

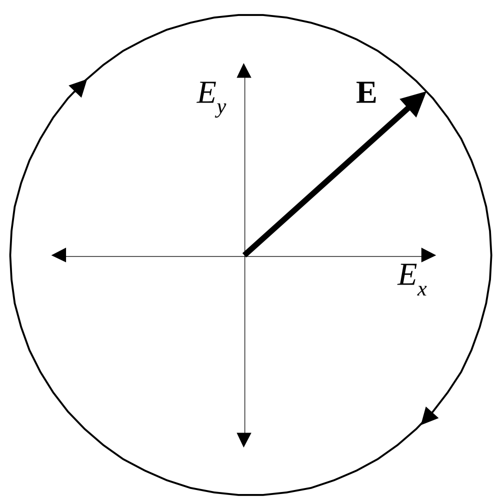


Figure 2.3 Right-circularly polarized light.

and

$$\mathbf{E}_y(z,t) = \mathbf{j}\, E_o \sin(kz_o). \qquad (2.8b)$$

The rotation rate is ◆ and kz=□/4.

At a latter time, $t = kz_o/\omega$, $\mathbf{E}_x = \mathbf{i}\, E_o$, $\mathbf{E}_y = 0$, and **E** is along the x-axis. The resultant electric

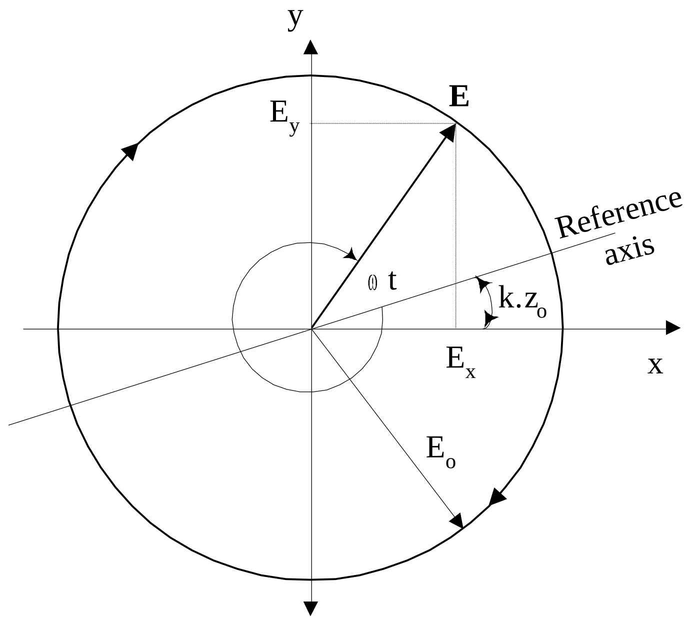


Figure 2.4 Rotation of electric vector in a right-circular wave.

field vector **E** is rotating clockwise at an angular frequency of ω, as seen by an observer toward whom the wave is moving (i.e., looking back at the source). Such a wave is said to be right-circularly polarized, and one generally simply refers to it as right-circular light. The **E**-vector makes one complete rotation as the wave advances through one wavelength. In comparison, if $\phi = \pi/2$, $5\pi/2$, $9\pi/2$, and so on (i.e., $\phi = \pi/2 \pm 2m\pi$, for m to be an integer), then

$$\mathbf{E}(z,t) = E_o\,[\mathbf{i}\cos(kz - \omega t) - \mathbf{j}\sin(kz - \omega t)]. \tag{2.9}$$

The amplitude is unaffected, but **E** in this case rotates counter-clockwise, and the wave is referred to as left-circularly polarized.

A linearly polarized wave can be raised from two oppositely polarized circular waves of equal amplitude. In particular, if we add the right-circular wave of equation (2.7) and to the left-circular wave of equation (2.9), we get

$$\mathbf{E}(z,t) = 2\,E_o\,\mathbf{i}\cos(kz - \omega t), \tag{2.10}$$

which has a constant amplitude vector of $2E_o\mathbf{i}$ and is therefore linearly polarized.

2.1.3 Elliptical Polarization

As for as the mathematical description is concerned, both linear and circular polarization light may be considered to be special cases of elliptically polarized light. This means that, in general, the resultant electric field vector **E** will rotate and change its magnitude as well. In such cases the endpoint of **E** will trace out an ellipse, in a fixed space perpendicular to **k**. We can see this better by actually writing an expression for the curve traced by the tip of **E**. We can recall he equations (2.1) and (2.2) as

$$E_x = E_{ox} \cos(kz - \omega t) \tag{2.11}$$

and

$$E_y = E_{oy} \cos(kz - \omega t + \phi). \tag{2.12}$$

The above equations can be rearranged to get rid of the (kz – ωt) dependence. Expand the expression for E_y into

$$E_y/E_{oy} = \cos(kz - \omega t).\cos\phi - \sin(kz - \omega t).\sin\phi \tag{2.13}$$

and combine it with E_x/E_{ox} to have

$$\frac{E_y}{E_{oy}} - \frac{E_x}{E_{ox}}\cos\phi = -\sin(kz - \omega t).\sin\phi. \tag{2.14}$$

From equation (2.11) we have

$$\sin(kz - \omega t) = [1 - (E_x/E_{ox})^2]^{1/2}, \tag{2.15}$$

therefore equation (2.14) becomes

$$\left(\frac{E_y}{E_{oy}} - \frac{E_x}{E_{ox}}\cos\phi\right)^2 = \left[1 - \left(\frac{E_x}{E_{ox}}\right)^2\right].\sin^2\phi. \tag{2.16}$$

On rearranging terms, we have

$$\left(\frac{E_y}{E_{oy}}\right)^2 + \left(\frac{E_x}{E_{ox}}\right)^2 - 2\left(\frac{E_x}{E_{ox}}\right).\left(\frac{E_y}{E_{oy}}\right).\cos\phi = \sin^2\phi. \tag{2.17}$$

This is an equation of an ellipse making an angle α with the (E_x, E_y) coordinate system (Figure 2.5) such that

$$\tan 2\alpha = \frac{2.E_{ox}.E_{oy}.\cos\phi}{E_{ox}^2 - E_{oy}^2}. \qquad (2.18)$$

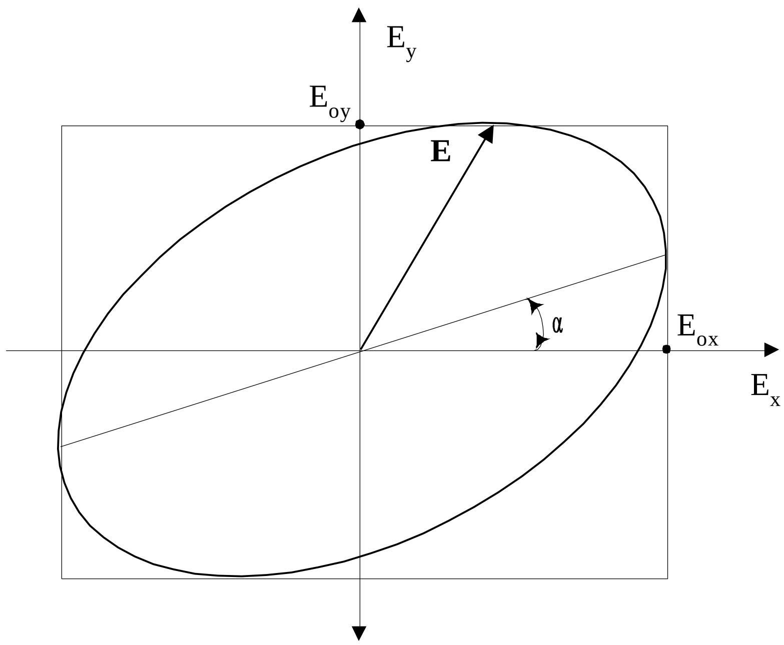


Figure 2.5 Elliptically polarized light.

Equation (2.17) will look like a simple equation of ellipse if the principle axes of the ellipse were aligned with the coordinate axes, that is α = 0 or equivalently ϕ = $\pm\pi/2$, $\pm 3\pi/2$, $\pm 5\pi/2$, …, in which case we have the familiar form

$$\left(\frac{E_y}{E_{oy}}\right)^2 + \left(\frac{E_x}{E_{ox}}\right)^2 = 0. \qquad (2.19)$$

Furthermore, if $E_{oy} = E_{ox} = E_o$, this equation can be reduced to

$$E_y^2 + E_x^2 = E_o^2, \qquad (2.20)$$

which is the equation of a circle, and is in complete agreement with our previous result. If ϕ is an even multiple of π, equation (2.17) results in

$$E_y = \frac{E_{oy}}{E_{ox}} E_x \tag{2.21}$$

and similarly for odd multiples of π, we have

$$E_y = -\frac{E_{oy}}{E_{ox}} E_x . \tag{2.22}$$

Both the equation (2.21) and (2.22) are equations of straight lines having slopes of $\pm E_{oy}/E_{ox}$; in other words we have linearly polarized light.

Figure 2.6 diagrammatically summarize most of these conclusions. This important diagram is labeled across the bottom "E_x leads E_y by: 0, $\pi/4$, $\pi/2$, $3\pi/4$,...", where these are the positive values of ϕ to be used in equation (2.2). The same set of curves will occur if "E_y leads E_x by: 2π, $7\pi/4$, $3\pi/2$, $5\pi/4$, ...", and that happens when ϕ equals -2π, $-7\pi/4$, $-3\pi/2$, $-5\pi/4$, and so on. Figure 2.7 illustrates how E_x leading E_y by $\pi/2$ is equivalent to E_y leading E_x by $3\pi/2$ (where sum of these two angles equals to 2π).

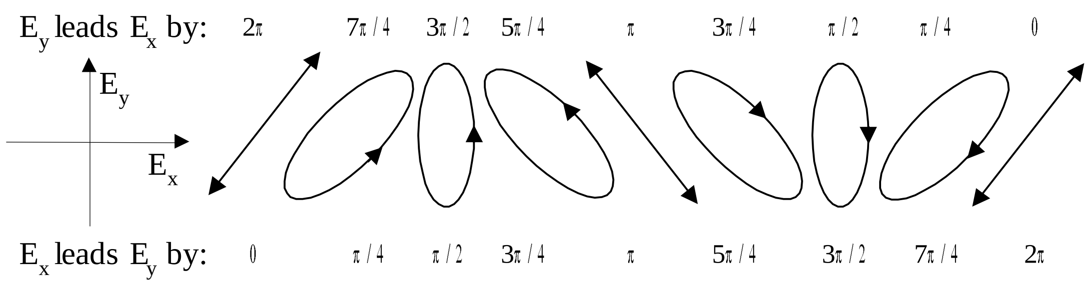


Figure 2.6 Various polarization configurations. The light would be circular with $\phi = \pi/2$ or $3\pi/2$ if $E_{ox} = E_{oy}$, but here for the sake of generality E_{oy} was taken to be larger than E_{ox}.

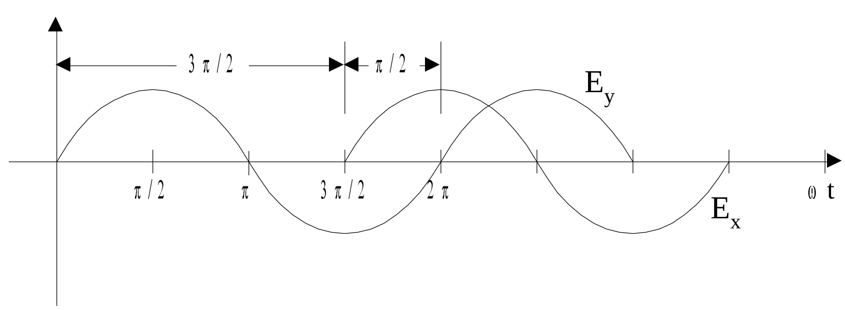


Figure 2.7 Phases of the electric field vector when E_x leads E_y (or E_y lags E_x) by $\pi/2$, or alternatively, E_y leads E_x (or E_x lags E_y) by $3\pi/2$.

2.1.4 Polarizers

An optical device whose input is natural light and whose output is some form of polarized light is known as polarizer. An instrument that separates the two components of electric field, discarding one and passing on the other, is known as linear polarizer. Depending on the form of the output, we could also have circular or elliptical polarizers. All these devices vary in effectiveness down to what might be called leaky or partial polarizer.

Polarizers come in many different configurations, but all of them are based on one of the following fundamental physical mechanisms.

2.1.4.1 Dichroism

Dichroism refers to the selective absorption of one of the two orthogonal electric field components of an incident beam. The dichroic polarizer itself is physically anisotropic, producing a strong asymmetric or preferential absorption of one field component while being essentially transparent to the other.

Birefringence: In some crystalline substances (i.e., solids whose atoms are arranged in some sort of regular repetitive array) the optical properties are not the same in all direction. They have different refractive indices in different directions. A material of this kind, which displays different indices of refraction, is said to be birefringent. A crystal so illuminated will be strongly absorbing for one polarization direction and transparent for the other. Thus a birefringent material in fact is dichroic.

2.1.4.2 Reflection

If a beam of white light is incident at one certain angle on the polished surface of ordinary glass, it is found upon reflection to be plane polarized. This is perhaps the simplest method of polarizing light and was discovered by Malus in 1808.

2.1.4.3 Scattering

The orientation of the electric field of the scattered radiation follows the dipole pattern. The vibrations induced in the atom are parallel to the **E**-field of the incoming light wave and so are perpendicular to the propagation direction. An oscillating dipole does not radiate in the direction of its axis. If the incident wave is unpolarized, the scattered light in the forward direction is completely unpolarized. It becoming increasingly more polarized as the angle increases. When the direction of observation is normal to the primary beam, the light is completely polarized.

2.1.5 Law of Malus

This law tells us how the intensity transmitted by the analyzer varies with the angle that its plane of transmission makes with that of polarizer. If natural light is incident on an ideal linear polarizer, as in Figure 2.8, only plane polarized light will be transmitted. The polarized light will have an orientation parallel to a specific direction, which may call the transmission axis of the polarizer. In other words, only the component of the optical field parallel to the transmission axis will pass through the polarizer unaffected.

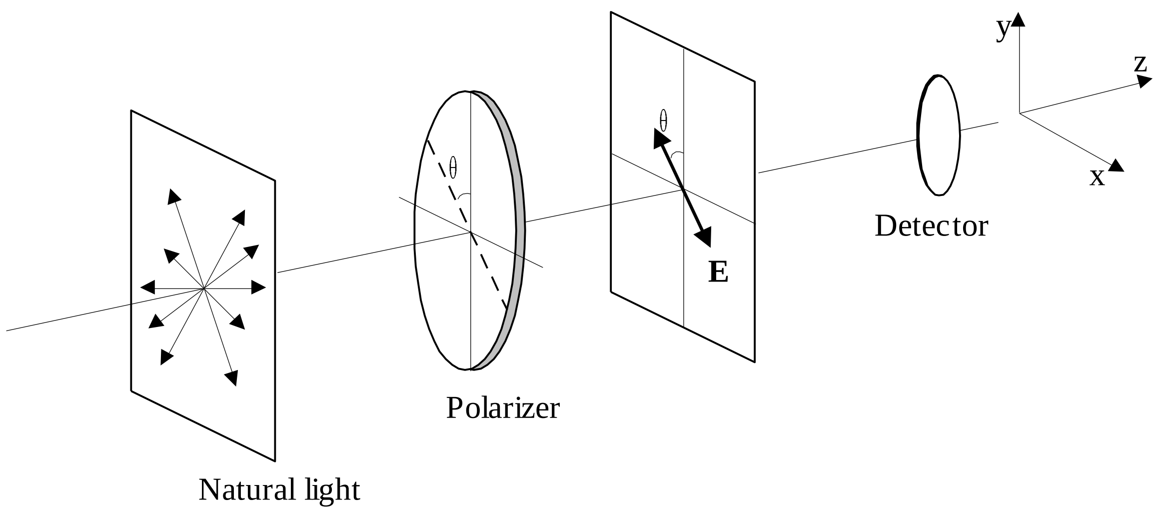


Figure 2.8. A linear polarizer.

Now introduce a second identical ideal polarizer, as analyzer, whose transmission axis is vertical. If the amplitude of the electric field transmitted by the polarizer is E_o, only its component, $E_o.\cos\theta$, parallel to the transmission axis of the analyzer will be passed on to the detector (Figure 2.9). The irradiance reaching to the detector is given by

$$I(\theta) = E_o^2 \cos^2 \theta \,. \qquad (2.23)$$

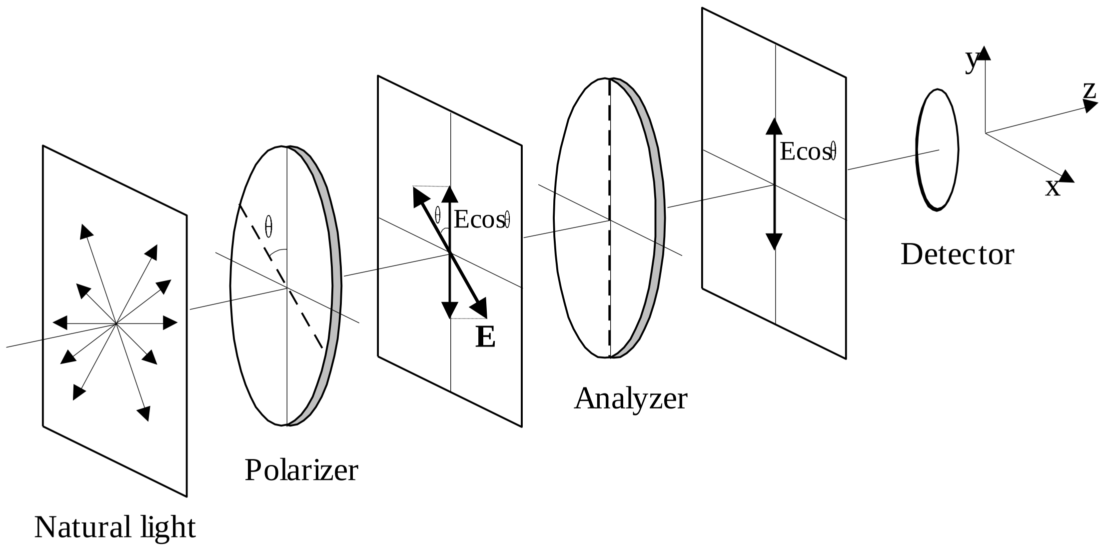


Figure 2.9 A linear polarizer and analyzer- Malus's law.

The maximum irradiance, I_o, will occur when the angle θ between the transmission axis of the analyzer and polarizer is zero. The above equation can be rewritten as

$$I(\theta) = I_o.\cos^2\theta. \qquad (2.24)$$

This is known as Malus's law. It is clear from equation (2.24) that for θ is equal to 90^o, the output intensity after the analyzer will be zero. This is due to the fact that the electric field that has passed through the polarizer is perpendicular to the transmission axis of the analyzer. The electric field is therefore parallel to the extinction axis of the analyzer and therefore, no component is along the transmission axis. We can use this setup along with Malus's law to determine the linear polarization of an optical beam.

2.1.6 Optical Activity

The phenomenon of optical activity was first observed by French scientist D.F.J. Arago in 1811. He discovered that the plane of vibration of a beam of linear light underwent a continuous rotation as it propagate along the optic axis of a quartz crystal (Figure 2.10). Latter the same effect was observed in vapors and liquid forms of various natural substances. Thus a material that causes the **E**-field of an incident linear plane wave to appear to rotate is said to be optically active. Moreover, there are some right-handed and left-handed rotations. If we look in the direction of the source and the plane of vibration appears to have revolved clockwise, the substance is referred to as dextroratory, *d-rotatory* (dextro meaning right). If **E**-field appears to rotate counterclockwise, the material is said to be *l-rotatory*.

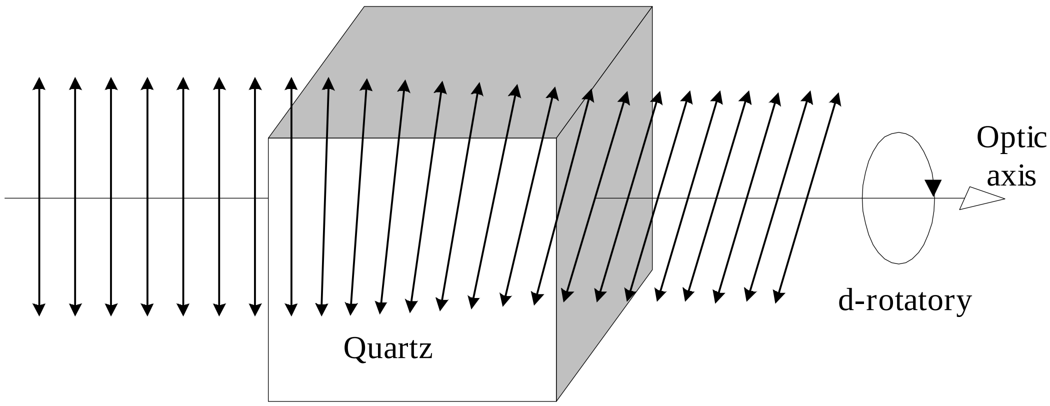


Figure 2.10 Optical activity displayed by quartz.

In 1825, Fresnel proposed a simple phenomenological description of optical activity. Since the incident light wave can be represented as a superposition of right and left circular light, he suggested that these two forms of circular light propagates at different speeds. Such material possesses two indices of refraction, one for right circular light (n_R) and one for left circular light (n_L). In traversing an optically active specimen, the two circular waves would get out of phase, and the resultant linear wave would appear to have rotated. Eqs. (2.7) and (2.9) described right and left circular light propagating in the z-direction. We can rewrite these equations as

$$\mathbf{E}_R = E_o\,[\mathbf{i}\cos(k_R z - \omega t) + \mathbf{j}\sin(k_R z - \omega t)] \qquad (2.25a)$$

and

$$\mathbf{E}_L = E_o\,[\mathbf{i}\cos(k_L z - \omega t) - \mathbf{j}\sin(k_L z - \omega t)] \qquad (2.25b)$$

where $k_R = k_o n_R$, and $k_L = k_o n_L$. The resultant disturbance is given by $\mathbf{E} = \mathbf{E}_R + \mathbf{E}_L$, therefore we have

$$\mathbf{E}_L = 2.E_o.\cos((k_R+k_L).z/2 - \omega t)\,[\mathbf{i}\cos(k_R-k_L)z/2) - \mathbf{j}\sin(k_R-k_L)z/2)] \qquad (2.26a)$$

At the position where the wave enters the medium (i.e., z = 0) it is linearly polarized along the x-axis, that is,

$$\mathbf{E} = 2.E_o.\mathbf{i}.\cos\omega t. \qquad (2.27)$$

At any point along the path, the two components have the same time dependence and are therefore in phase. This just means that anywhere along the z-axis the resultant is linearly polarized (Figure 2.11), although its orientation is a function of z. Moreover, if $n_R>n_L$ (or $k_R>k_L$), **E** will rotate counterclockwise, whereas if $k_L>k_R$, rotation is clockwise.

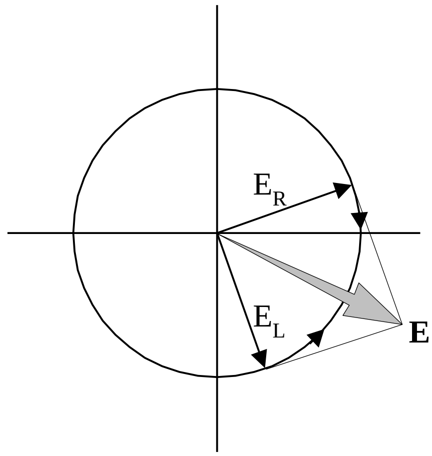


Figure 2.11 The superposition of a Right- and Left- circular light for $k_L>\mathbf{k}_R$.

2.2 Interference

In the previous chapter we know that two beams of light can be made to cross each other without either one producing any effect on the other after it passes beyond the region of crossing. In this sense the two beams do not interfere with each other. However, in the region of crossing, where both beams are acting at once, we observe that resultant amplitude and intensity is very different from the sum of the two beams acting separately. This modification of intensity obtained by the superposition of two or more beams of light we call interference. If the resultant intensity is zero or in general less than we expect from the separate intensities, we have destructive interference. While if the resultant is greater, we have constructive interference. After the dominating century of corpuscular theory of light, Thomas Young, in 1801, performed the historical experiment of the interference of light.

We have derived the expressions of the superposition of two scalar waves, and in many respects those results will again be applicable. But light is a vector phenomenon; the electric and magnetic fields are vector fields. In accordance with the principle of superposition, the electric field intensity **E**, at a point in space, arising from the separate fields $\mathbf{E}_1$, $\mathbf{E}_2$, … of various contributing sources is given by

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \ldots$$

The light field **E** varies in time at a very rapid rate (roughly ~5x10[14] Hz) making the actual field an impractical quantity to detect. But the irradiance 'I' can be measured directly with a wide variety of sensors (e.g., photocells, photographic emulsions, eye, etc.). Therefore, we can study the interference by measuring the irradiance.

To study interference, consider two point sources, S_1 and S_2, emitting monochromatic waves of the same frequency in a homogeneous medium. Furthermore, let their separation 'a' be much greater that λ. Locate the point of observation P far enough away from the sources so that at P the wavefronts will be planes (Figure 2.12). For the moment, we will consider only

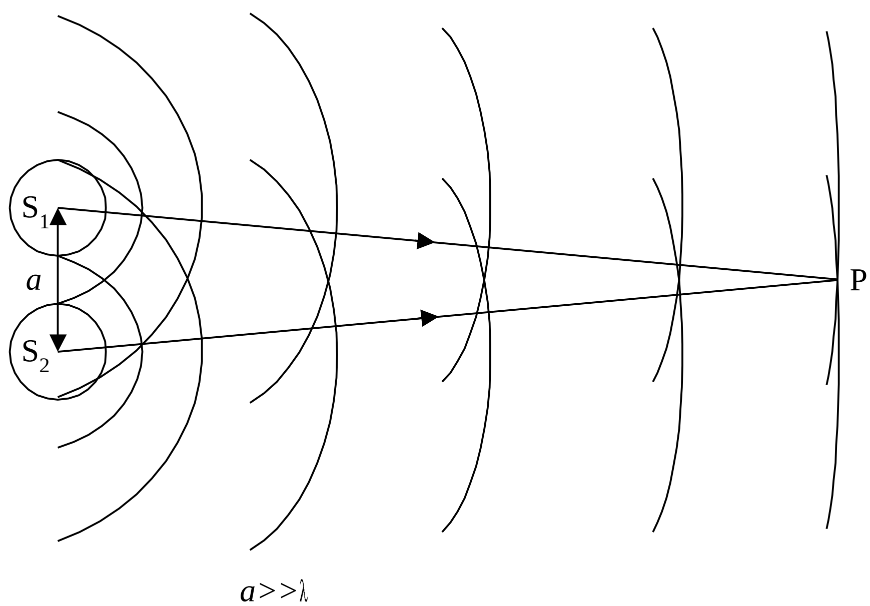


Figure 2.12 Waves from two point sources overlapping in space.

linearly polarized waves of the form

$$\mathbf{E}_1(\mathbf{r},t) = \mathbf{E}_{o1}\cos(\mathbf{k_1}.\mathbf{r} - \omega t + \phi_1) \tag{2.28a}$$

and

$$\mathbf{E}_2(\mathbf{r},t) = \mathbf{E}_{o2}\cos(\mathbf{k_2}.\mathbf{r} - \omega t + \phi_2). \tag{2.28b}$$

The irradiance at P is given by

$$I \propto \langle \mathbf{E}^2 \rangle \tag{2.29a}$$

Since we will be concerned only with relative irradiance within the same medium, we simply neglect the constant of proportionality and set

$$I = \langle \mathbf{E}^2 \rangle \tag{2.29b}$$

The $\langle \mathbf{E}^2 \rangle$ is the time average of the magnitude of the electric field intensity squared, or $\langle \mathbf{E.E} \rangle$. Accordingly

$$\mathbf{E}^2 = \mathbf{E.E} = (\mathbf{E}_1 + \mathbf{E}_2) . (\mathbf{E}_1 + \mathbf{E}_2) \tag{2.30a}$$

and thus

$$\mathbf{E}^2 = \mathbf{E}_1^2 + \mathbf{E}_2^2 + 2\, \mathbf{E}_1 \,.\, \mathbf{E}_2 \tag{2.30b}$$

Taking the time average on both sides, we find the irradiance becomes

$$I = I_1 + I_2 + I_{12}, \tag{2.31}$$

provided that

$$I_1 = \langle \mathbf{E}_1^2 \rangle,\ I_2 = \langle \mathbf{E}_2^2 \rangle, \text{ and } I_{12} = 2 \langle \mathbf{E}_1 \,.\, \mathbf{E}_2 \rangle. \tag{2.32}$$

The last term is known as the interference term. From equation (28) we can evaluate the last term as

$$\mathbf{E}_1 \,.\, \mathbf{E}_2 = \mathbf{E}_{o1} \,.\, \mathbf{E}_{o2} \cos(\mathbf{k_1.r} - \omega t + \phi_1) \times \cos(\mathbf{k_2.r} - \omega t + \phi_2) \tag{2.33a}$$

or equivalently

$$\mathbf{E}_1 \,.\, \mathbf{E}_2 = \mathbf{E}_{o1} \,.\, \mathbf{E}_{o2} [\cos(\mathbf{k_1.r} + \phi_1). \cos\omega t + \sin(\mathbf{k_1.r} + \phi_1). \sin\omega t]$$

$$\times [\cos(\mathbf{k_2.r} + \phi_2). \cos\omega t + \sin(\mathbf{k_2.r} + \phi_2). \sin\omega t] \tag{2.33b}$$

Using the time average values of <coswt> = $\frac{1}{2}$, <sinwt> = $\frac{1}{2}$, and <coswt. Sinwt> = 0, and simplifying the equation (2.30b), we have

$$\mathbf{E}_1 \,.\, \mathbf{E}_2 = \tfrac{1}{2}\, \mathbf{E}_{o1} \,.\, \mathbf{E}_{o2} \cos(\mathbf{k_1.r} + \phi_1 - \mathbf{k_2.r} - \phi_2). \tag{2.34a}$$

The interference term is then

$$I_{12} = \mathbf{E}_{o1} \,.\, \mathbf{E}_{o2} \cos\delta, \tag{2.34b}$$

where $\delta = (\mathbf{k_1.r} + \phi_1 - \mathbf{k_2.r} - \phi_2)$, is the phase difference arising from a combined path-length and initial phase angle difference. It is important to note that if $\mathbf{E}_{o1}$ and $\mathbf{E}_{o2}$ are perpendicular, then their dot product will be zero, i.e., $I_{12} = 0$ and $I = I_1 + I_2$. Two such orthogonal waves not interfere but yield some other polarized states.

The most common situation corresponds to $\mathbf{E}_{o1}$ parallel to $\mathbf{E}_{o2}$. In that case, the irradiance reduces to the value found in the scalar treatment of superposition of waves. Under those conditions

$$I_{12} = E_{o1}.\ E_{o2}.\ \cos\delta. \qquad (2.35)$$

Again by taking the time average, we can write

$$I_1 = \langle \mathbf{E}_1^2 \rangle = \tfrac{1}{2} E_{o1}^2, \text{ and} \qquad I_2 = \langle \mathbf{E}_2^2 \rangle = \tfrac{1}{2} E_{o2}^2. \qquad (2.36)$$

The interference term becomes

$$I_{12} = 2\sqrt{I_1 I_2}\,\cos\delta, \qquad (2.37)$$

and the total irradiance is

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\cos\delta. \qquad (2.38)$$

Equation (2.38) shows that, at various points in space the resultant irradiance can be greater, less than, or equal to $I_1 + I_2$, depending on the value of I_{12} or in other words δ. A maximum in the irradiance is obtained when $\cos\delta = 1$, so that

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \qquad (2.39)$$

when

$$\delta = 0,\ \pm 2\pi,\ \pm 4\pi,\ \ldots$$

In this case the path difference between the two waves is an integer multiple of 2π, and the disturbances are said to be in phase. This can also be called as total constructive interference. When $0 < \cos\delta < 1$ the waves are out of phase, $I_1 + I_2 < I < I_{max}$, and the result is known as constructive interference. At $\delta = \pi/2$, $\cos\delta = 0$, the optical disturbances are said to be 90^o out of phase, and $I = I_1 + I_2$. For $0 > \cos\delta > -1$, we have the condition of destructive interference,

$I_1 + I_2 > I > I_{min.}$ The minimum of the irradiance results when the waves are 180° out of phase, troughs overlap crests, i.e., cosδ = –1 and

$$I_{min.} = I_1 + I_2 - 2\sqrt{I_1 I_2} \qquad (2.40)$$

This occurs when $\delta = \pm\pi, \pm 3\pi, \pm 5\pi, \ldots$ and it is referred as total destructive interference.

When amplitudes of both the waves reaching at P are equal (i.e., $\mathbf{E}_{o1} = \mathbf{E}_{o2}$), the irradiance from both sources are then be equal. Let $I_1 = I_2 = I_o$, then equation (2.35) can be written as

$$I = 2.I_o\,(1 + \cos\delta) = 4.I_o.\cos^2(\delta/2) \qquad (2.41)$$

Equation (2.41) clearly shows that irradiance maxima occur (i.e., $I_{max.} = 4I_o$) when

$$\delta = 2\pi m, \qquad \text{for } m = 0, \pm 1, \pm 2, \ldots \qquad (2.42a)$$

Similarly minima, for which $I_{min.} = 0$, arises when

$$\delta = \pi m', \qquad \text{for } m' = \pm 1, \pm 3, \pm 5, \ldots \qquad (2.42b)$$

The above equations equally holds for spherical waves emitted by S_1 and S_2. Such waves can be expressed as

$$\mathbf{E}_1(r_1,t) = \mathbf{E}_{o1}(r_1)\exp[i(k.r_1 - \omega t + \phi_1)] \qquad (2.43a)$$

and

$$\mathbf{E}_2(r_2,t) = \mathbf{E}_{o2}(r_1)\cos[i(k.r_2 - \omega t + \phi_2)]. \qquad (2.43b)$$

The terms r_1 and r_2 are the radii of the spherical wavefronts overlapping at P; in other words they specify the distances from the sources to P. In this case

$$\delta = k(r_1 - r_2) + (\phi_1 - \phi_2) \qquad (2.44)$$

Using equation (44) and (42), the expression for maximum and minimum irradiance can be rewritten as

$$(r_1 - r_2) = [2\pi m + (\phi_2 - \phi_1)]/k \qquad \text{for maximum irradiance} \qquad (2.45a)$$

and

$$(r_1 - r_2) = [\pi m' + (\phi_2 - \phi_1)]/k \qquad \text{for minimum irradiance} \qquad (2.45b)$$

If the waves are in phase at the source, i.e., $(\phi_2 - \phi_1) = 0$, the equation (2.45) can be simplified as

$$(r_1 - r_2) = 2\pi m/k = m\lambda \tag{2.46a}$$

and

$$(r_1 - r_2) = \pi m'/k = \tfrac{1}{2} m'\lambda \tag{2.46b}$$

for maximum and minimum irradiance, respectively.

From the above discussion we can find the condition for interference of the two linearly polarized waves as

1. Two orthogonal coherent linearly polarized waves cannot interfere in the sense that $I_{12} = 0$ and no fringes result.
2. Two parallel, coherent linearly polarized waves will interfere in the same way as of natural light.

2.2.1 Wavefront-Splitting Interferometers

Interference apparatus may be divided into two main classes, (a) based on the division of wavefront and (b) based on the division of amplitude. The former class is the one, which was experimented first, in which the wavefront is divided laterally into segments by mirrors or diaphragms. It is also possible to divide a wave by partial reflection, the two resulting wavefronts maintaining the original width but having reduced amplitudes. Young's Experiment originally performed nearly two hundred years ago is one of the representative examples of the wavefront-splitting interferometer.

2.2.1.1 Young's Experiment

The original experiment performed by Young is shown schematically in Figure 2.13. Sunlight was first allowed to pass through a pinhole S and then, at a considerable distance away, through two pinholes S_1 and S_2. The two sets of spherical waves emerging from the two holes interfered with each other in such a way as to form a symmetrical pattern of varying intensity on the screen AC. If the circular lines represent crests of waves, the intersections of any two lines represent the arrival at those points of two waves with the same phase or with phases differing by a multiple of 2π. Such points are therefore those of maximum disturbance or brightness. A close examination of the light on the screen will reveal evenly spaced light and dark bands or fringes, similar to those as shown in Figure 2.14.

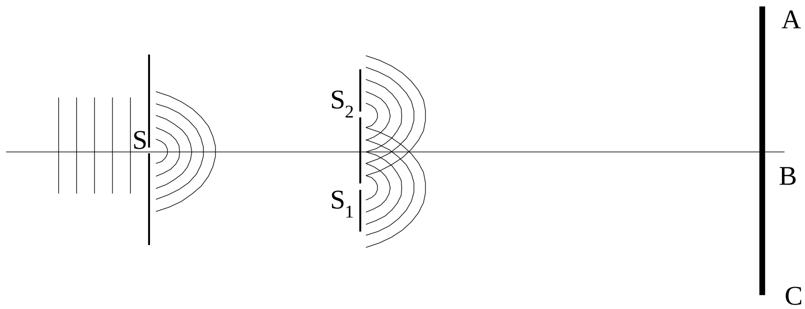


Figure 2.13 Experimental arrangement for Young,s double slit experimen

Figure 2.14 Interference pattern.

The equation (2.46a) gives the condition for maximum irradiance as

$$(r_1 - r_2) = m\lambda \qquad (2.46a)$$

Since the wavelength λ for light is very small, a large number of surfaces corresponding to the lower values of m will exist close to, and on either side of, the plane m = 0. A number of fairly straight parallel fringes will therefore appear on the screen in the vicinity of m = 0, and for this case the approximation $r_1 \sim r_2$ will hold.

Consider a hypothetical monochromatic plane wave illuminating a long narrow slit. From that primary slit a cylindrical wave will emerge. Suppose that this wave falls on two parallel, narrow, closely spaced slits, S_1 and S_2 (as shown in Figure 2.15). The segments of the primary wavefront arriving at the two slits will be exactly in phase, and the slits will constitute two coherent secondary sources. We expect that wherever the two waves coming from S_1 and S_2 overlap, interference will occur (provided that the optical path difference is less than the coherence length).

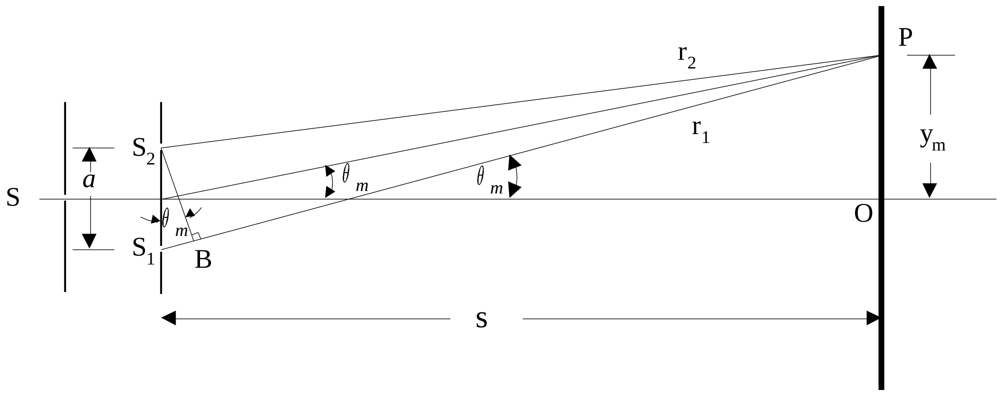


Figure 2.15. The geometry of Young's experiment.

In Figure 2.15 the distance between the two slits and the screen would be very large in comparison with the distance *a* between the two slits (e.g. several thousand times), and all the fringes would be very close to the center O of the screen. The paths difference between the rays along $\overline{S_1P}$ and $\overline{S_2P}$ can be determined by dropping a perpendicular from S_2 onto $\overline{S_1P}$. This path difference is given by

$$\overline{S_1B} = \overline{S_1P} - \overline{S_2P} \qquad (2.47a)$$

or

$$\overline{S_1B} = (r_1 - r_2). \qquad (2.47b)$$

We can express the path difference as

$$(r_1 - r_2) = a.\sin\theta, \qquad (2.48)$$

Since $\theta \sim \sin\theta$ for small angles, we can write

$$\theta = y/s \qquad (2.49)$$

or

$$(r_1 - r_2) = (a/s)y. \qquad (2.50)$$

Combining equations (2.46a) and (2.50) we obtained

$$y_m = (s/a).m\lambda \qquad (2.51)$$

This gives the position of the mth bright fringe on the screen, if we count the maximum at O as the zeroth fringe. The angular position of the fringe is obtained by substituting the equation (2.51) into equation (2.49) we get

$$\theta_m = m\lambda/a. \tag{2.52}$$

The spacing of the fringes on the screen can be obtained from equation (2.51). The difference in the positions of two consecutive maxima is

$$y_{m-1} - y_m = (s/a).(m+1)\lambda - (s/a).m\lambda \tag{2.50a}$$

$$\Delta y = (s/a).\lambda \tag{2.50b}$$

This pattern is equivalent to that obtained for two overlapping spherical waves (in the region $r_1 \sim r_2$). We can apply equation (41) using the phase difference $\delta = k.(r_1 - r_2)$ and get

$$I = 4.I_o.\cos^2(\delta/2) = 4.I_o.\cos^2[k.(r_1 - r_2)/2], \tag{2.51}$$

Provided that the two beams are coherent and have equal amplitudes I_o. Since

$$r_1 - r_2 = (a/s)y \tag{2.52}$$

the resultant irradiance becomes

$$I = 4.I_o.\cos^2\left(\frac{ya\pi}{s\lambda}\right) \tag{2.53}$$

Figure 2.16 shows the idealized intensity pattern, which should be observed at the screen, the consecutive maxima are separated by the Δy given in equation (2.50b). The actual pattern drops off with distance on either side of O because of diffraction.

There are a few more interferometer, which are based on the principle of wavefront-division, e.g., Fresnel's double mirror, Fresnel's biprism, Lloyd's mirror, etc. Details of these interferometers may be found in several textbooks, e.g. Optics by E. Hecht.

2.2.2 Amplitude-Splitting Interferometers

The second category of the interference apparatus is based on the division of amplitude of the wave. There are a good number of amplitude-splitting interferometers that utilize arrangements of mirrors and beam-splitters. The best known and historically the most prominent of these is the Michelson Interferometer.

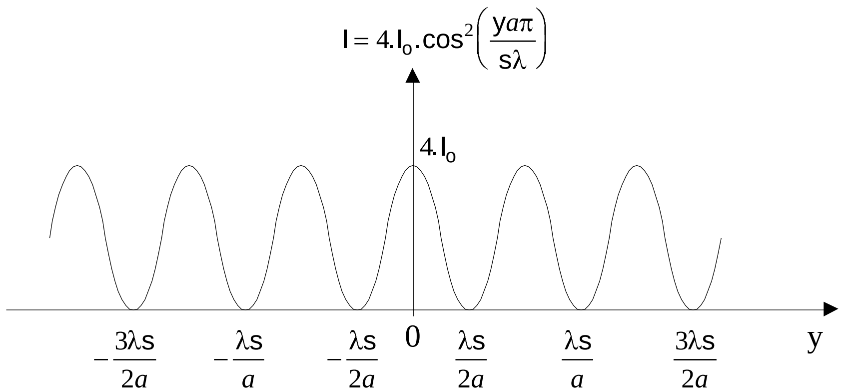


Figure 2.16. Idealized irradiance verses distance curve.

Other amplitude-division interferometers are Mach-Zehnder interferometer, Sagnac Interferometer for rotation measurements, etc. Details of these can be found in Optics by E. Hecht and Fundamentals of Optics by Jenkins and White.

2.2.2.1 Michelson Interferometer

Michelson interferometer is a device, based on the principle of interference of light, which can be used to measure lengths or change in length with great accuracy. We will describe the form originally built by A. A. Michelson in 1881.

The configuration of Michelson interferometer is illustrated in Figure 2.17. An extended light

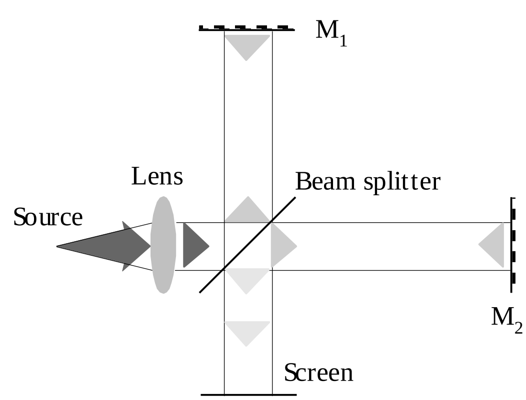


Figure 2.17 Experimental arrangement for Michelson interferometer

sources emits a wave which travels to the right. The beam-splitter at O divides the wave into two, one segment traveling to the right and one upward. The two waves are reflected by mirrors M_1 and M_2 and return to the beam-splitter. Part of the wave coming from M_2 passes through the beam-splitter going downward and part of the wave coming from M_1 is deflected by the beam-splitter downward to the screen. Thus the two waves are united, and interference can be observed.

If mirror M_2 is moved backward or forward, the effect is to change the thickness of the equivalent air film. Suppose that the center of the circular fringe pattern appears bright and that M_2 is moved just enough to cause the first bright circular fringe to move to the center of the pattern. The path of the light beam striking M_2 has changed by one wavelength. This means (because the light passes twice through the equivalent air film) that the mirror must have moved one-half a wavelength.

The interferometer is used to measure changes in length by counting the number of interference fringes that pass the field of view as mirror M_2 is moved. In 1961, an atomic standard of length was adopted by international agreement, which described the standard as the wavelength of the orange-red light of the krypton-86 has replaced the platinum iridium bar as standard of length. Now the meter is defined as a multiple (1,650,763.73) of the wavelength of the light emitted from krypton-86.

2.3 Diffraction

When a beam of light passes through a narrow slit, it spreads out to a certain extent into the region of the geometrical shadow. This is one of the simplest examples of diffraction, i.e., of the failure of light to travel in straight lines. It can be explained only by assuming a wave character of light. In this section we shall investigate quantitatively the diffraction pattern, or distribution of intensity of the light behind the aperture, using the principles of wave motion.

2.3.1 Fresnel and Fraunhofer Diffraction

Consider an opaque shield, A, containing a single small aperture, which is being illuminated by plane waves from a distant point source S, as shown if Figure 2.18. The plane of observation, B, is a screen parallel with, and very close to A. Under these conditions an image of the aperture is projected onto the screen, which is clearly recognizable despite some slight fringing around its periphery. If the plane of observation is moved further away from A, the image of the aperture becomes increasingly more structured as the fringes become more prominent. This phenomenon is known as *Fresnel* or *near-field* diffraction. If the plane of observation is slowly moved out still farther, a continuous change in fringes results. At a very great distance from A, the projected pattern will have spread out considerably, bearing little or no resemblance to the actual aperture. Further moving B essentially changes only the size of the pattern and not its shape. This is *Fraunhofer* or *far-field* diffraction. If the point source was now move toward A, spherical waves falls on the aperture, and a Fresnel pattern would exist, even on a distant plane of observation.

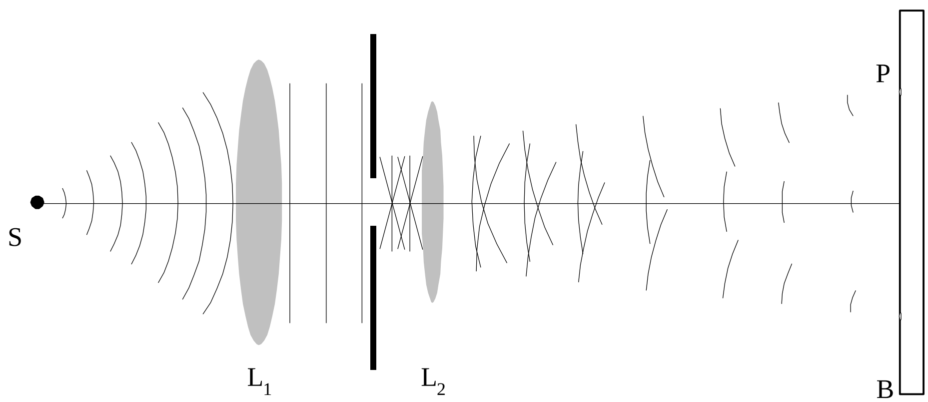


Figure 2.18. Fraunhoffer diffraction

In other words when the source of light and the screen on which the pattern is observed are effectively at infinite distance from the aperture causing the diffraction is called *Fraunhofer* diffraction.

On the other hand, when S or P or both are near to the diffracting aperture A, thus the curvature of the incoming and outgoing waves are not negligible, then *Fresnel* diffraction is dominant.

As a practical rule of thumb, Fraunhofer diffraction will occur at an aperture (or obstacle) of width α when

$$R > \alpha^2/\lambda \qquad (2.57)$$

where R is the smaller of the two distances from S to A and A to P. An increase in the wavelength λ clearly shifts the phenomenon toward the Fraunhofer extreme.

A practical realization of the Fraunhofer condition, where both S and P are effectively at infinity, is achieved by using an arrangement equivalent to that of Figure 2.18. The point source S is located at F_1, the principal focus of lens L_1, and the plane of observation is the second focal plane of L_2. The image at B would be a Fraunhofer diffraction pattern.

Although the Fraunhofer diffraction is a special case of the Fresnel diffraction but because of its simplicity, wide applications and limitations of the length of a Unit we will discuss the Fraunhofer diffraction only.

2.3.2 The Single Slit Fraunhofer Diffraction

A typical arrangement for single slit Fraunhofer diffraction is shown in Figure 2.19. A single slit has a width of several hundred λ and a length of a few centimeters. Diffraction pattern only appears due to small width, the long length, i.e., a few cm, does not play any significant role, therefore, the length of the coherent line source actually corresponds to the width of the slit. The irradiance resulting from an idealized coherent line source (of width b and wavelength λ) in the Fraunhofer approximation is given by[*]

$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2 \qquad (2.58)$$

where $\beta = (\pi b/\lambda)\sin\theta$ and θ is measured from the x-axis in the z direction. Figure 2.20 shows the graph of the intensity pattern. This pattern will be seen to have the form required by the experimental result in Figure 2.21. The maximum intensity of the strong central band comes at the point P_0 (Figure 2.22), where evidently all the secondary wavelets arrive in phase because the path difference is zero. For this point $\beta = 0$, and although the quotient $\sin\beta/\beta$ becomes indeterminate but $\sin\beta$ approaches β for small angles and is equal to it when β vanishes. Hence $\sin\beta/\beta = 1$ for $\beta \to 0$.

[*] See *Optics* by E. Hecht for derivation of the expression.

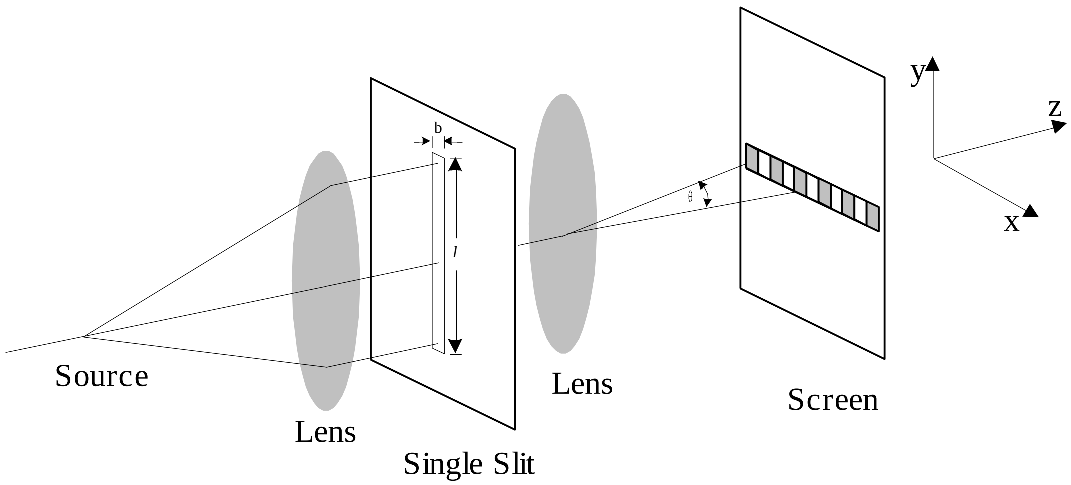


Figure 2.19 Experimental arrangement for obtaining the diffraction pattern of a sing
slit Fraunhofer diffraction.

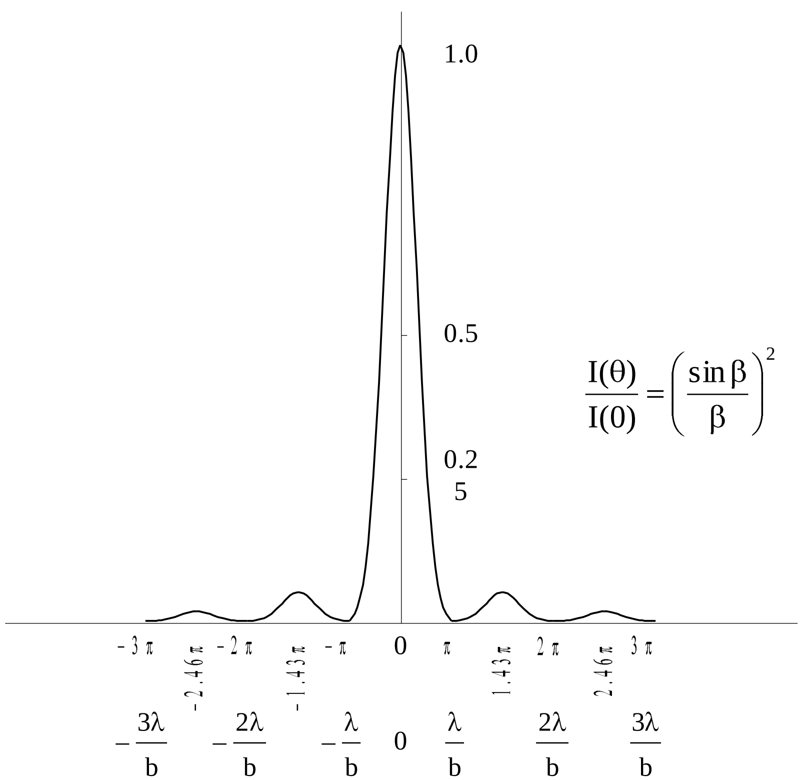


Figure 2.20 Intensity pattern for Fraunhofer diffraction of a single slit show positions of maxima and minima.

In equation (2.58), the line source (slit width b) is short, β is not large, and the irradiance falls off very rapidly but the higher order maxima are observable. The extrema of $I(\theta)$ occur at values of β that causes $dI/d\beta$ to be zero, that is,

$$\frac{dI}{d\beta} = I(0).\frac{2.\sin\beta.(\beta\cos\beta - \sin\beta)}{\beta^3} = 0. \qquad (2.59)$$

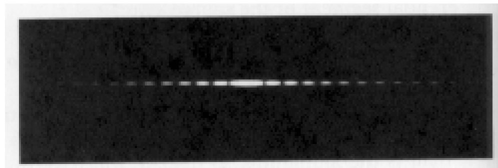


Figure 2.21 Single slit diffraction pattern.

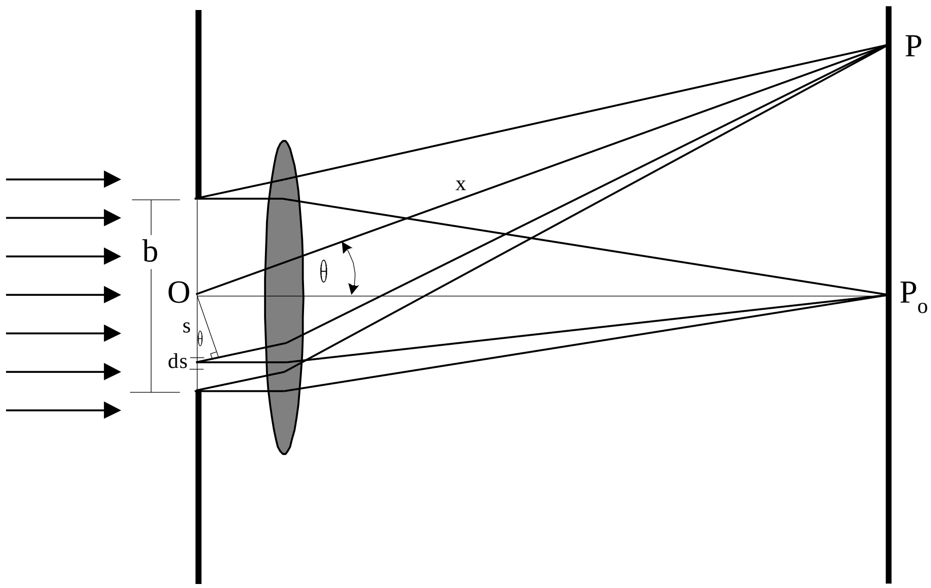


Figure 2.22 Geometrical construction for investigating the intensity in the sin slit diffraction pattern.

After the principal maxima at $\beta = 0$, the irradiance has minima, equal to zero, when

$$\sin\beta = 0, \text{ for } \beta = \pm\pi,\ \pm 2\pi,\ \pm 3\pi,\ \ldots \qquad (2.60)$$

The secondary maxima do not fall halfway between these points, but are displaced towards the center of the pattern. This can be obtained from the above equation (2.59) as

$$\beta.\cos\beta - \sin\beta = 0 \qquad \Rightarrow \tan\beta = \beta. \qquad (2.61)$$

The values of β satisfying this relation can be found graphically as the intersection of the curve $f_1(\beta) = \tan\beta$ and the straight-line $f_2(\beta) = \beta$. Only one maxima exists between adjacent minima, so that $I(\theta)$ has subsidiary maxima at the values of β ($\pm 1.43\pi$, $\pm 2.46\pi$, $\pm 3.47\pi$, ..).

For physical understanding of the phenomenon of the diffraction through single slit consider the light from the slit of Figure 2.23 coming to the point P_1 on the screen. The point P_1 is just

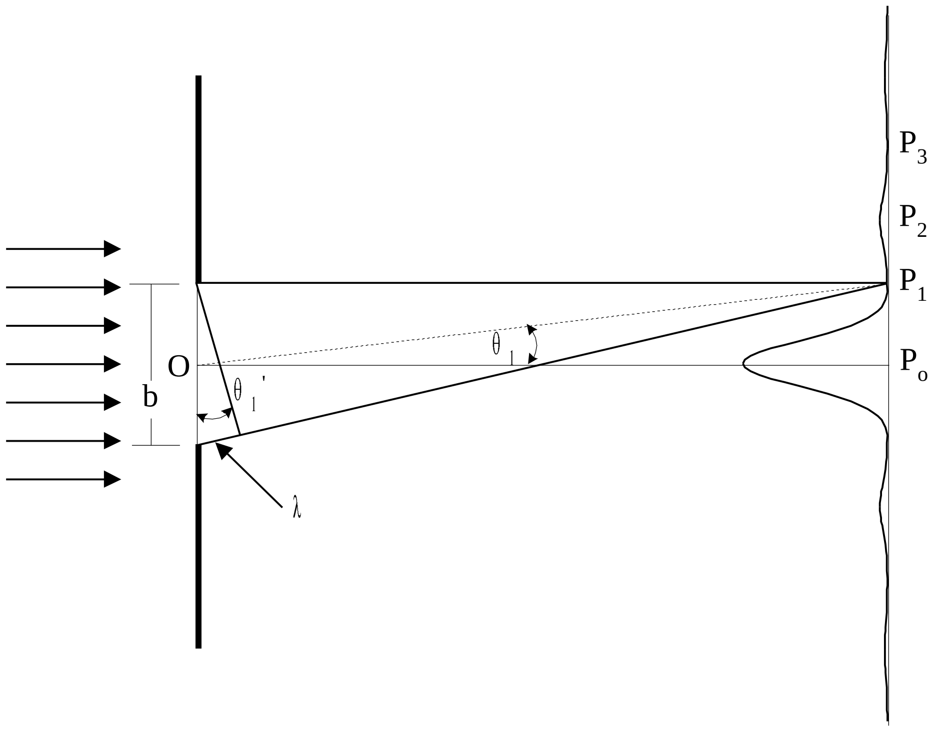


Figure 2.23 Angle of the first minimum of the single-slit diffraction patter

one wavelength farther from the upper edge of the slit than the lower edge. The secondary wavelet from the point in the slit adjacent to the upper edge will travel approximately $\lambda/2$ more than that from the point at the center, and so these two will produce vibrations with a phase difference of π and will give a resultant displacement of zero at P_1. Similarly the wavelet from the next point below the upper edge will cancel that from the next point below the center, and we can continue this pairing off to include all points in the wavefronts, so that the resultant effect at P_1 is zero. At P_3 the path difference is 2λ, and if we divide the slit into four parts, the pairing of points again gives zero resultant, since the parts cancel in pairs. For the point P_2, the path difference is $3\lambda/2$, and we divide the slit into three, two of which will cancel, leaving on third to account for the intensity at this point. The resultant amplitude at P_2 is for less than one-third that at P_0 due to some other reasons.

The above treatment is not very good if the screen is at a finite distance from the slit. As Figure 2.23 is drawn, the dotted line is drawn to cut off equal distances on the rays to P_1. It will be seen from this that the path difference to P_1 between the light coming from the upper edge and that from the center is slightly greater than $\lambda/2$ and that between the center and the lower edge slightly less than $\lambda/2$. Hence the resultant intensity will not be zero at P_1 and P_3, but it will be more nearly to it for greater the distance between slit and screen or for narrower slits. This corresponds to the transition from *Fresnel* diffraction to *Fraunhofer* diffraction. When the screen is at infinity, the relations become simpler. The two angles $\theta_1^{'}$ and θ_1 in Figure 2.23 become exactly equal, i.e., the two dotted lines are perpendicular to each other, and $\lambda = b.\sin\theta_1$ for the first minimum corresponding to $\beta = \pi$. In practice θ_1 is usually very small angle, so we may put the sine equal to the angle, then

$$\theta_1 = \lambda/b \tag{2.62}$$

The width of the pattern increases in proportion to the wavelength, so that for red light it is roughly twice as wide as for violet light. The angular width of the pattern for a given wavelength is inversely proportional to the slit width b, so that as b is made large, the pattern shrinks rapidly to a smaller scale. When width of the aperture is comparable to a wavelength the diffraction is significant. Sound waves will be diffracted through large angles in passing through an aperture of ordinary size, such as an open window.

2.3.3 The Double Slit Fraunhofer Diffraction

The interference of light from two narrow slits close together has already been discussed as a simple example of the interference of two beams of light. In double slit diffraction, the slits are assumed to have widths not much greater than a wavelength of light.

The intensity from the double slit diffraction is*

$$I(\theta) = 4I_0.\left(\frac{\sin^2\beta}{\beta^2}\right).\cos^2\alpha \tag{2.63}$$

where β is the same as defined for the single slit, $\alpha = (\pi d/\lambda)\sin\theta$, and d is the spacing between the two slits. The factor $(\sin^2\beta/\beta^2)$ in this equation is just the same for the single slit of width b, in the previous section. The second factor $\cos^2\alpha$ is characteristic of the interference pattern produced by the two beams of equal intensity and the phase difference δ as shown in the discussion of Young's experiment. There the resultant intensity was found to be proportional to $\cos^2(\delta/2)$, so that the expressions correspond if we put $\alpha = \delta/2$. The resultant intensity will be zero when either of the two factors is zero. For the first factor this will occur when $\beta = \pm\pi, \pm 2\pi, \pm 3\pi, \ldots$, and for the second factor when $\alpha = \pm\pi/2, \pm 3\pi/2, \pm 5\pi/2, \ldots$. The two variable β and α are not independent.

In equation (2.63), for $\theta = 0$ direction (i.e., when $\beta = \alpha = 0$), I_o is the flux-density contribution from either slit, and $I(0) = 4I_o$ is the total flux density. The factor of 4 comes from the fact that the amplitude of the electric field is twice what would be at that point with one slit covered. In the same equation, if d = 0, i.e., the two slits combines into one ($\alpha = 0$) and the equation (2.63) becomes $I(\theta) = 4.I_o.(\sin^2\beta/\beta^2)$. This is the equivalent equation for single slit diffraction with the source strength doubled. We might expect the total expression as being generated by a $\cos^2\alpha$ interference term modulated by a $(\sin^2\beta/\beta^2)$ diffraction term. If the slits are finite in width but very narrow, the diffraction pattern from either slit will be uniform over a broad central region and the bands resembling the idealized Young's fringes will appear within that region (Figure 2.24). In fact the intensity distribution in the double slit diffraction pattern is a combination of the interference and diffraction simultaneously, sharp maxima and minima is due to interference and the broader modulation is due to diffraction.

2.3.4 Diffraction Grating

An arrangement that is equivalent in its action to a number of parallel equidistant slits of the same width is called a diffraction grating. In the previous section we have discussed double slit diffraction, which may be considered as an elementary grating of only two slits.

The procedure for obtaining the irradiance function for a monochromatic wave diffracted by many slits is essentially the same as that used when considering two slits. In the case of N long parallel, narrow slits, each of width b and center to center distance d, the irradiance at an θ is given by*

* See again *Optics* by E. Hecht.

$$I(\theta) = I_0 \cdot \left(\frac{\sin\beta}{\beta}\right)^2 \cdot \left(\frac{\sin N\alpha}{\sin\alpha}\right)^2 \tag{2.64}$$

where α and β is the same as defined for the two slit case.

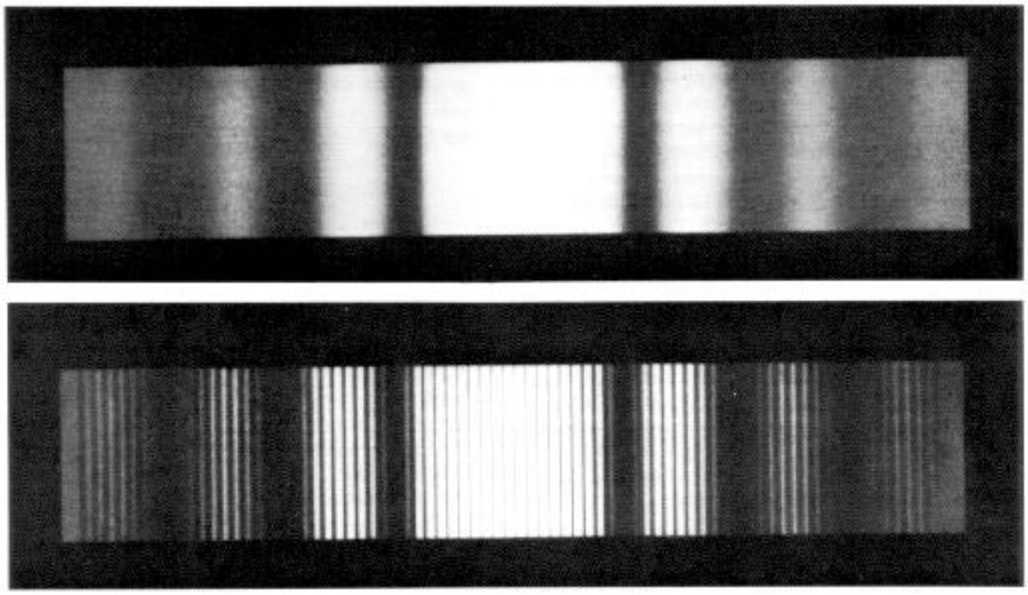


Figure 2.24 Double slit diffraction pattern.

The most striking modification in the pattern as the number of slits is increased consists of narrowing of the interference maxima as shown in Figure 2.25. For two slits these are diffuse, having intensity, which vary essentially as the square of the cosine. With more slits the sharpness of these principal maxima increases rapidly and for large N, they have become narrow lines.

In equation (2.64) the new factor ($\sin^2 N\alpha/\sin^2\alpha$) may be said to represent the interference term for N slits. It possesses maximum values equal to N^2 for $\alpha = 0, \pm\pi, \pm 2\pi, \pm 3\pi, \ldots$. Although the quotient becomes indeterminate at these values, this result can be obtained by noting that

$$\lim_{\alpha\to m\pi} \frac{\sin N\alpha}{\sin\alpha} = \lim_{\alpha\to m\pi} \frac{N.\cos N\alpha}{\cos\alpha} = \pm N \tag{2.65}$$

These maxima correspond in position to those of the double slit. For the above values of α, we have

$$d.\sin\theta = 0,\ \pm\lambda,\ \pm 2\lambda,\ \ldots = \pm m\lambda. \tag{2.66}$$

These maxima are more intense, however, in the ratio of the square of the number of slits. The relative intensities of the different orders, m, are in all cases governed by the single-slit diffraction envelope ($\sin^2\beta/\beta^2$). Hence the relation between β and α in terms of slit width and slit separation remains unchanged.

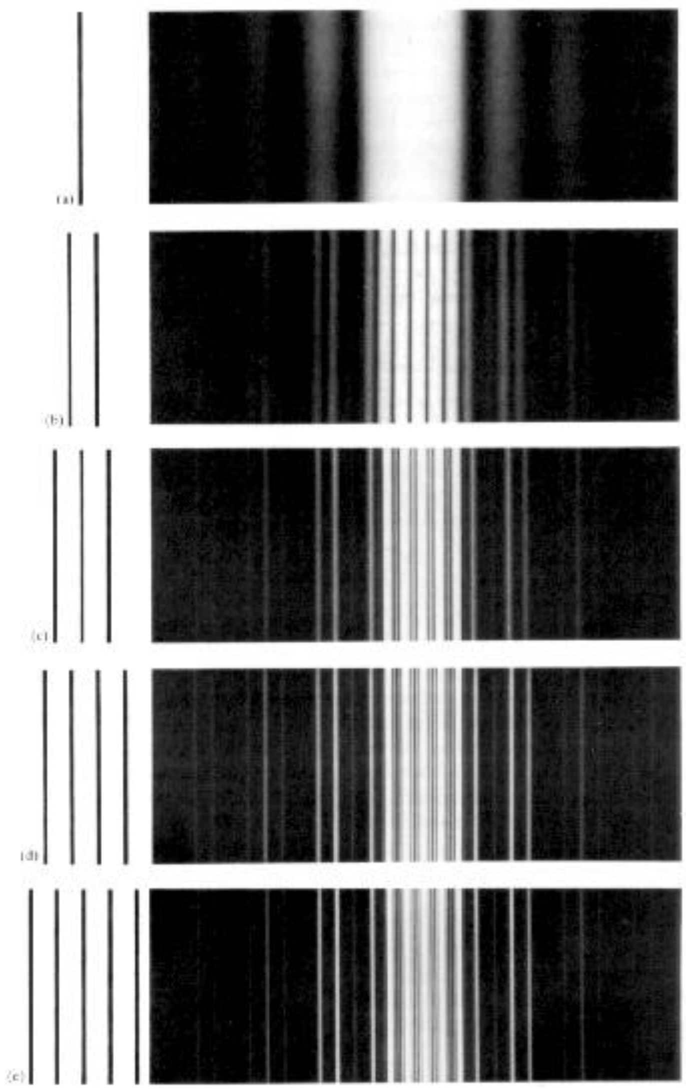


Figure 2.25 Multiple slit diffraction patterns.

Problems

2.1 Describe completely the state of polarization of each of the following waves:

(a) $\mathbf{E} = \mathbf{i}\, E_o.\cos(kz - \omega t) - \mathbf{j}\, E_o.\cos(kz - \omega t)$

(b) $\mathbf{E} = \mathbf{i}\, E_o.\cos(kz - \omega t) + \mathbf{j}\, E_o.\cos(kz - \omega t)$

(c) $\mathbf{E} = \mathbf{i}\, E_o.\cos(\omega t - kz) + \mathbf{j}\, E_o.\cos(\omega t - kz + \pi/2)$

(d) $\mathbf{E} = \mathbf{i}\, E_o.\, \sin(kz - \omega t) - \mathbf{j}\, E_o.\sin(kz - \omega t)$

2.2 Write an expression for a linearly polarized light wave of angular frequency ω and amplitude E_o propagating along the x-axis with its plane of vibration at angle of 30^o to the xy-plane. The disturbance is zero at $t = 0$ and $x = 0$.

2.3 Suppose that an ideal polarizer is rotated at a rate w between a similar pair of rotational crossed polarizers. Show that the emergent flux density will be modulated at four times the rotational frequency. In other words, show that

$$I = (I_1/8)\,(1 - \cos 4\omega t)$$

where I_1 is the flux density emerging from the first polarizer and I is the final flux density.

2.4 Is Young's experiment an interference experiment or a diffraction experiment, or both?

2.5 Design a double slit arrangement that will produce interference fringes 1^o apart on a distant screen. Assuming wavelength of light as 632.8 nm.

2.6 What changes occur in the pattern of interference fringes if the apparatus of the Young's experiment is placed under water?

2.7 In double slit experiment the distance between slits is 4.0 mm and the slits are 1 meter from the screen. Two interference patterns can be seen on the screen, one due to light of 480 nm and the other 632.8 nm. What is the separation on the screen between the third-order interference fringes of the two different patterns?

2.8 In a double slit arrangement the slits are separated by a distance equal to 100 times the wavelength of the light passing through the slits. (a) What is the angular separation between the first and second maxima? (b) What is the linear distance between the first and second maxima if the screen is at a distance of 1 meter from the slits?

2.9 A Michelson interferometer is illuminated with monochromatic light. One of its mirrors is then moved, and 1500 fringe-pairs shift past the hairline in a viewing telescope during the process. If the device is illuminated with 632.8 nm light, how far was the mirror moved.

2.10 A collimated beam of microwaves impinges on a screen that contains a long horizontal slit that is 25 cm wide. A detector moving parallel to the screen in the far field regions locates the first minimum of irradiance at an angle of 30o above the central axis. Determine the wavelength of the radiation.

Books for further reading

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed. (McGraw-Hill, New York, 1985)

E. Hecht, *Optics*, 2nd ed. (Addison-Wesley, Reading, Mass., 1990).

R. Guenther, *Modern Optics*, (John Wiley & Sons, New York, 1990).

M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1986).

Unit-3

Geometrical Optics

Objective

Geometrical optics is being studied for the last several centuries and still its application can be found in most of the experimental research, especially in lasers and its applications. To study the laser resonator, one needs the ray tracing. Light propagation through optical fibers is commonly used and has wide applications. This unit deals with the basic laws of the geometrical optics, matrix formulation for ray tracing and optical fibers.

3.1 Introduction

Geometrical optics is the study of the propagation of light which passes through systems with dimensions, which are large compared with the wavelength of light. The propagation of light can be described with a few simple geometrical relationships, which are the four basic laws of geometrical optics. Since the observations of these phenomena require no special scientific apparatus, therefore, these were studied a long time ago and geometrical optics is considered as the oldest branch of optics.

By the end of the twentieth century, a large number of applications of the geometrical optics have been found. Laser resonators and optical fibers are among the new applications. After the basic laws of the geometrical optics we will discuss the matrix formulation of optical components and finally the optical fibers will be described.

3.2 Laws of Geometrical Optics

The first law of geometrical optics is the law of rectilinear propagation. This is a scientific way of saying that in an isotropic, homogeneous medium, light travels in straight lines. An isotropic homogeneous medium is one whose physical properties are constant and the same in every direction. Examples of such media are vacuum, optical air, and calm air. A modern demonstration of rectilinear propagation is to shine a laser beam through chalk dust suspended in air. The laser beam clearly travels in straight line.

Another law of geometrical optics is again very trivial and known as law of reversibility. This law states that if the direction of a ray is reversed, it will trace exactly the same path backwards. The law of rectilinear propagation also agrees with the law of reversibility. After all, a straight line is a straight line in both directions.

We will discuss the laws of reflection and refraction and its present form, i.e., Fresnel's laws in a little more detailed.

3.2.1 Laws of Reflection and Refraction

When light encounters a surface between two media, its direction of propagation will be changed in two ways. The light will bounce off the surface, which we call reflection; and the light will be transmitted, which we call refraction (Figure 3.1). In the Figure θ_i, θ_r, and θ_t are

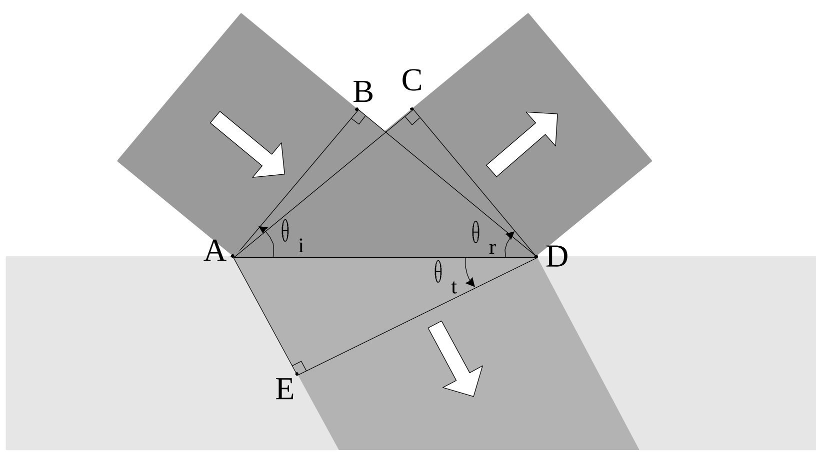


Figure 3.1 Reflected and transmitted wavefronts at a given insta

the angles of incidence, reflection, and refraction (or transmission), respectively. We can write from the figure that

$$\sin\theta_i = \frac{\overline{BD}}{\overline{AD}}, \tag{3.1a}$$

$$\sin\theta_r = \frac{\overline{AC}}{\overline{AD}}, \tag{3.1b}$$

$$\sin\theta_t = \frac{\overline{AE}}{\overline{AD}}. \tag{3.1c}$$

From the above three equations we have

$$\frac{\sin\theta_i}{\overline{BD}} = \frac{\sin\theta_r}{\overline{AC}} = \frac{\sin\theta_t}{\overline{AE}} = \frac{1}{\overline{AD}}. \tag{3.2}$$

If speed of the incident and reflected light is v_i, and for the transmitted light is v_t in the medium, we can write

$$\overline{BD} = v_i.t, \qquad \overline{AC} = v_i.t, \qquad \text{and} \qquad \overline{AE} = v_t.t, \tag{3.3}$$

Substituting these values in the equation (2) and cancelling t, we have

$$\frac{\sin\theta_i}{v_i} = \frac{\sin\theta_r}{v_i} = \frac{\sin\theta_t}{v_t}. \tag{3.4}$$

It follows from the first two terms that the angle of incidence equals the angle of reflection, that is,

$$\theta_i = \theta_r. \tag{3.5}$$

This is mathematical formulation of the *law of reflection*. Another important condition regarding the law is that the incident ray, reflected ray, and the normal to the interface lie on the same plane.

The first and last term of equation (3.4) yield

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{v_i}{v_t}, \tag{3.6}$$

since $v_i/v_t = n_t/n_i$ (n_t and n_i being the refractive indices of the two mediums)

$$n_i.\sin\theta_i = n_t.\sin\theta_t \tag{3.7}$$

This is very important *law of refraction*, the physical consequences of which has been studied some eighteen hundred years ago.

3.2.1.1 Fresnel Reflections

Ordinary glass is a good example of a dielectric, when light incident on it normally, a small portion of it is reflected back and a major portion is transmitted through it. At other angles of incidence the reflecting intensity increases with angle until at 90^o, that is, grazing incidence, at which all the light is reflected.

It was observed that the light reflected from the dielectric medium is partially plane-polarized. The incident unpolarized light can be considered as having two plane-polarized components, the vibrations of which are parallel and perpendicular to the plane of incidence, respectively. In the laboratory, this can be performed by examining the reflected light that passes through a polarizer (Figure 3.2). If the polarizer is oriented with its principal axis

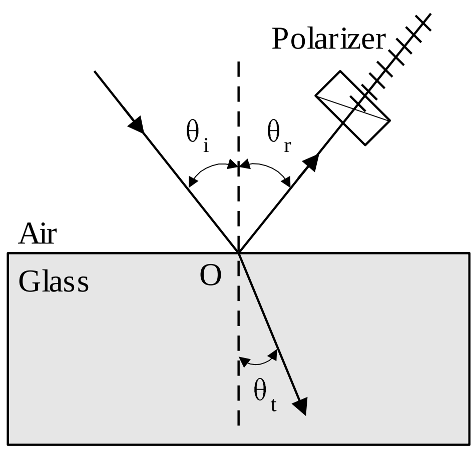


Figure 3.2 Analysis of the reflected light into its two plane-polarized components.

parallel to the plane of incidence, the p-vibrations, i.e., the vibrations parallel to the plane of incidence can be measured. Rotation of the polarizer through 90^o then allows the s vibrations perpendicular to the plane of incidence to be measured. The reflected intensities in the p and s polarized light is plotted in Figure 3.3 as two solid curves against the angle of incidence.

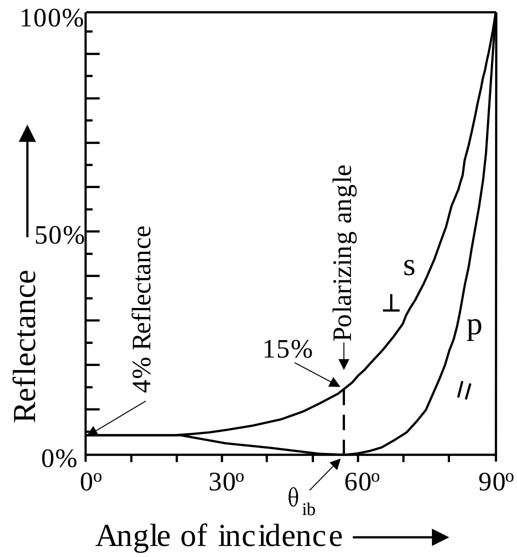


Figure 3.3 Reflected intensities for a dielectric having n =1.50.

The curves of Figure 3.3 are represented very accurately by theoretical equations, which were first derived by Fresnel and are known as Fresnel laws of reflection. The law may be written as

$$r_s = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)}, \tag{3.8a}$$

$$r_p = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)}, \tag{3.8b}$$

$$t_s = \frac{2.\sin\theta_t.\cos\theta_i}{\sin(\theta_i + \theta_t)}, \tag{3.8c}$$

$$t_p = \frac{2.\sin\theta_t.\cos\theta_i}{\sin(\theta_i + \theta_t).\cos(\theta_i - \theta_t)}, \tag{3.8d}$$

where r_s, and r_p are amplitude reflection coefficients, and t_s and t_p are amplitude transmission coefficients of s and p polarized light, respectively. The reflectance are given by r_s^2 and r_p^2 and these are the curve in the Figure 3.3. At normal incidence the parallel and perpendicular components must be equally reflected because here the plane of incidence is undefined and the two components are not distinguishable. With increasing θ_i, r_p^2 drops and r_s^2 rises until at the polarizing angle their values are zero and 15%, respectively. At grazing incidence both components are totally reflected. The value of the reflectance at normal incidence can be easily determined from the equation (3.8a and 3.8b). For small θ_i, we can set sine and tan are equal, as

$$r_p = -r_s = r = \frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \tag{3.9a}$$

$$r = \frac{\sin\theta_i \cos\theta_t - \cos\theta_i . \sin\theta_t}{\sin\theta_i \cos\theta_t + \cos\theta_i . \sin\theta_t} \tag{3.9b}$$

Dividing numerator and denominator of equation (3.9b) by $\sin\theta_t$ and replacing $\sin\theta_i/\sin\theta_t$ by n, we find that it reduces to

$$r = \frac{n.\cos\theta_t - \cos\theta_i}{n.\cos\theta_t + \cos\theta_i} \approx \frac{n-1}{n+1} \tag{3.10}$$

The approximate equality becomes exact in the limit when angles become zero. Hence reflectance at the normal incidence is

$$r^2 = \left(\frac{n-1}{n+1}\right)^2 \tag{3.11}$$

This very useful equation gives the reflectance at $\theta_i = 0$ for any single clean surface of a dielectric. Thus a glass having n = 1.50 has $r^2 = 0.04$, or exactly 4 percent as indicated in Figure 3.3.

3.2.1.2 Internal Reflection

In the previous discussion it was assumed that the light strikes the boundary from the side of the rare medium to denser medium, so that we are dealing with the external reflection. Fresnel’s laws apply equally well to the case of denser to rare medium, or internal reflection. If the same value of n is to be assumed for the dense medium, then in the previous discussions only involves the exchange of θ_i and θ_t in the equations. The resulting curves of reflectance are plotted in Figure 3.4. The reflectance varies in the similar fashion as of external reflection up to the critical angle θ_c (= 41^o for ordinary glass). The reflections of s and p components start at 4 percent on the normal incidence, and diverge until the polarizing

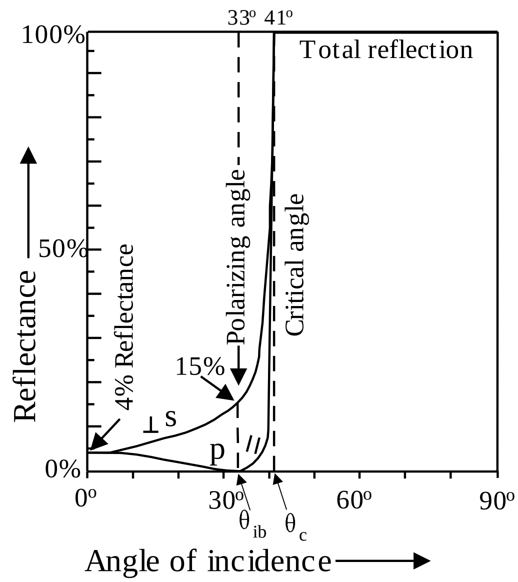


Figure 3.4 Intensity for internal reflection at a dielectric boundary, n = 1.5

angle is reached. At the critical angle the refracted ray comes out at grazing angle, and internal reflection becomes 100% just as for external reflection at grazing incidence. When θ exceeds the critical angle, Fresnel equations contains imaginary quantities, but reflection remains total.

3.2.1.3 Brewester's Law

When an unpolarized light falls on a dielectric medium, the reflected and refracted beams are partially polarized. By changing the incidence angle the polarization may increase or decrease and at certain incidence angle the reflected and refracted beams are completely polarized but their polarizations are orthogonal to each other. To get clearer picture, consider unpolarized light to be incident at an angle θ_i on a dielectric medium like glass, as shown in Figure 3.5. There will always be a reflected ray OR and a refracted ray OT. Experimental observation shows that the reflected ray OR is partially plane-polarized and that only at certain definite angle, about 57^o for ordinary glass, it is plane-polarized. It was Brewester who first discovered that at this polarizing angle 'θ_{ib}' the reflected and refracted rays are just 90^o apart. This remarkable discovery enables one to correlate polarization with the refractive index. According to the law of refraction

$$\frac{\sin\theta_i}{\sin\theta_t} = n \tag{3.12}$$

where θ_i and θ_t are angles of reflection and refraction, respectively. Since at Brewester angle θ_{ib}, the angle ROT = 90^o, we have $\sin(\theta_{tb}) = \cos(\theta_{ib})$, giving

$$\frac{\sin\theta_{ib}}{\sin\theta_{tb}} = \frac{\sin\theta_{ib}}{\cos\theta_{ib}} = n$$

$$\Rightarrow \quad \tan(\theta_{ib}) = n \tag{3.13}$$

This is Brewester's law, which shows that the angle of incidence for maximum polarization depends only on the refractive index. It therefore varies a little with wavelength, but for ordinary glass the dispersion is such that the polarizing angle (θ_{ib}) does not change much over the whole visible spectrum.

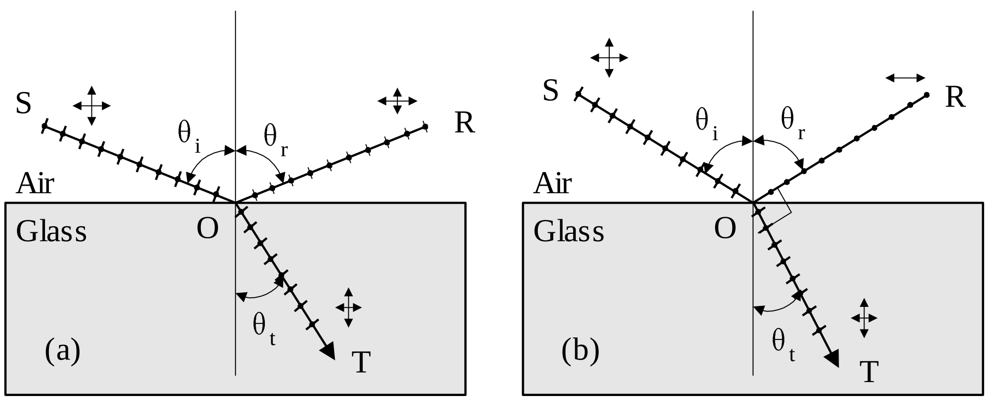


Figure 3.5 (a) Polarization by reflection and refraction. (b) Brewester's law for the polarizing angle.

The phenomenon of polarization at Brewester's angle can be understood as follows. The incident beam sets the electrons in the atoms of the material into oscillation, and it is the re-radiation from these that generates the reflected beam. When the latter is observed at 90^o to the refracted beam, only the vibrations that are perpendicular to the plane of incidence can contribute. Those in the plane of incidence have no component traverse to the 90^o direction and hence can not radiate in that direction. The reason is the same as that which causes the radiation from a horizontal antenna to drop to zero along the direction of the wires.

3.3 Optical Components

The instruments/components used to control the path of the optical radiation may be termed as optical components. This section we will start with the simplest component, i.e., mirrors.

3.3.1 Mirrors

A mirror is an optical component, which reflects most of the radiation falls on it. In general the mirrors are formed by metallic coating on some substances, however, dielectric medium are also used for reflection. Mirrors can be subdivide according to the shape or curvature.

3.3.1.1 Plane Mirrors

Plane mirrors perhaps the simplest and extremely common image forming system. The nature, size, and location of plane mirror images can be determined by applying the law of reflection to a number of rays leaving an object point. Another interesting aspect of the images formed by the mirror is the inversion. In fact, the meaning of mirror image in everyday language is that the right and left hands are reversed.

3.3.1.2 Spherical Mirrors

These mirrors are formed from a small portion of a sphere. A spherical mirror in which reflecting surface bows outward is convex and the one that bows inward is called concave. An important characteristic of these mirrors is that they have axial symmetry, this implies that the centres of all spherical surfaces in the system lie along a straight line. This line is called the optical axis.

If we draw rays reflecting off a spherical mirror, we should realise that the rays do not all crosses at one image point (Figure 3.6) and a sharp image will not be formed. In fact, we need to restrict the rays. If we deal only with rays that make small angles with the optical axis and never get very far from the optical axis, then the rays take paths through the optical systems, which are similar to the path taken by the optical axis. These rays are called paraxial rays. Dealing only with paraxial rays means that all rays coming from a particular object point will meet again at an image point. In practical laser resonators (will be discussed in unit-6) this is a good assumption, because only those rays keep on oscillating which lies close to the optical axis.

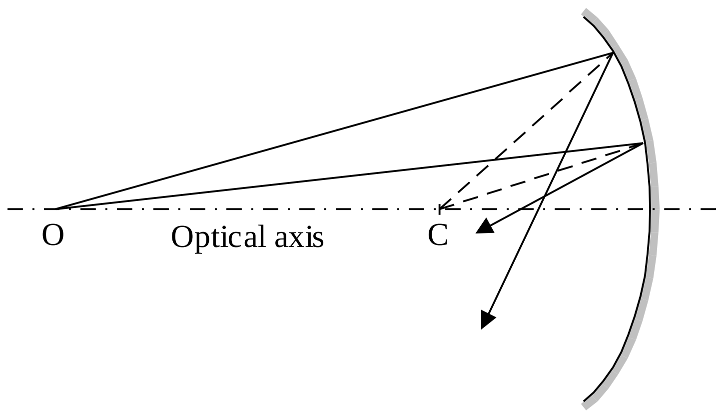


Figure 3.6 Nonparaxial rays reflecting off a spherical mirror.

3.3.2 Lenses

A lens is an optical system consisting of two or more refracting interfaces at least one of which is curved. A lens that consists of one element (i.e., it has only two refracting surfaces) is a simple lens. The presence of more than one element make it a compound lens. A lens is also classified as to whether its thickness is effectively negligible or not. We will limit ourselves for centred systems (for which all surfaces are rotationally symmetric about a common axis) of spherical surfaces. Lenses that are commonly known as *convex*, *converging*, or *positive* are thicker at the centre and so tend to decrease the radius of curvature of wavefronts. In other words, the wave converges more as it traverses the lens (assuming that the index of refraction of the lens is greater than that of the media in which it is immersed). *Concave*, *diverging*, or *negative* lenses, on the other hand, are thinner at the centre and tend to advance that portion of the wavefronts, causing it to diverge more then it did upon entry.

The image of an object can be formed from a thin lens by using the famous lens formula as;

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q} \tag{3.14}$$

where f is the focal length of the lens, p is the distance of the object from the lens, and q is the distance of the image from the lens. The value of f is +ve for convex lens and –ve for concave lens, and p has positive sign for real image and –ve for virtual image.

The same relation for image formation holds for imaging from spherical mirrors but focal length f is -ve for convex mirror and +ve for concave mirror (opposite to the lens system).

3.4 Ray Tracing

Ray tracing is one of the important tools of an optical designer. After completing an optical system on paper, one can shine rays through it (mathematically) to evaluate its performance. Any ray paraxial or otherwise, can be traced through the system exactly. Conceptually, it is a simple matter of applying the law of refraction at the first interface, locating where the transmitted ray then strikes the second surface, applying the law once again, and so on all the way through. This ray tracing can be analytically done but life can be a lot easier by using matrix method, introduced by T. Smith in 1930s.

3.4.1 Ray Matrix Formulation

In ray tracing by matrix method, each optical component has different matrix. For example, when a ray encounters reflection and refraction, one can define a reflection matrix and refraction matrix. By combining matrices that represent individual reflection and refraction in an optical system, one obtains a resultant matrix from the essential properties of the composite system. All the matrices will be derived under paraxial approximation. Paraxial rays are those rays, which make very small angles with the axis and lies close to the axis throughout in the optical system.

3.4.1.1 Matrix of a Component

Now consider a ray of light that is either transmitted by or reflected from an optical element e.g., a lens or a mirror. If the ray is travelling approximately along the z direction, then the ray-vector $\mathbf{r_1}$ at a given input plane $z = z_1$ of the optical element (Figure 3.7) can be characterised by two parameters, namely, its radial displacement $r_1(z_1)$ from the z-axis at the given plane and its angular displacement θ_1. Similarly the ray-vector $\mathbf{r_2}$ at a given output

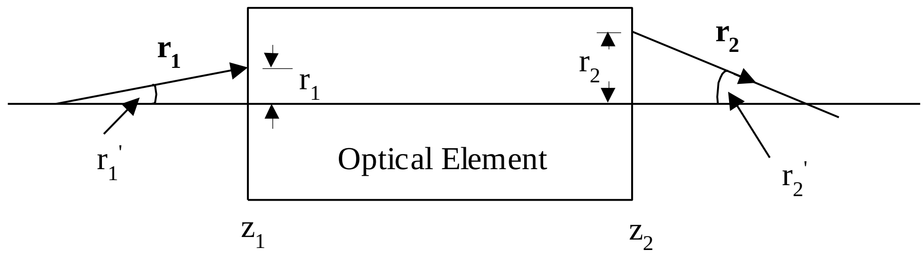


Figure 3.7 Matrix formulation of the optical components.

plane z = z_2 can be defined by its radial, $r_2(z_2)$, and angular, θ_2, displacements. Within the paraxial-ray approximation the angular displacements θ are assumed to be very small such that $\sin\theta \approx \tan\theta \approx \theta$ is a valid assumption. In this case the final position, (r_2, θ_2) and initial position, (r_1, θ_1), are related by a linear transformation. If we put $\theta_1 \cong \left(\frac{dr_1}{dz}\right)_{z_1} = r_1^{'}$ and $\theta_2 \cong \left(\frac{dr_2}{dz}\right)_{z_2} = r_2^{'}$ we can write

$$r_2 = Ar_1 + Br_1^{'} \tag{3.15a}$$

and

$$r_2^{'} = Cr_1 + Dr_1^{'} \tag{3.15b}$$

where A, B, C, and D are constants characteristic of the given optical element. In a matrix formulation it is therefore convenient to write the above equations as

$$\begin{bmatrix} r_2 \\ r_2^{'} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_1^{'} \end{bmatrix} \tag{3.16}$$

where ABCD matrix completely defines the given optical element within the paraxial ray approximation.

3.4.1.2 Translation Matrix

As a first and simplest case we will consider the free-space propagation of a ray along a length Δz = L of a given material with refractive index n (Figure 3.8). If the input and output planes lie just outside the medium, in a medium of refractive index equal to unit, we have

$$r_2 = r_1 + \frac{L \cdot r_1'}{n} \tag{3.17a}$$

and

$$r_2' = r_1' \tag{3.17b}$$

and the corresponding ABCD matrix is

$$\begin{bmatrix} 1 & L/n \\ 0 & 1 \end{bmatrix} \tag{3.18}$$

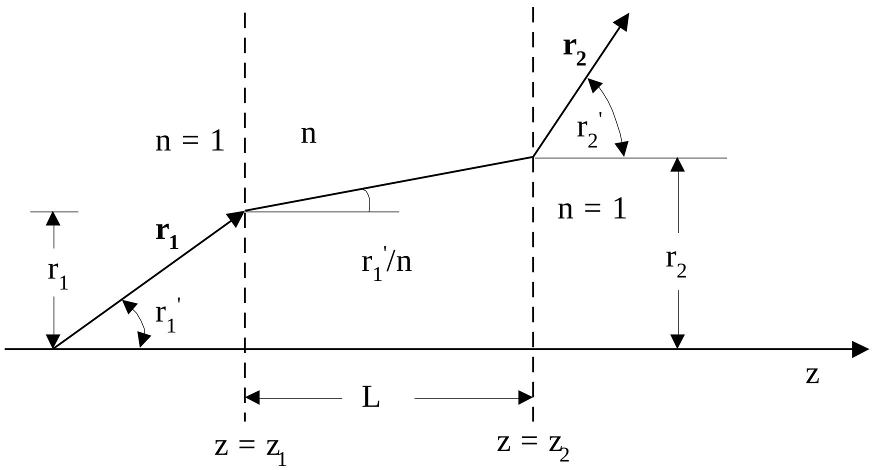


Figure 3.8 Calculation of the ABCD matrix for free space propagation

3.4.1.3 Matrix of a Lens

As a next case we consider ray propagation through a lens of focal length f (f is taken to be positive for a converging lens). For a thin lens, shown in Figure 3.9, we have

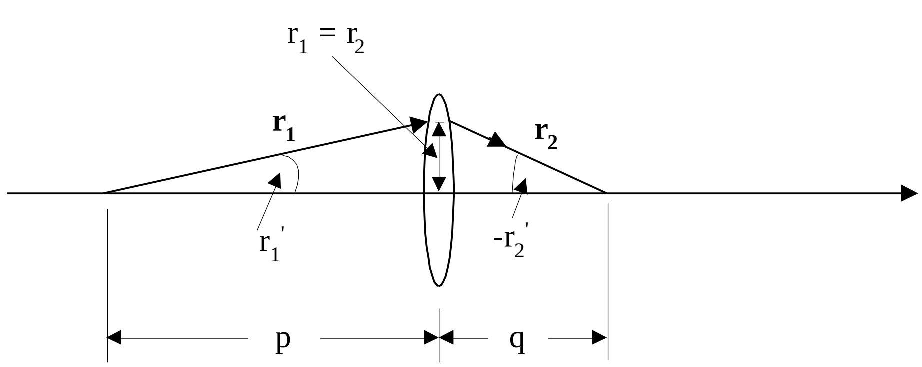


Figure 3.9 Derivation of the matrix for ray propagation through thin lenses

$$r_2 = r_1 \tag{3.19a}$$

The second relation is obtained from the well-known law of geometrical optics, i.e., the lens formula, $\frac{1}{p}+\frac{1}{q}=\frac{1}{f}$, and using the fact that $p=\frac{r_1}{r_1'}$ and $q=-\frac{r_2}{r_2'}$. We get

$$r_2' = -(1/f)\,r_1 + r_1' \tag{3.19b}$$

According to the equation (3.19a) and (3.19b) for a thin lens, the ABCD matrix is

$$\begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}. \tag{3.20}$$

3.4.1.4 Matrix of a Spherical Mirror

Here we consider reflection of a ray by a spherical mirror of radius of curvature R (R is taken to be positive for a concave mirror) as shown in Figure 3.10. In this case the z_1 and z_2 planes

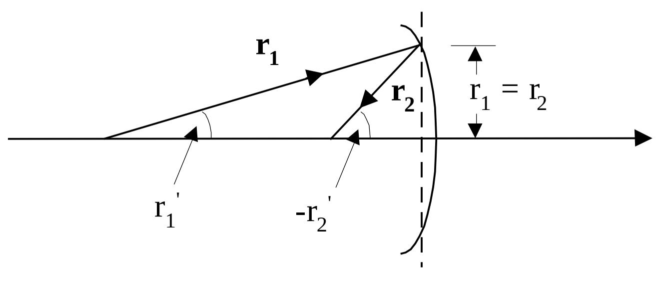


Figure 3.10 Ray matrix of reflection from a spherical mirro

are taken to be coincident and to be placed just in front of the mirror. The positive direction of the z-axis is taken to be that from left to right for the incident vector and from right to left for the reflected vector. Given these conventions the ray matrix of a concave spherical mirror of radius of curvature R and, hence, focal length f = R/2 becomes identical to that of a positive lens of focal length f. The ray matrix is therefore equal to

$$\begin{bmatrix} 1 & 0 \\ -2/R & 1 \end{bmatrix}. \tag{3.21}$$

An important property that holds for all the three cases that we have considered is that the determinant of the ABCD matrices is always unitary (i.e., AD - BC = 1) if the index of refraction at the input and output planes is the same.

3.4.1.5 Matrix of a Composite System

Once the matrices of the elementary optical elements are known one can readily obtain the overall matrix of a more complex optical element by subdividing it into these elementary matrices. Suppose that within a given optical element, there is an intermediate plane of co-ordinate z_i, as shown in Figure 3.11. The two elementary ABCD matrices between planes $z = z_1$ and $z = z_i$ and planes $z = z_i$ and $z = z_2$ are known. If we call r_i and r_i' the co-ordinates of the ray vector at plane $z = z_i$, we can write

$$\begin{bmatrix} r_i \\ r_i' \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} \tag{3.22a}$$

and

$$\begin{bmatrix} r_2 \\ r_2' \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \cdot \begin{bmatrix} r_i \\ r_i' \end{bmatrix} \tag{3.22b}$$

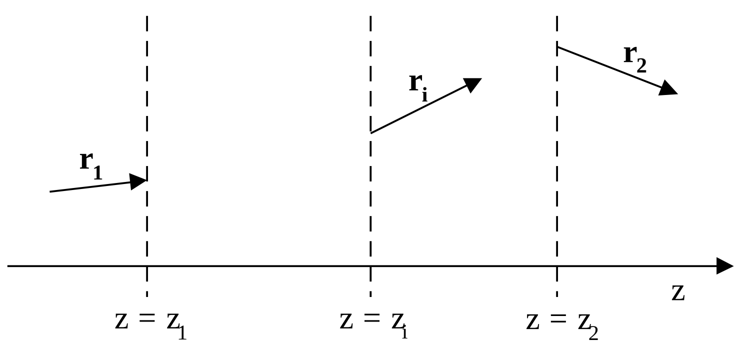


Figure 3.11 Ray propagation through distinct planes when two matrices betw planes $z = z_1$ and $z = z_i$ and between planes $z = z_i$ and $z = z_2$ are known.

By substituting $\mathbf{r_i}$ from the equation (3.22a) into the right hand side of the equation (3.22b), we have

$$\begin{bmatrix} r_2 \\ r_2' \end{bmatrix} = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \cdot \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} \tag{3.23}$$

The overall ABCD matrix can thus be obtained by the multiplication of the ABCD matrices of the elementary components. Note that the order in which the matrices appear in the product is the opposite of the order in which the corresponding optical elements are traversed by the light ray.

As a trivial example, consider the free space propagation through a length L_1 followed again by a free space propagation through a length L_2, in a medium of refractive index n. According to equation (3.23) the overall matrix equation can be written as

$$\begin{bmatrix} r_2 \\ r_2' \end{bmatrix} = \begin{bmatrix} 1 & L_2/n \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & L_1/n \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} \tag{3.24}$$

The product of the two matrices of the equation (3.24) can be shown as

$$\begin{bmatrix} 1 & (L_1 + L_2)/n \\ 0 & 1 \end{bmatrix} \tag{3.25}$$

This calculation confirms the result that overall propagation is equivalent to a free-space propagation over a total length $L = L_1 + L_2$.

In another example consider free propagation over a length L (in a medium of refractive index n=1) followed by reflection from a mirror of radius of curvature R. According to equations (3.18), (3.21) and (3.23) the overall ABCD matrix is given by

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -2/R & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & L \\ -(2/R) & 1-(2L/R) \end{bmatrix} \end{aligned} \tag{3.26}$$

From the above discussion, it is obvious that this matrix is also unitary, i.e., AD – BC = 1, and this formulation can be very helpful for describing the geometrical-optics behaviour of an optical resonator and it holds for any arbitrary cascade of optical elements.

From the previous examples and discussions, we found that the reflection of a ray from a curved mirror of radius R has the same effect as passage of a ray through a thin lens of focal length f = R/2, provided that slope of the ray is always defined with respect to its forward direction of travel. This is important in establishing the connection between optical resonators and lens waveguides.

3.5 Optical Fiber

In the previous section we have seen the law of refraction for denser to rare medium and at certain critical angle of incidence the refracted ray emerges at grazing angle. Mathematically, critical angle can be represented as

$$\theta_c = \arcsin(n_i/n_t) \tag{3.27}$$

where n_i is the refractive index of the medium in which light is reflected and n_t is the refractive index of the medium in which light is refracted at the grazing angle. When angle of incidence increases from this critical value, the total internal reflection takes place. These simple principles of refraction and reflection form the basis of light propagation through an optical fiber.

Example 3.1: *Show light propagation through a medium of refractive index 1.48 sand-witched in between medium of refractive index 1.46.*

Solution: *According to law of refraction, the critical angle is*

$$\theta_c = \arcsin(1.46/1.48)$$

$$= 80.6^o.$$

Therefore, the light striking the boundary between n_i and n_t at an angle greater than 80.6^o from the normal reflects back into the first medium. Again the angle of incidence equals the angle of reflection and light remains in the material of refractive index $n_i = 1.48$.

We saw that the reflection or refraction of light depends on the indices of refraction of the two media and on the angle at which light strikes the interface. The optical fiber works on this principle. Once light begins to reflect down the fiber, it will continue to do so under normal circumstances.

3.5.1 Numerical Aperture

Numerical aperture (NA) is the light-gathering ability of a fiber. Only light injected into the fiber at angles greater than the critical angle will be propagated. The numerical aperture, a dimensionless quantity, relates to the refractive indices of the core and cladding

$$NA = \sqrt{n_1^2 - n_2^2} \quad (3.28)$$

where n_1 is the refractive index of the core and n_2 is the refractive index of the cladding. We can also define the angles at which rays will be propagated by the fiber. These angles form a cone, called the acceptance cone, which gives the maximum angles of light acceptance. The acceptance cone is related to the NA as

$$\theta = \text{acrsin}(NA) \quad (3.29a)$$

$$\Rightarrow NA = \sin\theta \quad (3.29b)$$

where θ is the half angle of acceptance (Figure 3.12). The NA of a fiber is important because it gives an indication of how the fiber accepts and propagates light. A fiber with low NA requires highly directional light.

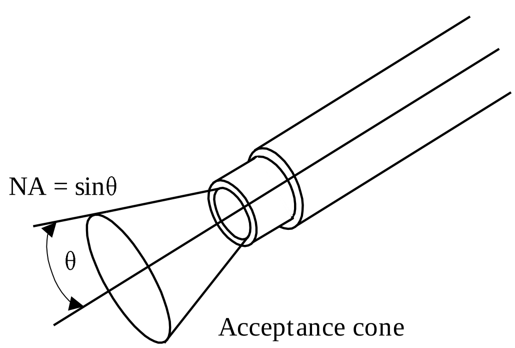


Figure 3.12 Numerical aperture of an optical fiber.

In general, fibers with a high bandwidth have a lower NA and allow fewer modes. Numerical apertures range from 0.5 for plastic fibers to 0.2 for graded-index fibers. A large NA promotes more modal dispersion, since more paths for the rays are provided.

3.5.2 Basic Fiber Construction

The optical fiber has two concentric layers called the core and cladding. The inner core is the light carrying part. The surrounding cladding provides the difference in refractive index that allows total reflection of light through the core. The index of cladding is less than 1% lower than that of the core.

The optical fiber is the signal-carrying member, similar in function to the metallic conductor in a wire. But the fiber is usually cabled, i.e., placed in a protective covering that keeps the fiber safe from environmental and mechanical damage.

Figure 3.13 shows the idea of light travelling through a fiber. Light injected into the fiber and striking the core-cladding interface at greater than the critical angle reflects back into the core. Since the angle of incidence and reflection is same, the reflected light will again be reflected. The light will continue to zigzag down the length of the fiber.

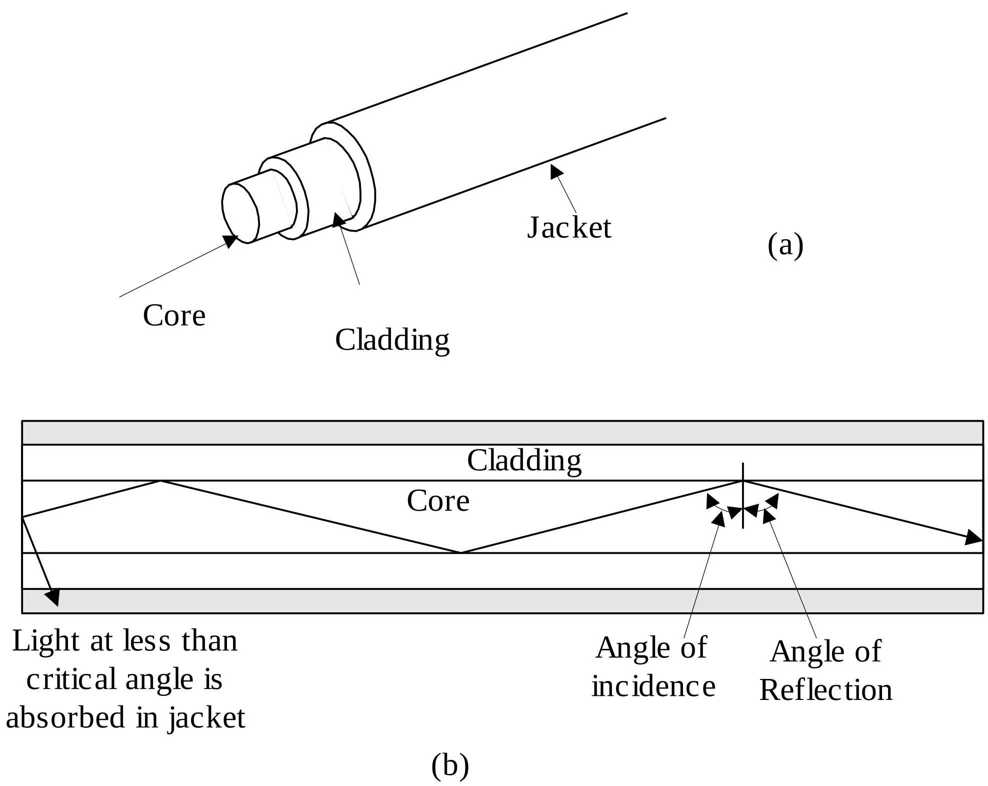


Figure 3.13 (a) An optical fiber (b) Total internal reflection in an optial fiber.

Light, however, striking the interface at less than the critical angle passes into the cladding, where it is lost over distance. The cladding is usually inefficient as a light carrier, and light in the cladding attenuated fairly rapidly.

Such total internal reflection forms the basis of light propagation through a simple optical fiber. The analysis, however, consider only meridional rays, those that pass through the fiber axis each time they are reflected. Other rays, called skew rays, those travel down the fiber without passing through the axis. The path of a skew ray is typically helical, wrapping around and around the central axis.

The specific characteristics of light propagation through a fiber depend on many factors, including (i) size, (ii) composition, and (iii) the light injected into the fiber. Fibers have exceedingly small diameters. For example; the diameter of the core/cladding of some commonly used fibers are: 8/125, 50/125, 62.5/125, and 100/140 (fiber sizes are usually expressed by first giving the core size, followed by the cladding size: thus, 50/125 means a core diameter of 50 μm and a cladding diameter of 125 μm). Just for comparison, one should know that human hair has a diameter of about 100 μm.

3.5.3 Fiber Classification

Optical fibers are classified in two ways, one way is by their material, i.e.;

1. Glass fibers have a glass core and glass cladding. The glass used in fibers is pure, highly transparent silicon dioxide or fused quartz. Impurities are purposely added to the pure glass to achieve the desired index of refraction. Germanium or phosphorus increases the index, while boron or fluorine decreases it.

2. Plastic-clad silica (PCS) fibers have a glass core and plastic cladding. Their performance, though not as good as all-glass fibers, is quite respectable.

3. Plastic fibers have a plastic core and plastic cladding. Compared with other fibers, plastic fibers are limited in loss and bandwidth. Their very low cost and easy use, however, make them attractive in applications where high bandwidth or low loss is not a concern. Their electromagnetic immunity and security allow plastic fibers to be beneficially used.

The second way to classify fibers is by the refractive index of the core and the modes that the fiber propagates. A mode is a set of guided electromagnetic waves propagating along a waveguide. For practical purposes, a mode is a path that a light ray can follow in travelling down a fiber. The number of modes supported by a fiber ranges from 1 to over 100,000. Thus, a fiber provides a path of travels for one or thousands if light rays, depending on its size and properties.

3.5.4 Refractive Index Profile

The refractive index profile describes the relation between the indices of the core and cladding. Two main relationships exist: step index and graded index. With this classification, there are three types of fibers:

1. Multimode step-index fiber (commonly called step-index fiber)
2. Multimode graded-index fiber (commonly called graded-index fiber)
3. Single-mode step-index fiber (commonly called single-mode fiber).

The characteristics of each type have important bearing on its suitability for particular applications.

3.5.4.1 Step-Index Fiber

The step-index fiber has a core with uniform index throughout. The profile shows a sharp step at the junction of the core and cladding (Figure 3.14). Mathematically it can be represented as

$$n(r) = \begin{cases} n_1 & r \langle a \qquad \text{(core)} \\ n_2 & r \geq a \qquad \text{(cladding)} \end{cases} \tag{3.30}$$

where *a* is the radius of the core. The multimode step-index fiber is the simplest type. It has a core diameter from 100 to 970 μm, and it includes glass or plastic constructions. As such, the step-index fiber is the most wide ranging, although not the most efficient in having high bandwidth and low losses.

Typical modal dispersion (will be discussed latter) figures for step-index fibers are 15 to 30 ns/km. This means that when rays of light enter a fiber at the same time, the ray following the longest path will arrive at the other end of 1 km-long fiber 15 to 30 nano-seconds after the ray following the shortest path. The delay of 15 to 30 ns may not seem like much, but

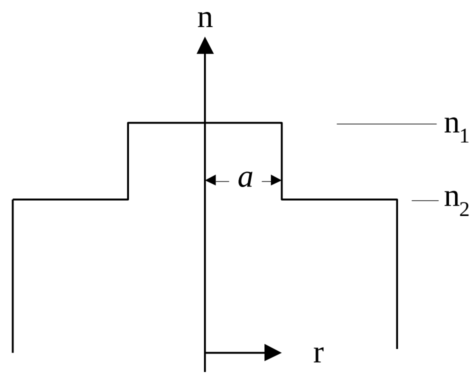


Figure 3.14 Refractive index profile for step index fibers; a is the core radius, n_1 is n_2 are the refractive indicies of the core and cladding, respectively.

dispersion is the main limiting factor on a fiber's bandwidth. Pulse spreading results in a pulse overlapping of adjacent pulses. Eventually, the pulses will merge so that one pulse cannot be distinguished from another. The information contained in the pulse is lost. Reducing the dispersion increases fiber bandwidth.

3.5.4.2 Graded-Index Fiber

The graded index has a non-uniform core. The index is highest at the centre and gradually decreases until it matches that of the cladding as shown in Figure 3.15. There is no sharp break between the core and the cladding. The variation in core refractive index is often expressed in the form:

$$n(r) = n_1.\left[1 - 2\Delta.\left(\frac{r}{a}\right)^{\alpha}\right]^{\frac{1}{2}} \qquad r<a \qquad (3.31a)$$

and

$$n(r) = n_1 (1 - 2.\Delta)^{\frac{1}{2}} \qquad r \geq a \qquad (3.31b)$$

where $\Delta = (n_1-n_2)/n_1$; here n_1 is the axial refractive index while n_2 is the refractive index of the cladding. The parameter α determines the core index variation. One can find that at r = *a*, n(*a*) = n_2, since Δ is usually much less than unity and $n(a) = n_1(1-\Delta)$.

One way to reduce modal dispersion is to use graded-index fibers. Here the core has numerous concentric layers of glass, somewhat like the annular rings of a tree. Each successive layer outward from the central axis of the core has a lower index of refraction.

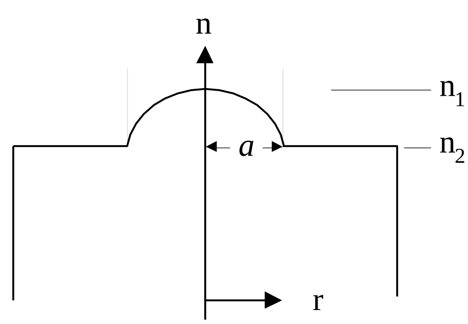


Figure 3.15 Refractive index profile for graded index fibers, where *a* and n's are defined in Figure 3.15.

3.5.4.3 Single-Mode Fiber

Another way to reduce modal dispersion is to reduce the core's diameter until the fiber propagates only one mode efficiently. The single-mode fiber has an exceedingly small core diameter of only 5 to 10 μm. Standard cladding diameter is 125 μm. Since this fiber carries only one mode, therefore, modal dispersion does not exist.

Single-mode fibers easily have a potential bandwidth of 50 to 100 GHz-km. The point at which a single mode fiber propagates only one mode depends on the wavelength of light carried. A wavelength of 820 nm results in multimode operation. As the wavelength is increased, the fiber carries fewer and fewer modes until only one remains. Single-mode operation begins when the wavelength approaches the core diameter. At 1300 nm, for example, the fiber permits only one mode. It becomes a single mode fiber.

The difference in propagation of light in single-mode and multimode fibers is that the optical energy in a single mode fiber travels in the cladding as well as in the cores, the cladding must be more efficient carriers of energy. In a multimode fiber, the light transmission characteristics of the cladding are basically unimportant. Indeed, because cladding modes are not desirable, a cladding with inefficient transmission characteristics can be tolerated. This situation does not hold for a single-mode fiber.

3.5.5 Dispersion in Fibers

The dispersion is the spreading of light pulse as it travels down the length of an optical fiber. Dispersion limits the bandwidth or information-carrying capacity of a fiber. The bit rate must be low enough to ensure that pulses do not overlap. A lower bit rate means that the pulses are farther apart and greater dispersion can be tolerated. There are three main types of dispersion:

3.5.5.1 Modal dispersion

In optical fiber light reflects at different angles for different paths (or modes), the path lengths of each mode is different. Thus, different rays take a shorter or longer time to travel the length of the fiber. The ray that goes straight down the centre of the core without reflecting arrives at the other end first. Other rays arrive latter. Thus, light entering the fiber at the same time exits the other end at different times. The light has spread out in time. This spreading of an optical pulse is called *modal dispersion*. A pulse of light that begins as a tightly and precisely defined shape has dispersed, i.e., spread over time.

Modal dispersion occurs only in multimode optical fibers and can be reduced by;

1. using a small core diameter, e.g., a 100 μm diameter core allows fewer paths than a 200 μm diameter core,
2. using a graded index fiber. Light travels faster in a lower index of refraction. So the farther the light is from the central axis, the greater its speed. Each layer of the core refracts the light. Instead of being sharply reflected as it is in a step-index fiber, the light is now bent or continually refracted in an almost sinusoidal pattern. Those rays that follow the longest path by travelling near the outside of the core have a faster average velocity. The light travelling near the centre of the core has the slowest average velocity. As a result, all rays tend to reach the end of the fiber at the same time. The graded index reduces the modal dispersion to 1 ns/km or less.
3. using single-mode fiber, which permits no modal dispersion.

3.5.5.2 Material dispersion

Different wavelengths travel at different velocities through a fiber, even in the same mode. Earlier we saw that the index of refraction is

$$n = \frac{c}{v} \tag{3.32}$$

where c is the speed of light in vacuum and v is the speed of the same wavelength of light in the material. According to the speed, v, for different wavelength, index of refraction changes. Dispersion from this phenomenon is called material dispersion since it arises from material properties of the fiber.

Material dispersion is of greater concern in single-mode fibers. In a multimode system, modal dispersion is usually significant enough that material dispersion is not a factor. The 820 nm to 850 nm region of the spectrum is used as transmission wavelength for many fiber-optic systems. In this region, material dispersion can be approximated as being roughly equal to 0.1 ns/nm of spectral width.

3.5.5.2 Waveguide dispersion

Waveguide dispersion, most significant in a single-mode fiber, occurs because optical energy travels in both the core and cladding, which have slightly different refractive indices. The energy travels at slightly different velocities in the core and cladding because of the slightly different refractive indices of the materials.

Problems

3.1 Calculate the lateral displacements of rays of light incident on a block of glass with parallel sides at the following angles (a) 2.0^o, (b) 5.0^o, (c) 20^o, (d) 40^o, (e) plot a graph of the displacements, d, verses incident angle.

3.2 A rectangular aquarium is to be filled with water. The sides are made of glass plates of 5.0 mm thick. Inside, the walls are 30 cm apart, and the refractive index of the glass is 1.517.

If a ray of light is incident on one side at an angle of 45^o, find the lateral displacement produced when the tank is (a) empty and (b) filled with water.

3.3 Compute the reflectance at normal incidence for the following materials: (a) diamond, n = 2.426; (b) quartz, n =1.547; (c) rutile, n = 2.946; (d) crown glass, n = 1.526.

3.4 Calculate the critical angles for water/air interface. The refractive index of water is 1.333.

3.5 Unpolarized light strikes a smooth glass (n = 1.781) surface at an angle 45^o. Calculate the intensities of the reflected p and s components. Also find the degree of polarization of the reflected ray.

3.6 What are the advantages of paraxial ray approximation for matrix formulation of geometrical optics? Which limitations arise due to this approximation?

3.7 Consider a ray propagating in free space is incident on a medium with a refractive index n = 1.53, as shown in Figure 3.16.

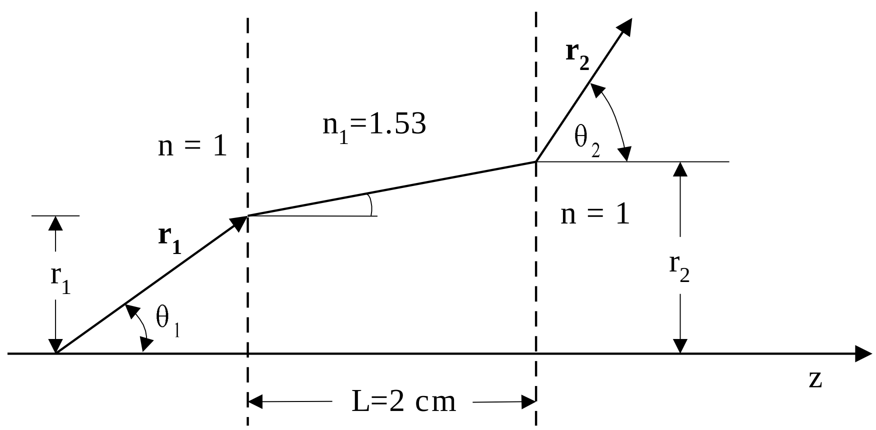


Figure 3.16

After travelling a distance of 2 cm, it emerges out of this medium into free space again. If θ_1 = 0.5 mrad, calculate θ_2 using (a) Snell Law (b) Ray matrix approach.

3.8 Show that the ABCD matrix for a ray entering a spherical dielectric interface from a medium of refractive index n_1 to a medium n_2 is

$$\begin{vmatrix} 1 & 0 \\ \dfrac{n_2 - n_1}{n_2}\dfrac{1}{R} & \dfrac{n_1}{n_2} \end{vmatrix}$$

where R is the radius of the spherical surface.

Books for further reading

R. Ditteon, *Modern Geometrical Optics* (John-Wiley & Sons, NY, 1998)

F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed. (McGraw-Hill, New York, 1985)

E. Hecht, *Optics*, 2nd ed. (Addison-Wesley, Reading, Mass., 1990)

R. Guenther, *Modern Optics*, (John Wiley & Sons, New York, 1990).

J. Wilson and J.F.B. Hawkes, *Optoelectronics: An Introduction*, (Prentice-Hall International, India, 1996)

O. Svelto, *Principles of Lasers*, (Plenum Press, New York, 1989)

P. W. Milonni and J. H. Eberly, *Lasers*, (John Wiley & Sons, New York, 1991).

A. E. Seigman, *Lasers and Masers*, (1971)

Unit-4

LASERS:

Introductory Concepts

Objective

The objective of this unit is to give the student introductory but fundamental concept regarding lasers. These concepts include the fundamental difference between spontaneous and stimulated emission, their origins, and basic requirements of laser emission, e.g., active medium, population inversion, and optical feedback. A brief history is also given for general interest.

4.1 Introduction and Brief History

Laser is a device that amplifies light and produces a highly directional, high intensity beam that typically has an almost pure frequency or wavelength. It comes in sizes ranging from approximately one-tenth the diameter of a human hair to the size of a very large building, in powers ranging from 10^{-9} to 10^{20} watts, and in wavelengths ranging from the microwave to the soft x-ray spectral regions with corresponding frequencies from 10^{11} to 10^{17} Hertz. Lasers have pulse energies as high as 10^{4} joules and pulse duration as short as 10^{-15} seconds.

Lasers are key components of some of our most modern communication systems and are the "phonographic needle" of compact disc players. They are used for heat treatment of high-strength materials, such as the pistons of automobile engines, and provide a special surgical knife for many types of medical procedures. They act as target designators for military weapons and are used in the checkout scanners. They can easily drill holes in the most durable of materials and can weld detached retinas within the human eye.

The word laser is an acronym for Light Amplification by Stimulated Emission of Radiation. One of the major processes in lasers is stimulated emission. Albert Einstein gave the idea of stimulated emission of radiation in 1917. Until that time, physicists had believed that a photon could interact with an atom in only two ways: it could absorbed and raise the atom to higher energy level or be emitted as the atom is dropped to a lower energy level. Einstein proposed a third possibility that a photon with energy corresponding to that of an energy level transition could stimulate the atom in the upper level to drop to the lower level by emission of another photon with the same energy and phase as the first one.

The stimulated emission is least known because at thermodynamic equilibrium more atoms are in lower energy levels than in higher ones. Thus a photon is much more likely to encounter an atom in a lower level and be absorbed than to encounter one in a higher level and stimulate emission.

The first efforts to use the idea of stimulated emission were a few decades latter in the microwave region and led to the invention of maser "microwave amplification by stimulated emission of radiation". The maser concept evolved nearly simultaneously in the United States and Soviet Union. A physicist Charles H. Townes of the Colombia University is known as the inventor of maser.

Townes thought that molecules in the excited state could be stimulated to emit microwaves when placed in a special resonant cavity designed to enhance the emission. He outlined the idea to post-doctoral fellow Herbert Zeiger and graduate student James P. Gordon in 1951, and by 1953 they had a working maser. Meanwhile, Alexander M. Prokhorov and Nikolai Basov (1954) of Lebedev Physics Institute in Moscow calculated the details of maser action and published shortly after Townes's results. The contributions of all these three men were recognised and they were awarded Nobel Prize in physics in 1964.

After maser, Townes and other physicists began looking beyond the microwave region to shorter wavelengths. They realised that at those wavelengths the physical conditions required to produce stimulated emission would be very different. Townes and Arthur L. Schawlow (of Bell Labs.) worked out many key parameters. Their results were published in a major paper in 1958 and they also filed a patent application before the paper was published.

Meanwhile, a graduate student at Columbia, Gordon Gould, was working out his own analysis of the conditions required for stimulated emission at visible wavelengths. Gould wrote his proposals in a set of notebooks in 1957 but did not try to publish his results promptly. He wanted to patent his work but due to some bad legal advice, he did not file a patent application until 1959, about nine months after the Schawlow-Townes patents application was submitted. The Schawlow-Townes patent was granted promptly, but Gould's application ran into a lengthy process. Finally, Gould got four patents in 1977, 1979, 1987 and 1988, based on divisions of his original application.

Schawlow and Townes have received many scientific honours for their work, but Gould received little recognition until his patents were issued. Although Gould loses the prestige race, yet he benefited financially. However, it was he, who first coined the word LASER in his notebooks. Schawlow and Townes described their idea as an "optical maser".

Publication of Schawlow-Townes paper stimulated many scientists to build lasers, and interest spread beyond the narrow scientific community. Schawlow, Gould and most researchers thought that gases were the best materials for lasers. However, Theodore H. Maiman, a young scientist at Hughes Research Laboratories in Malibu, California, quietly disagreed. He preferred synthetic ruby crystals to gases, although some theorists insisted that ruby would not work. Maiman, who had studied energy levels in ruby extensively, proved that the theorists were wrong. In mid-1960 he proudly demonstrated the world's first laser, the ruby laser. The laser era was born.

Maiman had to face a lot of difficulties; Hughes management told him to stop work on the ruby laser, the prestigious journal, Physical Review Letters, rejected his report of ruby laser but it was published in Nature. Today, however, Maiman is universally recognised as the person who built first laser and has received a number of honours.

Maiman's demonstration of the ruby laser opened the floodgates and was followed by the demonstration of helium neon laser by Ali Javan, W.R. Bennett Jr., and Donald R. Harriot at Bell Telephone Laboratories in Murray Hill, New Jersey. Their first helium-neon laser operated at 1.15 micrometers (μm) in the near infrared. Latter, other researchers found the 632.8 nanometer (nm) red line which made the helium-neon laser most popular.

The laser boom really got going. In 1961, L. F. Johnson and K. Nassau demonstrated the first solid-state neodymium laser in which the Nd ions were dopant in calcium tungstate, but the today's best choice of neodymium host for best commercial applications- yttrium aluminium garnet (YAG), was demonstrated as a laser material in 1964. Three separate groups demonstrated semiconductor diode lasers nearly simultaneously in fall 1962. All the teams demonstrated the same gallium arsenide diodes cooled to the 77 K temperature of liquid nitrogen and pulsed with high-current pulses lasting a few microseconds. The next few years saw the birth of several more important lasers. W. B. Bridges (1964) observed 10 laser transitions in the blue and green parts of the spectrum from singly ionised argon. C. Kumar and N. Patel (1964) obtained a 10.6 μm laser emission from carbon dioxide. Sorokin and J. R. Lankard (1966) demonstrated the first organic dye laser; hydrogen chloride emitting at 3.7 μm was demonstrated in 1965 by J. V. V. Kaspar and G. C. Pimentel.

4.2 Basic Principle of Lasers

In order for most lasers to operate, three basic conditions must be satisfied. First, there must be an active medium, that is, a collection of atoms, molecules, or ions that emit radiation in the optical part of the electromagnetic spectrum. Second, a condition known as population inversion, i.e., more number of atoms available at the higher energy level as compare to lower energy level, must exist. This condition is highly abnormal in nature. It is created in a laser by an excitation process known as pumping. Finally, for true laser oscillation to take place there must be some form of optical feedback present in the laser system. If this was not present, the laser might serve as an amplifier of narrow-band light, but it could never produce the highly collimated, monochromatic beam that makes the laser so useful.

4.2.1 Active Medium

The heart of a laser system is a material capable of emitting radiation of the required energy. This material, known as active medium, may have any form e.g. solid, liquid, or gas but must contain a set of energy levels in which it can absorb or emit energy in the form of optical radiation.

4.2.1.1 Energy levels

In the macroscopic world, energy might seem to vary continuously, like the level of sand in a pail. If we look closely at the sand, we can see that it is made up of many separate grains, and that we can add or subtract only one grain at a time. Likewise, atoms and molecules can only have certain amounts of energy. We call these energy states or levels.

For example, let's look at the simplest atom (hydrogen) in which a single electron circles a nucleus that contains a single proton. At first glance, the hydrogen atom looks like a very simple solar system, with a single planet (the electron) orbiting a star (the proton). The force that makes the atom stable is the attraction between the positive charge of the proton and the negative charge of the electron.

The electron can occupy only certain orbits, as shown in Figure 4.1. We show the orbits as circles for simplicity, but we can't really measure exactly what the orbit looks like. If we add energy to our simple planetary system, the planet would move farther from the star. The same happens in the hydrogen atom. As we add more energy to the atom, the electron moves to more and more distant orbits. However, there is crucial difference between the behaviour of a hydrogen atom and our imaginary planetary system. The planet can be at any distance from the sun, or could even fall into it. However, the electron can only occupy certain orbits also called energy levels. These energy levels are plotted in Figure 4.1. Atom is, in general, lies in the lowest possible energy level, the ground state.

The energy levels in the hydrogen atom get closer together as they get higher above the ground state. Eventually the differences become vanishingly small. If the electron gets too much energy, it escapes from the atom altogether, a process called ionization. If we define the energy of the ionized hydrogen atom to be zero, we can write the energy of the atom E as a negative number using the simple formula:

$$E = -R/n^2 \tag{4.1}$$

where, R is a constant ($2.178\text{x}10^{-18}$ Joules), n is the quantum number of the orbit (counting outwards, with one the innermost level).

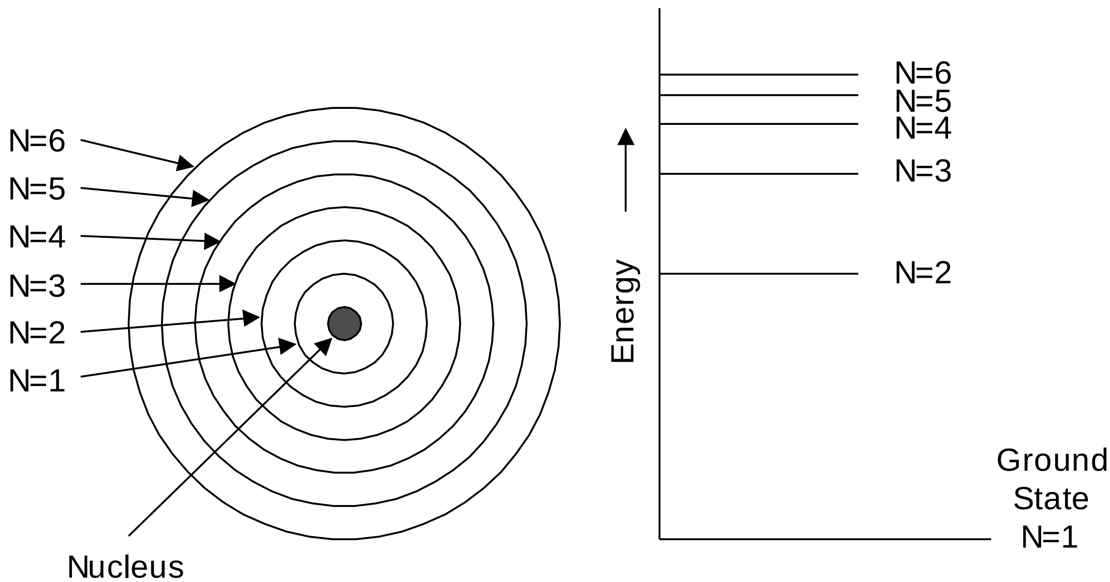


Figure 4.1 The hydrogen atom and the corresponding energy levels.

4.2.1.2 Transitions

The process of making changes or "transitions" between energy levels is very important to laser physics. To look at the process, let's start again with the hydrogen atom. The electron needs to gain energy to move from ground state to higher energy level. Conversely, it must release energy when it drops from a higher level to lower one. One of the most convenient ways (although by no means the only way) for an electron to absorb or release energy is as a photon, a quantum of electromagnetic energy. The photon energy equals the difference in energy between the two energy levels.

Suppose we start with the electron in the ground state and want to raise it up one step to the first excited level. To do so, we must give the electron exactly as much extra energy as the difference in energy between the ground state and the first excited level. Conversely, for the electron to drop from the first excited level to the ground state, it must emit a photon with as much energy as the difference in energy between the two levels. In short, the photon energy equals the transition energy. The energy of a photon can be represented by $h\nu$; in the frequency term, where h (= 6.626×10^{-34} J.s) is the Plank's constant and ν is the frequency of the transition, or in the wavelength form hc/λ, where c is the velocity of light in vacuum and λ is the wavelength of the photon. An electron in a hydrogen atom can emit or absorb light at only certain wavelengths.

This neat ordering of energy levels is evident only in hydrogen atom, where there is just a single electron. Add more electrons, and the energy level picture quickly becomes more complicated. Electrons interact with each other, and with the nucleus, shifting energy levels slightly. Electrons can occupy subshells within each shell. The more complex the energy level structure, the more transitions between energy levels is possible. The more transitions, the larger number of possible spectral lines. Superimpose them all on a single spectrum and look almost like a set of random lines.

Another complication is that all transitions are not equally likely. One reason is that more atoms are in some states than in others. For example, under normal circumstances, more atoms are in ground state than in excited levels. Another is that quantum mechanical rules make some transitions much more likely than others. That means that an atomic or molecular species will absorb or emit some wavelengths much more strongly than others. This effect shows up both in absorption spectra (which shows that wavelengths the material absorbs when light from another source passes through it) and emission spectra (the wavelengths the material emits when it is itself de-excited).

4.2.1.3 Types of Transitions

So far we have concentrated on electronic transitions, partly because we picked the hydrogen atom as our introductory example. Electronic transitions can cover a wide range of wavelengths. These occur at ultraviolet, visible, or infrared wavelengths from 100 nm in the ultraviolet through to near infrared wavelengths.

The shortest wavelengths come from inner-shell electronic transitions in heavy elements, which involve much more energy than outer-shell transitions. These short wavelengths are considered to be X-rays. On the other hand, transitions between high lying electronic energy levels (say, level 18 and 19 of hydrogen) involve very little energy, putting them deep in the infrared, microwave, or even radio-frequency range. Because these are qualitatively different than higher-energy transitions of outer electrons, they are put into special class called

Rydberg transitions.

Neither Rydberg transitions nor X-ray emission falls into the optical region, which is a very small portion of the electromagnetic spectrum, therefore, these are not likely events under normal laser conditions. Transitions between nuclear energy levels can produce even higher energy photon, called gamma rays.

On the other end of the wavelength spectrum are transitions between vibrational and rotational energy levels of molecules. Vibrational transition energies typically correspond to wavelengths of a few to tens of micrometers; rotational transitions have less energy, typically corresponding to wavelengths of at least 100 micrometers.

Transitions in two or more types of energy levels can occur simultaneously. For example, a molecule can undergo a vibrational and rotational transition simultaneously, with the resulting wavelength close to that of more energetic vibrational transition. Many infrared lasers emit families of closely spaced wavelengths on such vibrational-rotational transitions.

Remember in considering transitions that longer wavelengths correspond to lower energy, and shorter wavelengths correspond to higher energy ($E = hc/\lambda$). Thus, the energy of a vibrational transition is much larger than that of a rotational transition, even though the rotational wavelength is much larger. A combination of a rotational and vibrational transition thus has only slightly different energy than the original vibrational transition. Transition energies or frequencies add together in a straightforward manner:

$$E_{1+2} = E_1 + E_2 \tag{4.2}$$

where, E_{1+2} is the combined transition energy, E_1 and E_2 are the energies of the separate transitions. The same rule holds for frequencies, with ν substituted for the energy. However, wavelength (λ) of combined transitions add by an inverse rule:

$$1/\lambda_{1+2} = 1/\lambda_1 + 1/\lambda_2 \tag{4.3a}$$

or $$\lambda_{1+2} = 1/(1/\lambda_1 + 1/\lambda_2) \tag{4.3b}$$

4.2.1.4 Spontaneous Emission

We know that atoms have well defined energy levels and they can be pushed to the higher energy states (excited states) by absorption of a photon or by some other means e.g. electric discharge. Atoms spent some time in the excited state and then decay to the lower energy state. The average time required to de-exite the 1/e number of atoms from the upper level to the lower level is called lifetime of the state. This lifetime can be very small e.g. in nano-seconds and up to a few seconds. Typically, for electronic transitions it is in the order of ten nano-seconds. After spending this period in the excited state, atoms come down to the lower energy state by emitting a photon. This emission of photon from an excited atom is called spontaneous emission. This phenomenon one sees in daily life. Sun, bulb, tubelight and all the fluorescent devices emit photons spontaneously in the visible region. These photons are emitted in all the direction and illuminate the whole area, probably, this is the blessing of spontaneous emission.

Example 4.1*: Describe spontaneous emission from excited helium*

When we see helium discharge lamp directly, it appears to emit pinkish white light as shown in Figure 4.2. In fact in the discharge, helium atoms are being excited in the upper energy levels. They soon give up their energies by dropping down to lower energy level with emission of photons. This spontaneous emission involves quite a number of different energy levels and thus produces the photons of different colours. When viewed the same discharge through a diffraction grating (an optical element to

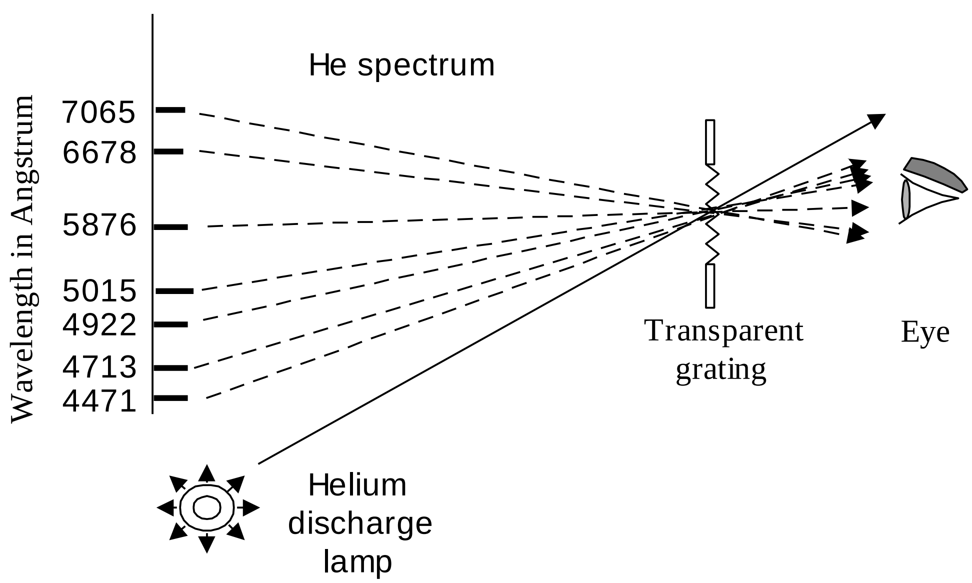


Figure 4.2 Helium discharge spectrum observed through a transmission grating.

separate different wavelengths of light) multiple images of different colours of the lamp appear at different angles. A strong yellow line at 588 nm is prominent but violet, green, blue, red and deep red lines can also be seen.

To understand the spontaneous emission a little more, let us consider two energy levels, 1 and 2, of some given material, their energies being E_1 and E_2 ($E_1<E_2$) and population densities N_1 and N_2 as shown in Figure 4.3. It is convenient, however, to take level 1 to be the ground level. Let us now assume that an atom (or molecule) of the material is initially in level 2. Since $E_2>E_1$, the atom will tend to decay to level 1. The atom must therefore release the energy, corresponding energy difference (E_2-E_1). This energy is delivered in the form of a photon of frequency ν given by

$$\nu = \frac{(E_2 - E_1)}{h} \tag{4.4}$$

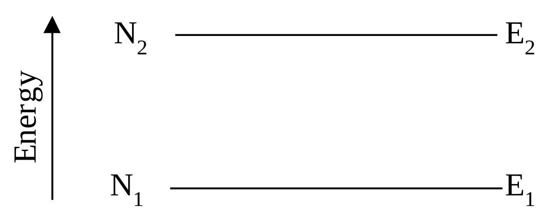


Figure 4.3 Two level energy system.

Spontaneous emission is therefore characterized by the emission of a photon of energy $h\nu = E_2-E_1$, when the atom decays from level 2 to level 1 (Figure. 4a). Note that radiative emission is just one of the two possible ways for the atom to decay. The decay can also occur in a non-radiative way. In this case the energy difference E_2-E_1 is delivered in some form other than a photon (e.g., it may go into kinetic energy of the surrounding molecules).

The probability of spontaneous emission can be characterised in the following way: Let us suppose that, at time t, there are N_2 atoms (per unit volume) in level 2. The rate of decay of these atoms due to spontaneous emission, i.e., $(dN_2/dt)_{sp}$, will obviously be proportional to N_2. We can therefore write

$$\left(\frac{dN_2}{dt}\right)_{sp} = -A_{21}N_2 \tag{4.5}$$

The coefficient A_{21} is called the spontaneous emission probability or the Einstein coefficient. The quantity $\tau_{sp}=1/A_{21}$ is called the spontaneous emission lifetime. The numerical value of A_{21} (and τ_{sp}) depends on the particular transition involved.

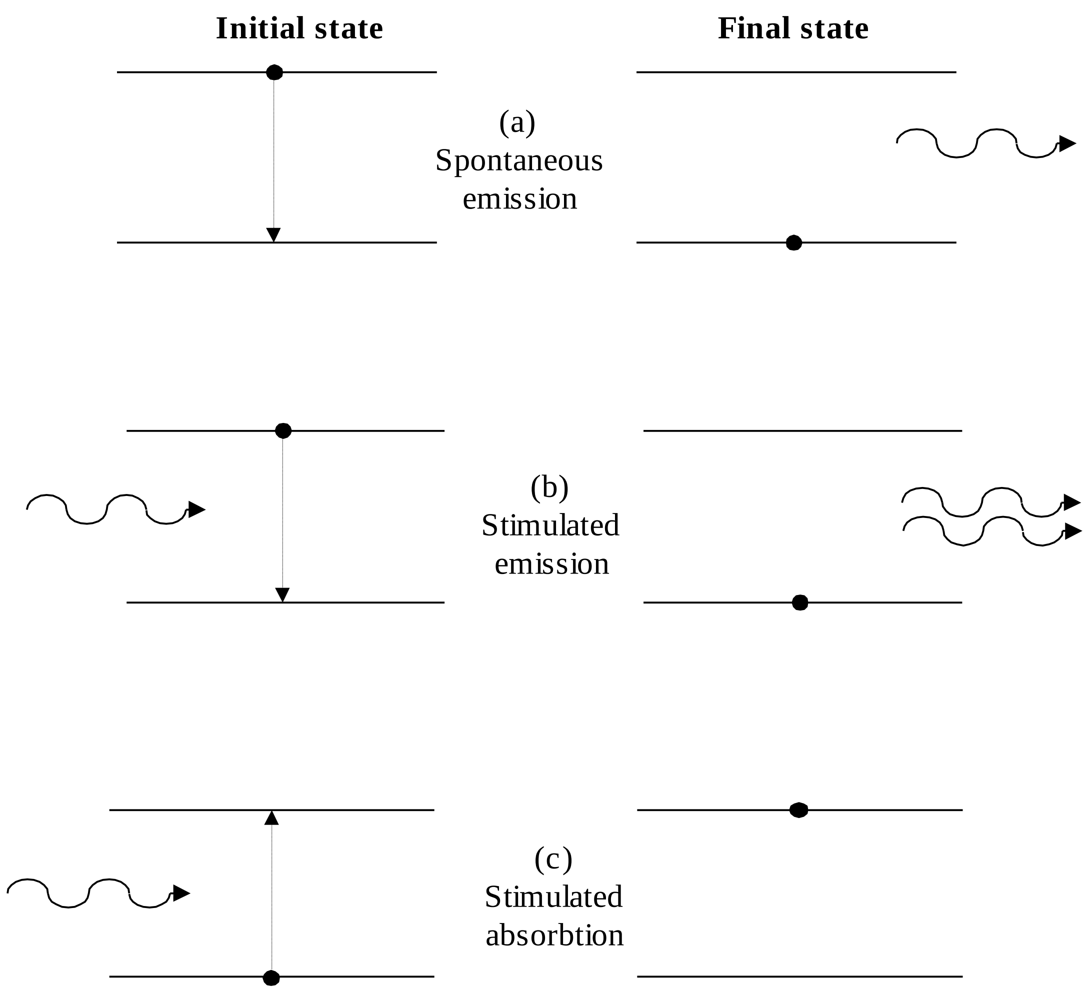


Figure 4.4 Energy level diagram illustrating (a) spontaneous emission, (b) stimulated emission and (c) stimulated absorption.

In spontaneous emission each individual atom acts like a small randomly oscillating antenna emitting at the transition frequency. Therefore, the total emission from a collection of spontaneously emitting atoms exhibit noise-like character (Figure 4.5).

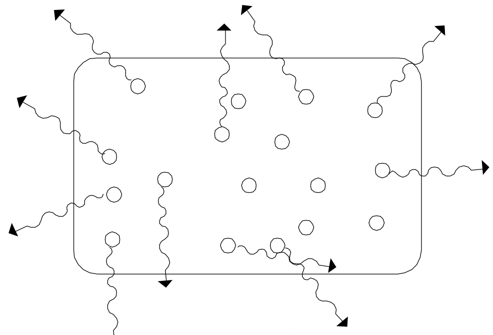


Figure 4.5 Spontaneous emission is incoherent or noise-like character, emerging randomly in all directions.

4.2.1.5 Stimulated Emission

In stimulated emission, the atom is triggered to undergo the transition in the presence of photons of energy (E_2-E_1). In these transitions each individual atom acts like a passive resonant antenna that is set oscillating by the applied signal. Therefore, the internal motion or oscillation in the atom is not random, but is driven by an applied signal. Let us again assume that the atom is found initially in level 2 and a photon of frequency ν as of equation (1) is incident on the material. Since this wave has the same frequency as the atomic frequency, there is a finite probability that this wave will force the atom to undergo the transition $2 \rightarrow 1$. In this case the energy difference E_2-E_1 is delivered in the form of a photon that adds to the incident one as shown in Figure 4b. There is, however, a fundamental distinction between the spontaneous and stimulated processes. In the case of spontaneous emission, the atom emits a photon that has no definite phase relation with that emitted by another atom. Furthermore, the photon can be emitted in any direction. In the case of stimulated emission, since the process is forced by the incident photon, the emission of any photon adds in phase to that of the incoming wave. This wave also determines the direction of the emitted wave. In this case, too, we can characterise the process by means of the equation

$$\left(\frac{dN_2}{dt}\right)_{st} = -\rho_\nu B_{21} N_2 \tag{4.6}$$

where $(dN_2/dt)_{st}$ is the rate at which transitions $2 \rightarrow 1$ occur as a result of stimulated emission, B_{21} is called the Einstein coefficient for stimulated emission, the energy density $\rho_\nu = Nh\nu$, where N is the number of photons per unit volume having frequency ν.

4.2.1.6 Absorption

All the atoms and molecules have discrete energy levels. An atom can absorb a photon and become excited. Let us now assume that the atom is initially in level 1. If this is the ground level, the atom will remain in this level unless some external stimulus is applied to it. We shall assume that a photon of frequency ν given by equation (1) is incident on the material. In this case there is a finite probability that the atom will be raised to level 2 (Figure 4.4c). The energy difference E_2-E_1 required by the atom to undergo the transition be obtained from the energy of the incident photon. This is the absorption process.

In a similar fashion to equations 2 & 3, we can write an equation for absorption rate,

$$\frac{dN_1}{dt} = -\rho_\nu B_{12} N_1 \qquad (4.7)$$

where N_1 is the number of atoms per unit volume that at the given time are lying in level 1 and B_{12} is called the Einstein coefficient for stimulated absorption.

In the absorption process, the incident photon is simply absorbed to produce the $1 \rightarrow 2$ transition. In the beginning of the century, Einstein also showed that the coefficients of stimulated emission and absorption are equal, i.e. $B_{21}=B_{12}$ (assuming that degeneracy in both the upper and lower level is same)

4.2.1.7 The Einstein Relations

Einstein showed that the parameters describing the above three processes are related through the requirement that for a system in thermal equilibrium, the rate of upward transitions (E_2 to E_1) must be equal to the rate of downward transition processes.

We can write the upward transition rate as $N_1\rho_\nu B_{12}$. The total downward transition rate is the sum of the induced and spontaneous contributions i.e. $N_2\rho_\nu B_{21}+N_2A_{21}$. In the preceding discussions A_{21}, B_{21} and B_{12} are called the Einstein coefficients. The relationship between them can be established as follows.

For a system in equilibrium, the upward and downward transition rates must be equal and hence we have

$$N_1\rho_\nu B_{12} = N_2\rho_\nu B_{21} + N_2 A_{21} \qquad (4.8)$$

Thus

$$\rho_\nu = \frac{N_2 A_{21}}{N_1 B_{12} - N_2 B_{21}} \qquad (4.9a)$$

or

$$\rho_\nu = \frac{A_{21}/B_{21}}{N_1/N_2 - 1} \qquad (4.9b)$$

The populations of various energy levels of a system in thermal equilibrium are given by Boltzmann statistics to be:

$$N_j = \frac{N_0 \exp(-E_j/kT)}{\sum_i \exp(-E_i/kT)} \qquad (4.10)$$

where N_j is the number of atoms in the jth level with energy E_j, N_0 is the total number of atoms, k is Boltzmann constant, and T is temperature in Kelvin.

Hence
$$\begin{aligned} \frac{N_1}{N_2} &= \exp((E_2 - E_1)/kT) \\ &= \exp(h\nu/kT) \end{aligned} \qquad (4.11)$$

From equations 4.9b and 4.11 we get ratio of spontaneous emission to stimulated emission in thermodynamic equilibrium as

$$\frac{A_{21}}{\rho_\nu B_{21}} = \exp(h\nu / kT) - 1 \qquad (4.12)$$

Equation (4.9) shows that whenever there is emission of photons, each in thermal equilibrium, stimulated emission is always present. ;However, the ratio of stimulated emission to spontaneous emission depends on the temperature (in thermal equilibrium) frequency of the photons and the number of photons available for stimulation.

Example 4.2: *Calculate the ratio of spontaneous emission to stimulated emission for tungsten filament lamp operating at a temperature of 1500 K (assume the average frequency* ν *to be* $5x10^{14}$ *Hz).*

Solution: *The ratio R of spontaneous emission to stimulated emission (* $R = \frac{A_{21}}{\rho_\nu B_{21}}$ *) is as follows*

$$R = \exp(h\nu / kT) - 1$$

Taking $k=1.38x10^{-23}$ JK^{-1}, $h=6.6x10^{-34}$*J.s,* ν $=5x10^{14}$ *Hz.*

$$R = \exp\left(\frac{6.6x10^{-34}.5x10^{14}}{1.38x10^{-23}.1500}\right) - 1$$
$$\approx e^{16} \approx 8.4x10^{6}$$

This confirms that under conditions of thermal equilibrium stimulated emission is not very significant. For sources operating at lower temperatures and higher frequencies stimulated emission is even less likely.

Example 4.3*: At what temperature are the rates of spontaneous and stimulated emission equal for* λ*=550 nm radiation? At what wavelength are they equal at room temperature (T=300 K)?*

Solution*: For spontaneous and stimulated emission rates to be equal at wavelength of 550 nm, we have*

$$R = \exp\left(\frac{h.c}{\lambda kT}\right) - 1 = 1$$
$$\Rightarrow T = \frac{h.c}{\lambda k}\frac{1}{\ln(1+R)}$$

Substituting values of $h=6.6x10^{-34}$ J.s, $k=1.38x10^{-23}$ JK^{-1}, $c=3x10^{8}$ m/s, λ=550 nm, and R=1, we get T=37635 K. Hence to achieve the equal spontaneous and stimulated emission temperature of more that thirty seven thousand Kelvin is required.

Similarly for equal spontaneous and stimulated emission rates at room temperature we can find the required wavelength, λ, as

$$\lambda = \frac{h.c}{k.T}\frac{1}{\ln(1+R)}$$

Again substituting values of all the constants and temperature T=300 K, we found λ~69 μm which lies in the microwave region. Thus radiation of 69 μm wavelength have equal spontaneous and stimulated emission rates at room temperature.

The above discussion indicates that the process of stimulated emission competes with the processes of spontaneous emission and absorption. Clearly if we wish to amplify a beam of light by stimulated emission then we must increase the rate of this process in relation to the other two processes. To achieve this for a given pair of energy levels we must increase i) radiation density and ii) the population density N_2 of the upper level in relation to the population density N_1 of the lower level. We shall show that to produce laser action we must create a condition in which $N_2>N_1$, even though $E_2>E_1$ that is we must create a so-called population inversion.

4.2.2 Laser Pumping

The population inversion required for light amplification constitutes an abnormal distribution of atoms among the various available energy levels. To understand how light amplification can be achieved in a medium, it is necessary to consider the Boltzmann distribution and then pumping mechanism to achieve the population inversion.

4.2.2.1 Population Inversion

The population inversion condition required for light amplification is a non-equilibrium distribution of atoms among the various energy levels of the atomic system. The Boltzmann distribution, which applies to a system in thermal equilibrium, is given by equation 7, it is obvious that as E_j increases N_j decreases for a constant temperature. We note that if the energy difference between E_1 and E_2 is nearly equal to kT (~0.025 eV at room temperature) then the population of the upper level would be 1/e or 0.37 times of the lower level. For an energy difference large enough to give visible radiation (~2.0 eV), however, the population of the upper level is almost negligible.

Example 4*: An atom has two energy levels with a transition wavelength of 694.3 nm. Assuming that all of the atoms in an assembly are in one or other of these levels, calculate the percentage of the atoms in the upper level at room temperature (T=300 K) and at T= 500 K.*

Solution: *The energy of the radiation of wavelength 694.3 nm,*

i.e. $E=hc/\lambda=6.6\text{x}10^{-34}.3\text{x}10^{8}/694.3\text{x}10^{-9}=2.85x10^{-19}$ *Joules*

We have

$$\frac{N_2}{N_1}=\exp\left(\frac{-E}{kT}\right)=\exp\left(\frac{-2.85\text{x}10^{-19}}{1.38\text{x}10^{-23}.T}\right)$$

At room temperature i.e. T=300 K, $N_2/N_1=1.2\text{x}10^{-30}$*, i.e. population of the upper level is* $1.2\text{x}10^{-28}$*% of the lower level in thermal equilibrium.*

At 500 K, $N_2/N_1=1.12\text{x}10^{-17}$*, i.e., population of the upper level is* $1.12\text{x}10^{-15}$*% of the lower level in thermal equilibrium.*

This shows that population of the upper level increases about 10^{13} *times by increasing the temperature up to 500 K from room temperature, but in both cases* $N_2 << N_1$.

Example 4.5: *The relative number of atoms* N_1 *and* N_2 *in two energy levels* E_1 *and* E_2 *separated by an energy gap* E_2-E_1 *are given at thermal equilibrium by Boltzmann ratio. Evaluate the ratio* N_2/N_1 *for the following cases:*

an optical transition, λ*=550 nm, at room temperature, 300 K;*

a microwave transition, ■ = 3 GHz, at room temperature;

A 10 GHz transition at liquid-helium temperature, 4.2 K.

For an optical transition at λ *=550 nm to have* N_2/N_1 *= 0.1, what temperature would be* *required?*

Solution:

The energy difference for λ = 550 nm is $E_2 - E_1 = hc/550\text{x}10^{-9} = 3.6\text{x}10^{-19}$ *J*

$$\frac{N_2}{N_1} = \exp\left(\frac{-3.6\text{x}10^{-19}}{1.38\text{x}10^{-23} \times 300}\right)$$
$$\approx e^{-87} \approx 10^{-37}$$

The number of atoms present in the upper energy level for a visible transition at room temperature is negligibly small.

The energy difference for ■ *= 3 GHz,* $h\nu = 6.6\text{x}10^{-34}.3\text{x}10^{9} = 1.98\text{x}10^{-24}$ *J*

$$\frac{N_2}{N_1} = \exp\left(\frac{-1.98\text{x}10^{-24}}{1.38\text{x}10^{-23} \times 300}\right)$$
$$\approx 0.99952$$

In the microwave region, number of atoms present in the upper energy level is a little less than the lower energy level at room temperature. Therefore to create a population inversion in microwave region is easier than in optical region.

The energy difference for ■ *= 10 GHz,* $h\nu = 6.6\text{x}10^{-24}$ *J*

$$\frac{N_2}{N_1} = \exp\left(\frac{-6.6\text{x}10^{-24}}{1.38\text{x}10^{-23} \times 4.2}\right)$$
$$\approx 0.89237$$

At low temperature the number of atoms present in the upper energy level decreases.

To have $N_2/N_1 = 0.1$ *at* λ *= 550 nm, the temperature required is 11329 K.*

This shows that to increase the number of atoms in the upper energy level at thermal equilibrium requires very high temperatures.

From the above example it is obvious that population can not be achieved by increasing the temperature of the medium. To create a population inversion in the optical region, we must supply energy selectively to the lower energy level, to excite atoms into the upper energy level. This excitation process is called pumping. Pumping produces a non-thermal equilibrium situation.

One of the methods used for pumping is stimulated absorption, that is the energy levels which one hopes to use for laser action are pumped by intense irradiation of the system. Now as B_{12} and B_{21} are equal once atoms are excited into upper level the probabilities of further stimulated absorption or emission are equal so that even with very intense pumping we can not get more atoms in the excited state. The best we can achieved is the equal population in both the levels, therefore, to make a laser amplifier with just two energy levels is not possible, i.e. a population inversion in a two level system can never be achieved by optical pumping.

Therefore, we must look for materials with either three or four energy levels system. This is not a problem as atomic systems generally have a large number of energy levels. A three level system is illustrated in Figure 4.6. Initially the distribution obeys the Boltzmann law. If the collection of atoms is intensely illuminated by photons, they can be excited into the level E_3 from the ground level E_1. From the level E_3, the atoms decay by non-radiative process to the level E_2 and a population inversion may be created between E_2 and E_1. Ideally the transition from level E_3 to E_2 should be very rapid to keep E_3 level almost empty. The transition from E_2 to E_1 should be slow, that is E_2 should have relatively longer lifetime. This allows a large build-up in the number of atoms in level E_2. Hence N_2 may become greater than N_1 and then population inversion will be achieved.

The level E_3 should preferably consist of a large number of closely spaced levels so that pumping uses a wide range of the radiation. This increases the pumping efficiency. Three level lasers, for example ruby, require very high pumping powers because the terminal level of the laser transition is the ground state. This means that more than half of the ground state atoms has to be pumped to the upper state to achieve population inversion.

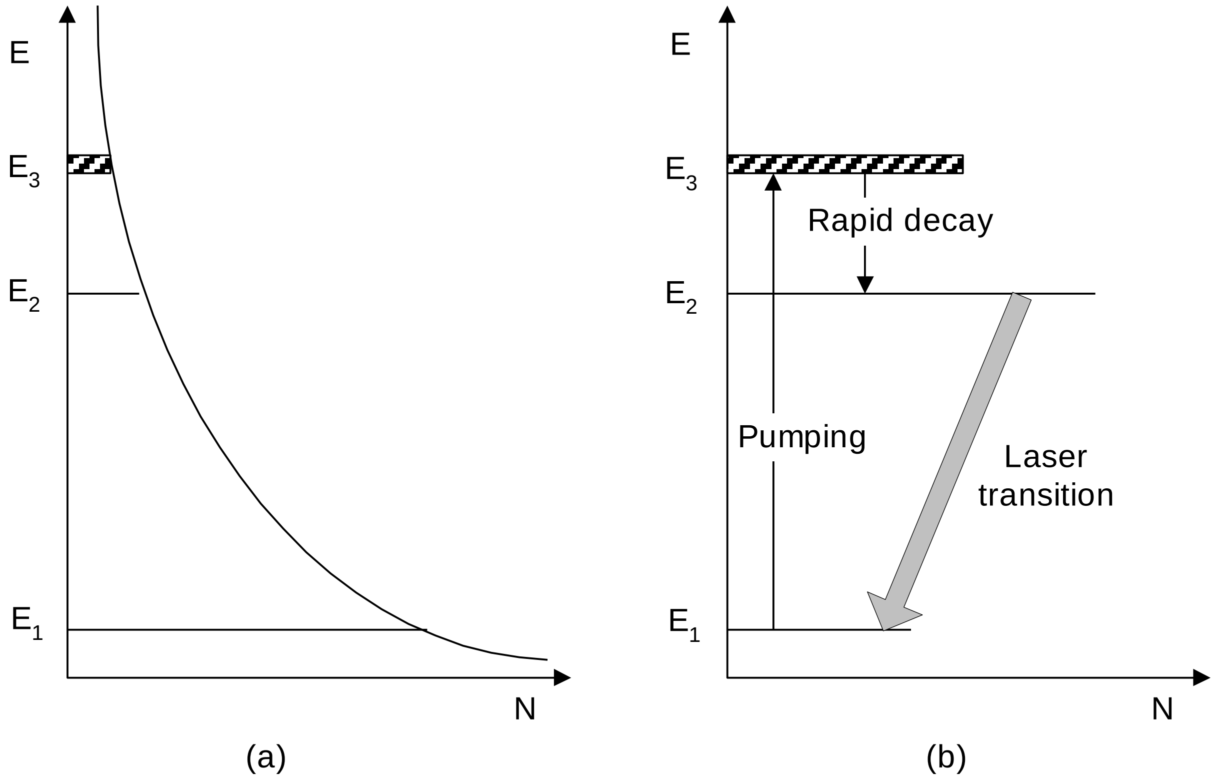


Figure 4.6 Population of energy levels by pumping in a three level system: (a) Boltzmann distribution before pumping and (b) distribution after pumping and the transitions involved.

The four-level system shown in Figure 4.7 has much lower pumping requirements. The populations of the levels E_4, E_3, and E_2 are all very close in conditions of thermal equilibrium. Thus, if atoms are pumped from the ground state to the level E_4 from which they decay very rapidly to the level E_3. The population inversion is quickly created between levels E_3 and E_2. The energy level schemes of the media used in lasers are often complex but they can usually be approximated by either three or four level schemes.

4.2.3 Optical Resonator

The population inversion is not all it takes to make a laser. A material with an inverted population can emit light in every direction. The light may be due to stimulated emission, and it may be at a single wavelength, but is not converted into a laser beam. To extract energy efficiently from a medium with a population inversion and make a laser beam, we need a resonant cavity that helps building (or amplifying) stimulated emission by feedback, i.e., reflecting some of the light back into the laser medium.

Before discussing the feedback, let us start by looking at the process of amplification.

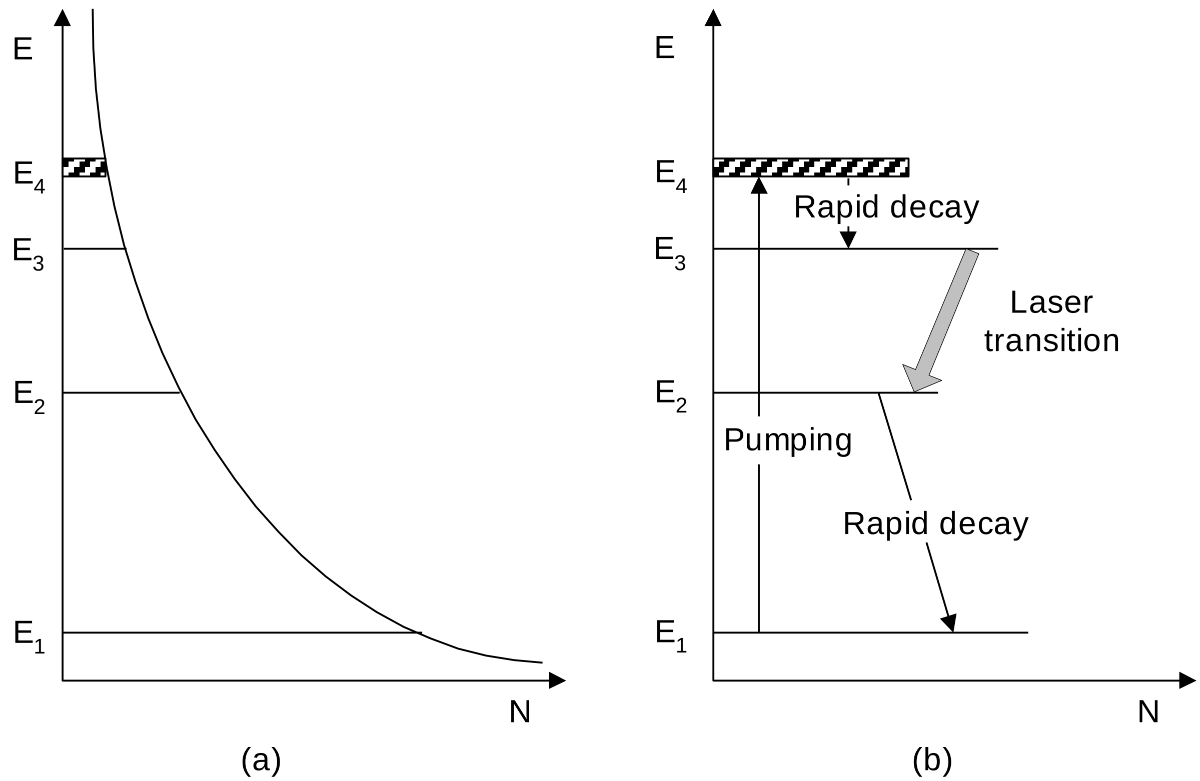


Figure 4.7 Population of the energy levels in a four level system: (a) before pumping and (b) after pumping.

4.2.3.1 Amplification of Light

In lasers, light is amplified. To understand it let us consider a collimated beam of perfectly monochromatic radiation passing through an absorbing medium (Figure 4.8). We assume for simplicity that there is only one relevant electron transition, which occurs between the energy levels E_1 and E_2. For homogeneous medium, the change in irradiance of the beam dI is proportional to the distance travelled dx and to the incident intensity I, i.e. $dI = -\alpha I.dx$. Here the constant of proportionality, α, is the absorption coefficient. The negative sign indicates the reduction in beam irradiance due to absorption. Writing this equation in differential form we have

$$\frac{dI}{dx} = -\alpha I \tag{4.13}$$

By integrating Eq. 4.10 we have

$$I = I_o e^{-\alpha x} \tag{4.14}$$

where I_o is the incident intensity.

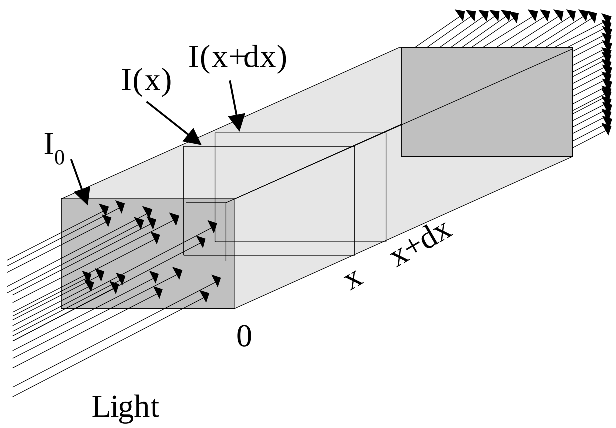


Figure 4.8 Collimated beam of light traversing an absorbing medium. The change in the irradiance across a small slab of the medium is proportional to the irradiance at the slab and to the thickness of the slab dx.

Example 4.6: *If 1% of the light incident into a medium is absorbed per millimetre, what fraction is transmitted if the medium is 0.1 m long?*

Solution: *Transmission of light passing through an absorbing medium of length x is given by*

$$I = I_0\, e^{-\alpha x}$$

For $\alpha = 0.01\ mm^{-1}$, $x = 0.1\ m = 100\ mm$, *we have*

$$I = I_0\, e^{-0.01x100} \quad = I_0 . e^{-1} = 0.368 x I_0$$

Therefore, 36.8% of the incident light will transmit through the medium and rest will be absorbed.

Let us consider the absorption coefficient in more detail. Clearly the degree of absorption of the beam will depend upon N_1, number of atoms with electrons in the lower energy level E_1, and N_2, number of atoms with electrons in upper energy level E_2. If N_2 is zero then the absorption would be maximum, while if all of the atoms are in the upper level the absorption would be zero and the probability of stimulated emission would be large.

When a beam having N number of photons per unit volume pass through a medium then net rate of loss of photons can be determined from the stimulated transitions, i.e. absorption and emission.

$$\frac{dN}{dt} = N_2\rho_\nu B_{21} - N_1\rho_\nu B_{12} \tag{4.15}$$

For the same degeneracy of both levels (i.e. $B_{12}=B_{21}$), we can write

$$\frac{dN}{dt} = (N_2 - N_1)\rho_\nu B_{21} \tag{4.16}$$

In this discussion we have ignored photons generated by spontaneous emission as these are emitted randomly in all direction and do not contribute to the initial collimated beam. Similarly we have ignored scattering losses.

Irradiance of a beam is the energy crossing a unit area in unit time. In mathematical form it can be written as

$$I=N.h\nu_{21}.u \tag{4.17}$$

where u is the velocity of light in the medium. If n is the refractive index of the medium then u=c/n, where c is the velocity of light in free space. The rate of change of irradiance when passing through a medium is given by

$$\frac{dI}{dx} = \frac{dN}{dx} h\nu_{21}.u \tag{4.18}$$

since

$\frac{u}{dx} = \frac{1}{dx/u} = \frac{1}{dt}$, we can write equation 4.18 as

$$\frac{dI}{dx} = \frac{dN}{dt} h\nu_{21} \tag{4.19}$$

Combining equations 4.16 and 4.19 we get

$$\frac{dI}{dx} = (N_2 - N_1)\rho_\nu.B_{21}.h\nu_{21} \qquad (4.20)$$

Now from equations 4.13, 4.20, ρ_ν = Nhν, and equation (4.17)

$$(N_1 - N_2).\rho_\nu.B_{21} = \alpha.\rho_\nu.u\frac{1}{h\nu_{21}} \qquad (4.21)$$

Re-arranging for absorption coefficient α, and substituting the value of u = c/n, we have

$$\alpha = (N_1 - N_2)\frac{B_{21}.h\nu_{21}.n}{c} \qquad (4.22)$$

The absorption coefficient, α, depends on the difference in the populations of the two energy levels E_1 and E_2. For a collection of atoms in thermal equilibrium, N_1 will always be greater than N_2. If, however, we can create a situation in which N_2 becomes greater than N_1 then α is negative and the quantity (-αx) in the exponent of equation 11 becomes positive. Thus the irradiance of the beam grows as it propagates through the medium in accordance with the equation:

$$I = I_o e^{kx} \qquad (4.23)$$

where k is referred to as the small signal gain coefficient and is given by

$$k = (N_2 - N_1)\frac{B_{21}.h\nu_{21}.n}{c} \qquad (4.24)$$

In this situation the medium acts as an amplifier and may be called active medium.

Example 4.7*: If the irradiance of light doubles after passing once through a laser amplifier 0.5 m long, calculate the small signal gain coefficient assuming no losses in the medium. If the increase in irradiance were only 5% what would be k?*

Solution: In an optical amplifier, the light intensity increases as

$$I = I_0 e^{kx} \Rightarrow k = \frac{1}{x}\ln\left(\frac{I}{I_0}\right)$$

(a) for I/I_0=2 and x=0.5 meter; *k=1.386 m^{-1}*

The value of signal gain coefficient is 1.386/m

(b) for I/I_0=1.05 and x=0.5 meter; *k=0.098 m^{-1}*

The value of signal gain coefficient is 0.098/m.

4.2.3.2 Optical Feedback

The gain (amplification) per unit length of most active media is so small that very little amplification of a beam of light results from a single pass. In the multiple passes through the same active media may result a large amplification.

To make an oscillator from an amplifier, it is necessary to introduce a suitable positive feedback. In the laser, the positive feedback may be obtained by placing the active material between two highly reflecting mirrors (e.g., plain-parallel mirrors as shown in Figure 4.9). The initial stimulus is provided by any spontaneous transitions between appropriate energy levels in which the emitted photon travels along the axis of the system. The signal is amplified as it passes through the medium and fedback by mirrors. Saturation is reached when the gain provided by the medium exactly matches the losses incurred during a complete round trip.

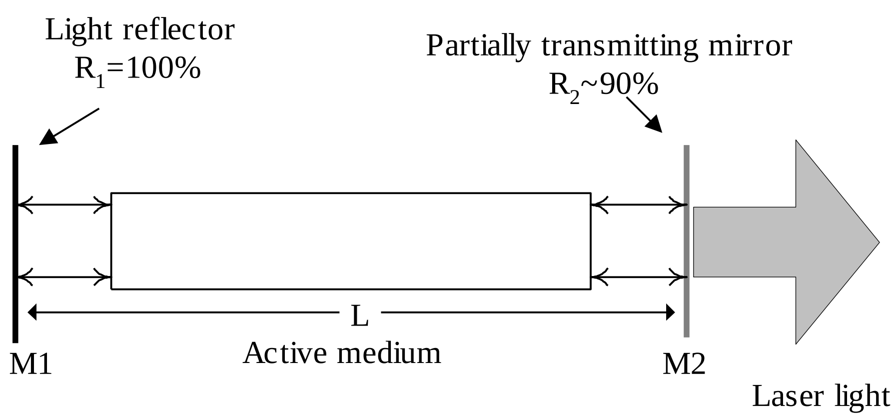


Figure 4.9. Scheme of laser oscillations.

If one of the two mirrors is made partially transparent, a useful output beam can be extracted. It is important to note that for laser, a certain threshold condition must be fulfilled. For laser, the oscillation will start when the gain of the active medium compensates the losses. Thus, while a population inversion is a necessary condition for laser action it is not the sufficient one because the minimum or threshold value of gain coefficient must be large enough to overcome the losses and sustain oscillations. The threshold gain specifies the minimum population inversion required.

The total loss of the system is due to a number of different processes, the most important ones include:

transmission at the mirrors, in fact transmission from one of the mirrors usually provides the useful output, the other mirror is made as reflective as possible to minimise losses;

absorption and scattering at mirrors;

absorption in the laser medium due to transmissions other than the desired transitions;

scattering at optical inhomogeneities in the laser medium;

diffraction losses at the mirrors.

To simplify, let us include all the losses except those due to transmission at the mirrors in a single effective loss coefficient γ, which reduces the effective gain coefficient to (k-γ). We can determine the threshold gain by considering the change in irradiance of a beam of light undergoing a round trip within the laser cavity. We assume that the laser medium fills the space between the two mirrors M_1 and M_2 which have reflectance R_1 and R_2 and a separation L. In travelling from mirror M_1 to M_2 the beam irradiance increases from I_0 to I according to equation 4.23 as,

$$I = I_0 . e^{(k-\gamma).L} \qquad (4.25)$$

After reflection at M_2, the beam irradiance will be $R_2.I_0.\exp[(k-\gamma).L]$ and after a complete round trip the final irradiance will be such that the round trip gain G is

$$G = R_1 . R_2 . e^{2(k-\gamma)L} \qquad (4.26)$$

If G is greater than unity then there will be a net amplification and the oscillation will grow; if G is less than unity, the oscillation will die out. Therefore, we can write the threshold condition as

$$G = R_1 . R_2 . e^{2(k_{th}-\gamma)L} = 1 \qquad (4.27)$$

where k_{th} is the threshold gain. It is important to realise that the threshold gain is equal to the steady-state gain in continuous output lasers. This equality is due to gain saturation, which can be explained as follows. Initially, when laser action starts the gain may be well above the threshold value. The effect of stimulated emission reduces the population of the upper level of the laser transition so that the degree of population inversion and consequently the gain will decrease. Thus the round trip gain may vary. It is only when G has been equal to unity for a period of time that the laser output power settles down to a steady-state value, that is when the gain just balances the losses in the medium. In terms of the population inversion there will be a threshold value

$$N_{th}= (N_2-N_1)_{th} \tag{4.28}$$

In the steady state situation (N_2-N_1) remains equal to N_{th} regardless of the amount by which the threshold-pumping rate is exceeded. The small signal gain required to support steady state operation depends on the laser medium through k and γ, and on the laser construction through R_1, R_2 and L.

$$k_{th} = \gamma + \frac{1}{2L}\ln\left(\frac{1}{R_1.R_2}\right) \tag{4.29}$$

The above equation shows that k can have a wide range of values, depending not only (N_2-N_1) but also on the intrinsic properties and design parameters. If k is high then it is relatively easy to achieve laser action. With low gain media, mirrors must have high reflectivities, be very clean and carefully aligned.

Example 8: *Calculate the mirror reflectance required to sustain laser oscillations in a laser which is 0.1 m long given that the small signal gain coefficient is 1* m^{-1} *(assume one mirror has 100% reflectance).*

Solution: *Gain of a laser amplifier is given by* $G=R_1R_2e^{2kL}$ *and for threshold it should be unity, therefore,*

$$R_1R_2\, e^{-2k_{th}L} =1$$

or

$$R_1R_2 = e^{-2k_{th}L}$$

It is given that $k_{th}=1\ m^{-1}$, $R_1=1$, *and* $L=0.1\ m$. *The value of* R_2 *is:*

$$R_2=e^{-0.2}=0.82$$

The reflectance of the second mirror (i.e. output coupler) is 82 %.

The efficiency of laser system is the ratio of the output light power to input pump power. It therefore depends on how effectively the pump power is converted into producing a population inversion.

Problems

4.1 The part of the electromagnetic spectrum that is of interest in the laser field starts from the submillimeter wave region and goes down in wavelength to the x-ray region. This covers the following regions in succession: (1) far infrared; (2) near infrared; (3) visible; (4) ultraviolet (uv); (5) vacuum ultraviolet (vuv); (6) soft x-ray; (7) x-ray. Draw a chart (to the scale) indicating all the regions in the units of (i) Angstroms, (ii) Hertz, (iii) electron volts (ev), and (iv) meters.

4.2 When in thermal equilibrium (at T=300 K), the ratio of the level populations N_2/N_1 for some particular pair of levels is given by 1/e. Calculate the frequency ν for this transition. In what region of the e.m. spectrum does this frequency fall?

4.3 The value of signal gain coefficient of a certain laser amplifier is 0.29/m. What's physical meanings of it?

4.4 Calculate the small signal gain co-efficient required to sustain laser oscillations in a medium which is 0.15 meter long, given that the effective loss coefficient is 0.15 m^{-1}, the reflectivity of rear mirror is 100% and that of output mirror is 80%.

4.5 For a two level system, suppose $N_2 = 10N_1$, where N_1 and N_2 denote the number of atoms (per unit volume) in level 1 and level 2 respectively. The transition between these two levels corresponds to the frequency of $5x10^4$ Hz. Calculate the ratio of spontaneous emission and stimulated emission. Also calculate the temperature required to produce this population inversion.

4.6 In a two levels system, the levels are separated by an energy E_2 - E_1. Find the ratio of population inversions at room temperature if transition between levels occur at (i) λ = 0.55 μm (ii) λ = 3 μm

4.7 When two quantum energy levels, E_1 and E_2 of an atom are separated by an energy gap $\Delta E = E_2 - E_1$, and a large number of such atoms are in thermal equilibrium at temperature T, then the relative number of atoms N_1 and N_2 in the two energy levels are given by the Boltzmann ratio $N_2/N_1 = e^{-\Delta E/kT}$. Evaluate this ratio for the following cases:

(a) transition occurs at the frequency of 3100 MHz, and the temperature is 300 K. What is the fractional population difference $(N_2-N_1)/ N_1$?

(b) Consider the same situation, except that ν = 9000 MHz and the temperature is 4 K. What is $(N_2-N_1)/ N_1$?

(c) Calculate the Boltzmann ratio N_2/ N_1 for λ = 5500 A^o, and T = 300 K.

(d) What temperature required to make (N_2 equal to 8 % of N_1 in part (c) ?

4.8 The irradiance of light becomes half after passing through a laser media. Calculate the small gain coefficient, explain for obtaining negative answer in part.

4.9 If the irradiance of light becomes double after passing once through a laser amplifier 0.5 m long., assuming no losses in the media:

(a) Calculate the small signal gain co-efficient.

(b) Calculate the small signal gain co-efficient, if the length of the media is doubled.

(c) By doubling the length of the media, is the small signal gain co-efficient doubled or not? Give comprehensive comments in either case.

Books for further reading

J. Wilson and J.F.B. Hawkes, *Optoelectronics: An Introduction*, (Prentice-Hall International, London, 1983)

O. Svelto, *Principles of Lasers*, (Plenum Press, New York, 1989)

P. W. Milonni and J. H. Eberly, *Lasers*, (John Wiley & Sons, New York, 1991).

A. E. Seigman, *Lasers and Masers*, (1971).

D. C. O’Shea, W. R. Callen, and W. T. Rhodes, *Introduction to Lasers and Their Applications,* (Addison-wesley, London, 1978)

Unit-5

Pumping Processes

Objective

The pumping processes are one of the essentials of a laser system. In this unit a couple of these processes are discussed. Attainment of population inversion through rate equation treatment has been described for three and four-level systems. Gain saturation behavior is also discussed. The student is required to understand the rate equation analysis, such that pumping rate and gain of a system can be calculated.

5.1 Introduction

The process by which atoms are raised from lower energy level to higher energy level is called pumping process. The most common pumping schemes are: optical or electrical. In Optical pumping the light from a powerful source is absorbed by the active material and the atoms are thereby pumped into the upper energy level. This method is particularly suited to solid-state (e.g., ruby or neodymium) and liquid dye lasers. Energy bands in solids and in liquids absorb a sizeable fraction of the broadband light emitted by the flash lamp.

Electrical pumping is particularly suited to gas and semiconductor lasers. In gas lasers, sufficiently intense electrical discharge is responsible for pumping. In gases, optical pumping is not preferred because of small widths of absorption line and usually broadband emission of the flash lamp. In the case of semiconductor lasers, optical pumping could be used but electrical pumping is very convenient, since it is very easy to pass a current through the semiconductor.

5.2 Pumping Mechanisms

There are quite a number of ways for pumping the active medium but most common are the optical and electrical. Following is the list of some common pumping processes; however, only the optical and electrical pumping with a little more details will be discussed.

5.2.1 Chemical Pumping

Energy from an exothermic chemical reaction in a gas can excite a laser medium. In the simplest way, two species react to produce a third species, which carries at least part of the exothermic energy of reaction as vibrational energy. If the reaction is rapid, the result can be a population inversion of the excited species produced by the reaction. The most common chemical lasers are the family of hydrogen halides. The chemical reaction can be shown as

$$A + B \rightarrow (AB)^{*}$$

where (AB)* is the resultant molecule in the excited vibrational state.

5.2.2 Gas Dynamic Pumping

Another way to produce a population inversion is to expand a hot, high-pressure gas into a vacuum. The gas dynamics expansion cools the gas, but this cooling is not same for all the energy levels. At high temperature sufficient number of molecules can reach to the higher energy level, if enough of the molecules in this energy state remain there even after expansion of the gas and the molecules in the lower energy state relax immediately after expansion, a population inversion can be achieved. The inversion lasts only until the gas can start approaching equilibrium. Within a gas dynamic laser, the population inversion exists only for a short distance downstream of the expansion nozzle. In practice, only one type of carbon dioxide laser, called gas dynamic CO_2 laser, is pumped with this technique, and produces powers of tens or hundreds of kilowatts.

5.2.3 Nuclear Pumping

Laboratory demonstrations have shown that gas laser can be excited by transferring energy from the products of nuclear reaction to the excited species. In experimental devices, ions produce by the fission reactions then transfers their energy to the laser gas, producing population inversion. Research on the concept revived in the mid-1980s by the Strategic Defense Initiative, but little progress has been made to the practical devices.

Now we will discuss the two popular pumping schemes in some details.

5.2.4 Optical Pumping

Optical pumping can be subdivided into two branches, one is laser pumping and the other is flashlamp pumping. The former is less common as compare to the latter.

5.2.4.1 Laser Pumping

It is a special kind of optical pumping where a laser beam is used to pump another laser. The directional properties of a laser beam make it very convenient for pumping another laser. The monochromaticity of the pump laser means that laser pumping is not limited to solid-state and liquid lasers but can also be applied to gas lasers. In the latter case, the line emitted by the pumping laser must coincide with an absorption line of the laser to be pumped. This situation applies in most far infrared gas lasers, e.g. methyl alcohol and CH_3OH lasers, which are usually pumped by a suitable line of a CO_2 laser.

In laser pumping the wavelength of the pumped laser is higher than the pumping laser (i.e., $\lambda_{emission} > \lambda_{absorbed}$) but the pumping efficiency is significantly higher as compare to the other scheme.

5.2.4.2 Flashlamp Pumping

In the case of optical pumping, the light from a powerful incoherent lamp is focussed, by a suitable optical system, to the active material. Figure 5.1 shows three most commonly used pumping configurations. In all three cases the active material is taken to be the form of a cylindrical rod, as usually applied in the practical systems. Its diameter ranges from a few millimeters to a few centimeters and lengths from a few millimeters up to a few tens of centimeters. This can be operated in a pulsed or continuous wave regime depending on whether the lamp is pulsed or continuous. In Figure 5.1a the lamp has a helical shape, and the

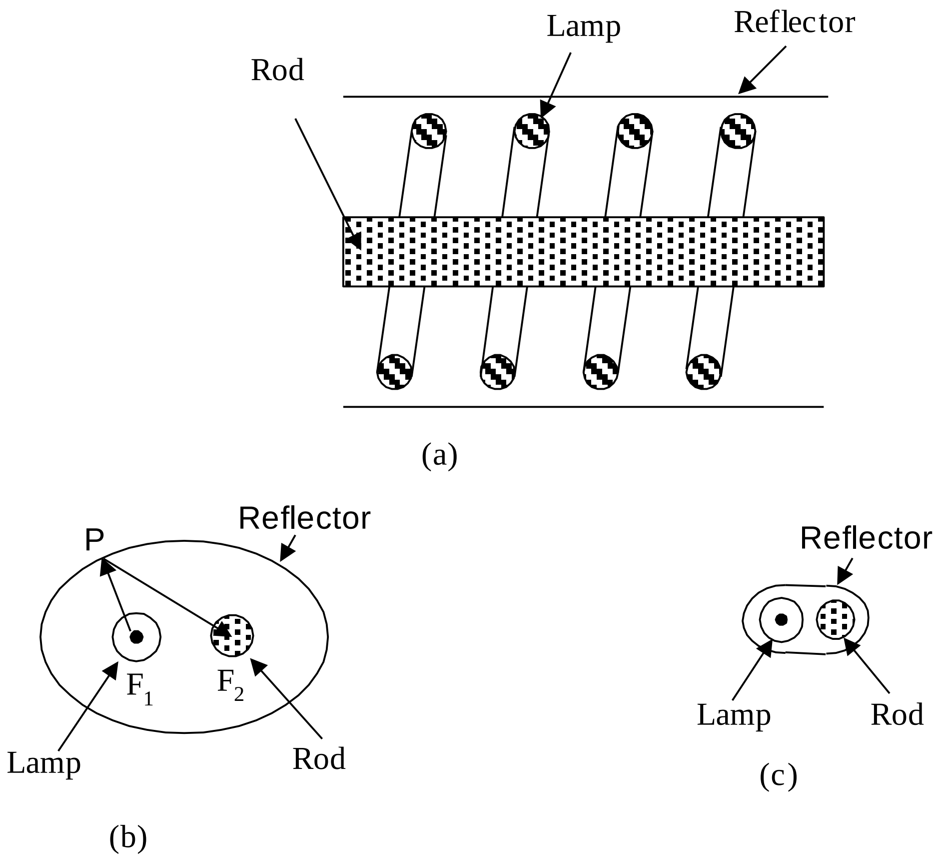


Figure 5.1. Most commonly used optical pumping systems. (a) The lamp is helical shape, light reaches to the rod directly or after reflection from cylinderical surface. (b) The lamp is a linear cylinderical shape, it is placed at focus of the elliptical reflecting cyilider and rod is at the other focus. (c) The and linear lamp are placed as closed as possible and are surrounded by a clo coupled cylinderical reflector.

light reaches the active material either directly or after reflection at the specular cylindrical surface. This was the arrangement used for ruby laser, and it is still in use for some type of pulsed lasers. In Figure 5.1b the lamp is in the form of a cylinder (linear lamp) having radius and length roughly equal to those of the active rod. The lamp is placed along one of the two focal axes, F_1, of a specularly reflecting elliptical cylinder and the laser rod is placed along the second focal axis F_2. A well-known property of an ellipse is that all the rays emitting from one focus converge at the second focus after reflection from the surface of the elliptical cylinder. Therefore, a large fraction of the light emitted by the lamp illuminate the active material. Figure 5.1c shows an example of what is known as a close-coupled configuration. The rod and the linear lamp are placed as closed as possible and are surrounded by a close-coupled cylindrical reflector. The efficiency for close-coupled configurations is usually not much lower than for the elliptical cylinders.

Pumping Efficiency

For continuous wave lasers, pumping efficiency η_p can be defined as the ratio between the minimum power P_m that would be needed to produce a given pumping rate and the electrical pump power P actually delivered to the lamp. The minimum power can be written as

$$P_m = \left(\frac{dN_2}{dt}\right).V.h\nu_p = W_p N_1.V.h\nu_p \qquad (5.1)$$

where W_p is the pumping rate, V is the volume of the active material and ν_p is the frequency difference between the ground level and the upper laser level. Therefore, the efficiency can be written as

$$\eta_p = \frac{W_p.N_1.V.h\nu_p}{P}\ . \qquad (5.2)$$

The pumping efficiency η_p can be written as the product of four terms corresponding to: (1) the emission of radiation from the lamp, (2) the transfer of this radiation to the active rod, (3) the absorption in the rod, and (4) the transfer of the absorbed energy to the upper laser level. Mathematically,

$$\eta_p = \eta_r.\eta_t.\eta_a.\eta_{pq} \qquad (5.3)$$

where η_r is the lamp radiative efficiency, i.e., the efficiency of conversion from electrical input to light output in the wavelength range corresponding to the pump bands of the laser medium. η_t is transfer efficiency, which can be defined as the ratio of the pump power actually entering the rod to that emitted by the lamp in the useful pump range. η_a is the absorption efficiency, i.e., the fraction of light entering the rod that is actually absorbed by the material. η_{pq} is the power quantum efficiency, i.e., the fraction of the absorbed power that leads to population of the upper laser level. Typical values of the pumping efficiency for Nd:YAG laser ranges from 0.1 to 1 percent.

5.2.5 Electrical Pumping

Electrical pumping is used for gas and semiconductor lasers. In this section we will limit our discussion for gas lasers only, pumping for semiconductor lasers may be discussed in the description of diode lasers.

Electrical pumping of gas lasers is achieved by allowing a current to pass through the gas mixture. For this purpose, a high dc voltage is applied to break down the gases so it will conduct the electricity. Generally, there are two types of electrical pumping (i) longitudinal discharge pumping (ii) transverse discharge pumping as shown in Figure 5.2. In longitudinal discharge lasers, the electrodes often have an annular structure with the cathode surface usually much larger than that of the anode, which helps in reducing the degradation due to ion collisions. In a transverse discharge, the electrodes extend over the whole length of the laser material. Various electrode structures are used depending upon the types of laser involved. Usually longitudinal discharge arrangements are used for continuous wave (cw) lasers, while transverse discharges are used for both types of lasers i.e., cw and pulsed. The longitudinal discharge, when confined in a glass tube (or any other dielectric tube) provides a more uniform and stable pumping configuration.

In an electrical discharge, ions and free electrons are produced. From the applied electric field, they acquire kinetic energy and are able to excite a neutral atom or molecule by collision. The positive ion, due to their heavier mass, are accelerated to lower velocities than the electrons and therefore do not play any significant role in the excitation process. Electrical pumping of a gas occurs via one, or both, of the following processes:

(i) For a gas consisting of only one species, the excitation is only produced by electron impact, i.e., according to the process

$$e + x \rightarrow x^* + e$$

where x and x* represent the atom in the ground and excited states, respectively.

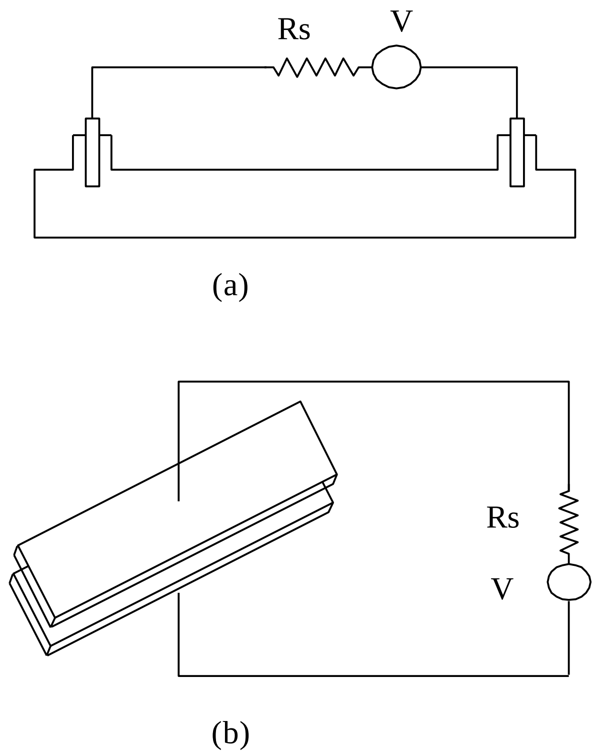


Figure 5.2 Frequently used pumping configuration for gas-discharge excitation; longitudinal discharge, and (b) transverse discharge.

(ii) For a gas consisting of two species (say A and B), excitation can also occur as a result of collisions between atoms of different species through a process known as resonant energy transfer as shown in Figure 5.3. Let us assume that species B is in ground state and species A is in the excited state brought by the electron impact. It is also assume that the energy difference ΔE between the two transitions is less than kT (which is the thermal energy at room temperature). In this case, there is an appreciable probability that, after collision, species A will be found in its ground state and species B in its excited state. This process can be denoted by

$$A^{*}+B \rightarrow A+B^{*} \pm \Delta E \qquad (5.4)$$

where the energy difference ΔE will be added or subtracted from the translational energy, depending on its sign. This process provides a particularly attractive way of pumping species B, if the upper state of A is metastable (forbidden transition). In this case, once A is excited to its upper level, it will remain there for a long time, thus constituting an energy reservoir for excitation of the B species.

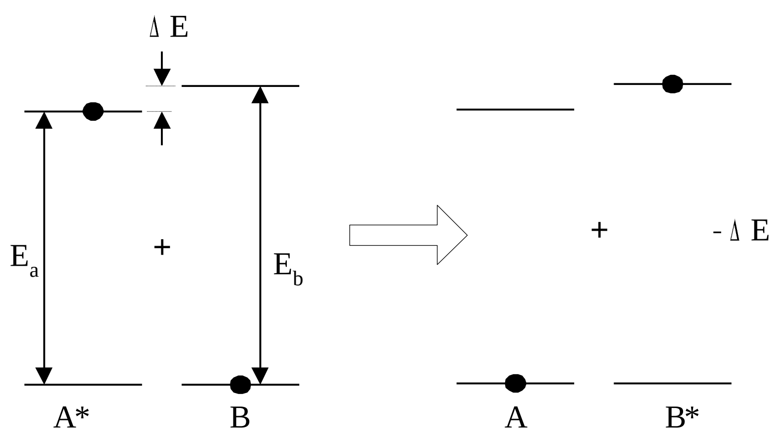


Figure 5.3 Near resonant energy transfer from A* to B

Helium-neon and carbon dioxide lasers are good examples of this kind of pumping process. In helium neon lasers, helium has a metastable state close to the upper laser level of neon and therefore, it plays the role of species A in the mixture. In carbon dioxide lasers, a mixture of CO_2, N_2 (nitrogen) and helium gases is used, here nitrogen molecule has a metastable level close to the upper laser level in the CO_2 molecule. Resonant energy transfer from N_2 to CO_2 increases the pumping efficiency.

5.2.5.1 Physical Characteristics of Discharges

In electrical discharge electrons play the main role. They acquire energy from the applied electric field and lose or exchange energy through three processes:

(1) Inelastic collisions with the atoms (or molecules) of the gas mixture, which either raise the atom to one of its excited states or ionize it. These electron-impact excitation or ionization phenomena are perhaps the most important processes for laser pumping.

(2) Elastic collisions with the atoms. If we assume these atoms to be at rest before the collision (the mean atomic velocity is much smaller than that of the electrons), the electron will lose energy upon collision. It can be shown by a straightforward analysis of the elastic collision that, if the direction of scattered electron is random, the

electron loses on average a fraction 2(m/M) of its energy, where m is the electron mass and M is the mass of the atom. This loss is very small since m/M is small, e.g., $m/M = 1.3 \times 10^{-5}$ for Ar atoms.

(3) Electron-electron collisions. For less weakly ionized gas, the frequency of such collisions is usually high since both particles are charged and exert forces on one another at considerable distance. Moreover, since both colliding particles have the same mass, the energy exchange in the collision is considerable. As a result of the collision phenomena mentioned above, the electron "gas" in the plasma acquires a distribution of velocities and hence of energies.

5.3 Steady State Laser Pumping and Population Inversion

One of the most common applications for rate equations is the analysis of laser pumping. In this section, therefore, we will develop and solve the rate equations to analyze steady-state laser pumping in simplified four-level and three-level laser systems.

5.3.1 Elementary Four-Level Laser System

For the analysis of steady-state laser pumping and population, we consider a neodymium laser (Nd:YAG or Nd:Glass), which is a typical example of four-level laser system. The complicated energy levels of Nd^{+3} have been simplified into the idealized four-level laser system shown in Figure 5.4. This four-level model will in fact provide a simple but surprisingly accurate analytical model for many laser systems. In this model level 4 represents the combination of all the levels lying above the upper laser level in the real atomic system. It is desirable that many of these levels be in fact broad absorption bands, so that the optical pumping into these levels by a broadband pump lamp can be very efficient. Level 3 represents the upper laser level, usually a fairly sharp and long-lived level, with a large gap below it. Level 2 then represents the lower laser level and level 1 is the lowest or ground level. Other low-lying levels that may be present inside the material, both above and below the lower laser level, are ignored in the model because they play no role in the laser action. They act only as temporary way stations through which atoms may pass in relaxing from the other levels to the ground level E_1.

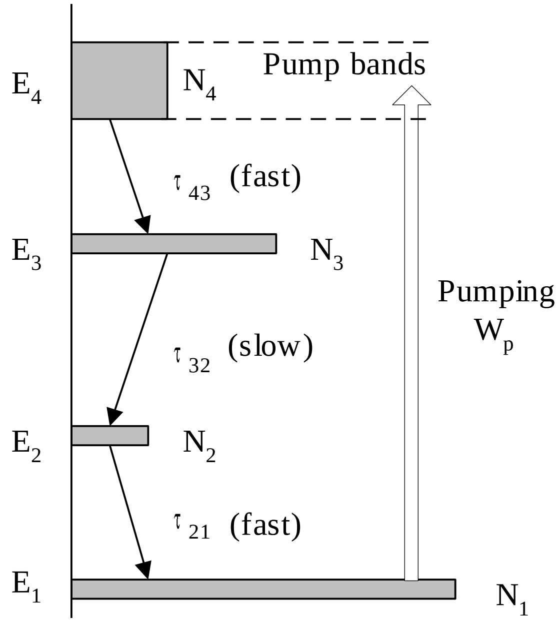


Figure 5.4 An idealized four-level pumping scheme suitabl for Nd:YAG laser system.

5.3.1.1 Four-Level Pumping Analysis

To analyze this system we will write down the relevant rate equations, using the model shown in Figure 5.4 and then solve for their steady-state solutions, making reasonable approximations during the analysis. Since the condition ($h\nu/kT$)>>1 is usually very well satisfied for all transitions in a visible laser system, we will write all of the following rate equations using the "optical frequency approximation". In this approximation, the thermally stimulated terms in the relaxation rates, either upward or downward, are totally negligible compared to the spontaneous emission rates, this is due to the fact that Boltzmann ratio at optical frequencies is always very small, i.e., on the order of 10^{-36} at room temperature.

We begin by assuming that the laser pumping process produces a stimulated transition probabilities $W_{14} = W_{41} = W_p$ between levels 1 and 4. The rate equation for level 4 in the optical approximation is then

$$\frac{dN_4}{dt} = W_p.(N_1 - N_4) - \frac{N_4}{\tau_4} \tag{5.5}$$

where the lifetime τ_4 is give by

$$\frac{1}{\tau_4} = \frac{1}{\tau_{43}} + \frac{1}{\tau_{42}} + \frac{1}{\tau_{41}} \qquad (5.6)$$

is the total lifetime for decay of level 4 to all lower levels. The steady-state population of level 4, when $dN_4/dt = 0$, is then given by

$$N_4 = \frac{W_p.\tau_4}{1 + W_p.\tau_4}.N_1 \qquad (5.7)$$

The normalized pumping rate $W_p\tau_4$, which will appear in many of the following expressions, will in fact have a value much less than unity in many (though not all) practical laser systems.

Direct pumping up from the ground level into the upper laser level 3 in this model can very often be assumed negligible. This may be due to either of the following two reasons:

(1) The $1 \rightarrow 3$ transition have a weaker absorption cross section than the $1 \rightarrow 4$ transitions, where level 4 corresponds to a group of levels.

(2) The transition is much narrower than the strong absorption bands from the ground level to the groups of levels that make up level 4. The rate equations for the two levels N_3 and N_2 are then:

$$\frac{dN_3}{dt} = \frac{N_4}{\tau_{43}} - \frac{N_3}{\tau_3} \qquad (5.8)$$

and

$$\frac{dN_2}{dt} = \frac{N_4}{\tau_{42}} + \frac{N_3}{\tau_{32}} + \frac{N_4}{\tau_{21}} \qquad (5.9)$$

Here, τ_3 is the total lifetime of the level 3, τ_{32} and τ_{21} are average time of transitions from level $3 \rightarrow 2$ and $2 \rightarrow 1$, respectively. Equation (5.8) then gives at steady state (d/dt = 0)

$$N_3 = \frac{\tau_3}{\tau_{43}}.N_4 \qquad (5.10)$$

In a good laser system, the $4 \rightarrow 3$ relaxation rate will be very fast, but the upper laser level 3 will have a long lifetime by comparison, so that $\tau_3 >> \tau_{43}$ and hence $N_3 >> N_4$.

Combining equation (5.9) and (5.10) then gives the result

$$N_2 = \left(\frac{\tau_{21}}{\tau_{32}} + \frac{\tau_{43}.\tau_{21}}{\tau_{42}.\tau_3} \right) . N_3 = \beta . N_3 \tag{5.11}$$

where the parameter β is defined to be

$$\beta = \frac{\tau_{21}}{\tau_{32}} + \frac{\tau_{43}.\tau_{21}}{\tau_{42}.\tau_3} \tag{5.12}$$

This parameter β thus depends only on relaxation-time ratios, not absolute values. If this quantity is less than unity, the steady-state result will be $N_2 < N_3$, which means there will be the desired population inversion on the $3 \rightarrow 2$ transition.

In a good laser system, the upper levels E_4 will relax primarily into the upper laser level E_3, so that $\tau_{42} \approx \infty$. In this case $\beta \approx \tau_{21}/\tau_{31}$, and the condition for population inversion becomes simply

$$\beta = \frac{N_2}{N_3} \approx \frac{\tau_{21}}{\tau_{32}} << 1 \tag{5.13}$$

$$\Rightarrow N_2 << N_3 .$$

In other words, to have good inversion on the $3 \rightarrow 2$ transition, atoms should relax out of the lower laser level E_2 down into lower levels much faster than atoms relax into E_2 from above. Even if level 4 does not relax only into level 3, if the upper laser level has a long lifetime (both τ_{32} and τ_3 are long) and the lower laser level has a short lifetime (τ_{21} is short), then population inversion on the $3 \rightarrow 2$ transition is virtually certain.

Whether this population inversion will be large enough to give sufficient gain to achieve laser action in a practical cavity is another mater. Nonetheless, these conditions are met and laser action can be produced on many transitions in many real laser systems.

5.3.1.2 Fluorescence Quantum Efficiency

Another dimensionless parameter often used in evaluating laser materials is the fluorescent quantum efficiency η, defined as the number of fluorescent photons spontaneously emitted on the laser transition divided by the number of pump photons absorbed on the pump transitions when the laser material is below threshold. For the four-level system this quantum efficiency is given by

$$\eta = \frac{\tau_4}{\tau_{43}} \times \frac{\tau_3}{\tau_{rad}} \tag{5.14}$$

$\tau_{rad} = \tau_{rad}(3 \rightarrow 2)$ is the radiative life time for $3 \rightarrow 2$ transition. The first ratio in this expression tells what fraction of the total atoms excited to level 4 relax directly into the upper laser level 3, rather than by passing 3 and dropping to lower levels. The second ratio tells what fraction of the total decay out of level 3 is purely radiative decay to level 2.

5.3.1.3 Four-Level Population Inversion

Using the parameters β and η plus the conservation of atoms condition that is

$$N = N_1 + N_2 + N_3 + N_4 \tag{5.15}$$

We can solve for the population inversion N_3-N_2 for the four-level system by substituting values of N's from equation (5.7), (5.10), and (5.11) into equation (5.15) and taking the ratio with N_3-N_2, we get

$$\frac{N_3 - N_2}{N} = \frac{(1-\beta)\eta.W_p.\tau_{rad}}{1 + \left[1 + \beta + 2\tau_{43}/\tau_{rad}\right]\eta.W_p.\tau_{rad}} \tag{5.16}$$

where $\tau_{rad} = \tau_{rad}(3 \rightarrow 2)$ is the radiative decay time of the laser transition itself. In a good laser material the lifetime τ_{43} from the pump level into the upper laser level will be short compared to the radiative decay time, τ_{rad}, and this expression can then be simplified into

$$\frac{N_3 - N_2}{N} \approx \frac{(1-\beta)\eta.W_p.\tau_{rad}}{1 + [1 + \beta]\eta.W_p.\tau_{rad}} \approx \frac{W_p.\tau_{rad}}{1 + W_p.\tau_{rad}} \qquad \text{if } \beta \rightarrow 0 \tag{5.17}$$

The optimum situation is obviously $\beta \approx \tau_{21}/\tau_{32} \to 0$.

Figure 5.5 shows a plot of the inversion $N_3 - N_2$ on the four-level laser transition verses the

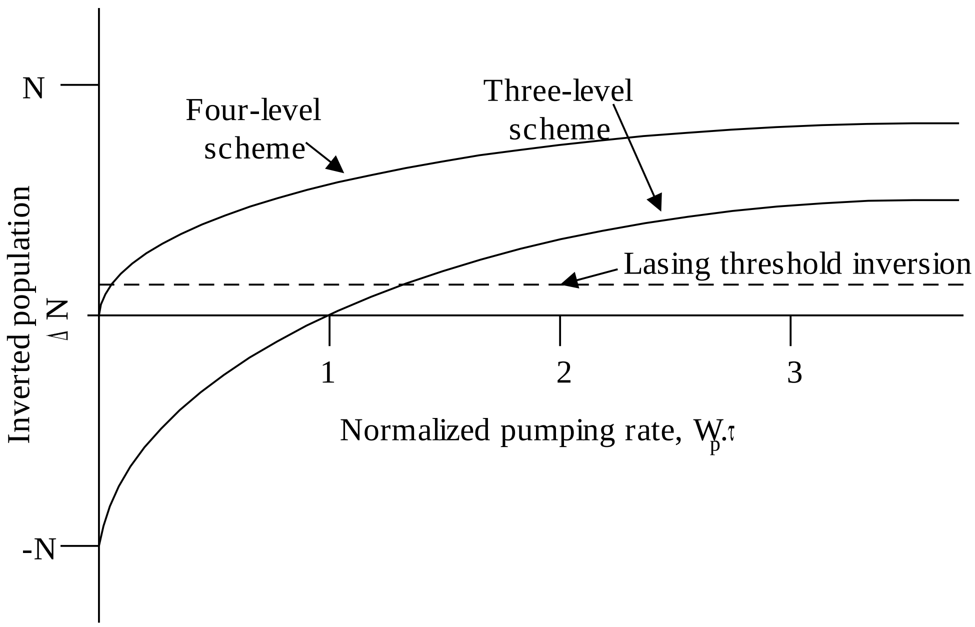


Figure 5.5 Laser population inversion verses normalized pumping rate for idealized four-level and three-level laser systems.

normalized pumping rate $W_p.\tau$, assuming $\beta=0$. For a four-level system, the population inversion on the $3 \to 2$ transition increases linearly with the pumping rate W_p at lower pump levels, but then approaches a limiting value for $W_p.\tau >> 1$ as the ground state E_1 is depleted and a large fraction of the atoms are lifted into the upper laser level.

This four-level pumping model provides an excellent analytical model for understanding the behavior of a large number of real laser systems.

5.3.2 Three-Level Laser System

Figure 5.6 illustrates how a three-level laser system can be similarly employed as a model for the real energy levels of the famous ruby laser. Similar to the model of four-level system for Nd:YAG, a model for three-level system can be formed.

A three-level system differs from the four-level system in that the lower laser level is the ground level E_1. This is a serious disadvantage, since more than half the atoms initially in the

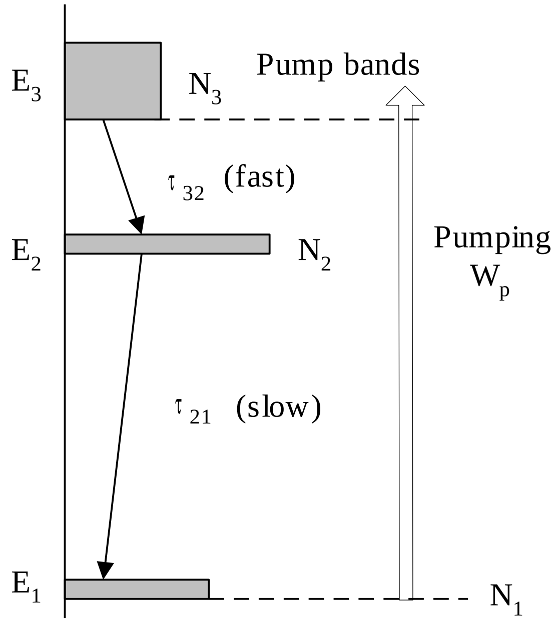


Figure 5.6 An idealized three-level model for laser pumping resembles closely with ruby laser system.

ground state must be pumped through the upper pumping level E_3 into the upper laser level E_2 before any inversion at all is obtained on the $2 \rightarrow 1$ transition. Three-level lasers are, therefore, usually not as efficient as four-level lasers. One reason for analyzing the three-level system, in addition to the general background knowledge, is that the 694 nm ruby laser, the first ever to be operated and still a useful solid state laser, is a nearly ideal three-level laser system.

Suppose the pumping process in a three-level system produces a stimulated transition probability $W_{13}=W_{31}=W_p$. Then the rate equations for the two upper levels are

$$\frac{dN_3}{dt} = W_p.(N_1 - N_3) - \frac{N_3}{\tau_3} \tag{5.18}$$

and

$$\frac{dN_2}{dt} = \frac{N_3}{\tau_{32}} - \frac{N_2}{\tau_{21}} \tag{5.19}$$

There is also the usual conservation equation $N_1 + N_2 + N_3 = N$, and it is again useful to define fluorescence quantum efficiency given by

$$\eta = \frac{\tau_3}{\tau_{32}} \times \frac{\tau_2}{\tau_{rad}(2 \to 1)} \tag{5.20}$$

The important relaxation-time ratio in this model is given by

$$\beta = \frac{N_3}{N_2} = \frac{\tau_{32}}{\tau_{21}} \tag{5.21}$$

The steady-state population difference on the $2 \to 1$ transition can be found to be

$$\frac{N_2 - N_1}{N} = \frac{(1-\beta).\eta.W_p.\tau_{rad} - 1}{(1+2\beta).\eta.W_p.\tau_{rad} + 1} \tag{5.22}$$

Inversion in the three-level system can be obtained only if $\beta < 1$, and even then inversion can occur only when the pumping rate exceeds a threshold value given by

$$W_p \tau_{rad} \geq \frac{1}{\eta.(1-\beta)} \tag{5.23}$$

The optimum situation occurs when the relaxation from the energy level 3 into the upper laser level 2 is very fast, so that $\beta \to 0$, and when the relaxation from the upper laser level 2 down to the ground level 1 is purely radiative, so that $\eta \to 1$. The inversion verses normalized pumping strength then reduces to

$$\frac{N_2 - N_1}{N} \approx \frac{W_p.\tau_{rad} - 1}{W_p.\tau_{rad} + 1} \qquad \text{If } \eta \to 1 \text{ and } \beta \to 0 \tag{5.24}$$

The significant differences in inversion verses pumping for a three-level and a four-level system are illustrated in the Figure 5.5. A four-level laser system should have a much lower pumping threshold than a three-level laser system.

***Example 1**. It is possible to have a three-level laser system which is pumped on the $1 \rightarrow 3$ transition and in which cw laser action takes place on the $3 \rightarrow 2$ rather than the $2 \rightarrow 1$ transition (no such real system is known). Suppose that level 3 in such a system is long lived with lifetime τ_3; level 2 has a short relaxation time to the ground state; and the system is pumped with transition probability W_p on the $1 \rightarrow 3$ transition.*

Derive the expression for population inversion between level 3 and 2.

***Solution**: Consider a three-level system as shown in Figure 5.7. The rate equations for the population of upper laser level E_3, and lower laser level E_2 are:*

Suppose the pumping process in a three-level system produces a stimulated transition probability $W_{13} = W_{31} = W_p$. Then the rate equations for the two upper levels are

$$\frac{dN_3}{dt} = W_p (N_1 - N_3) - \frac{N_3}{\tau_3} \qquad (5.a)$$

and

$$\frac{dN_2}{dt} = \frac{N_3}{\tau_{32}} - \frac{N_2}{\tau_{21}} \qquad (5.b)$$

In steady-state condition the rate of change of atoms in the two energy levels will be zero, thus

$$\frac{dN_3}{dt} = W_p (N_1 - N_3) - \frac{N_3}{\tau_3} = 0$$
$$\Rightarrow N_3 = \left(\frac{W_p . \tau_3}{1 + W_p . \tau_3} \right) . N_1 \qquad (5.c)$$

similarly

and
$$\frac{dN_2}{dt} = \frac{N_3}{\tau_{32}} - \frac{N_2}{\tau_{21}} = 0$$
$$\Rightarrow N_2 = \frac{\tau_{21}}{\tau_{32}} . N_3 \qquad (5.d)$$

From equation (5.c) and (5.d), we have

$$N_3 - N_2 = N_3 - \frac{\tau_{21}}{\tau_{32}} N_3 = \left(1 - \frac{\tau_{21}}{\tau_{32}}\right) \cdot \left[\frac{W_p.\tau_3}{1 + W_p.\tau_3}\right] . N_1 \qquad (5.e)$$

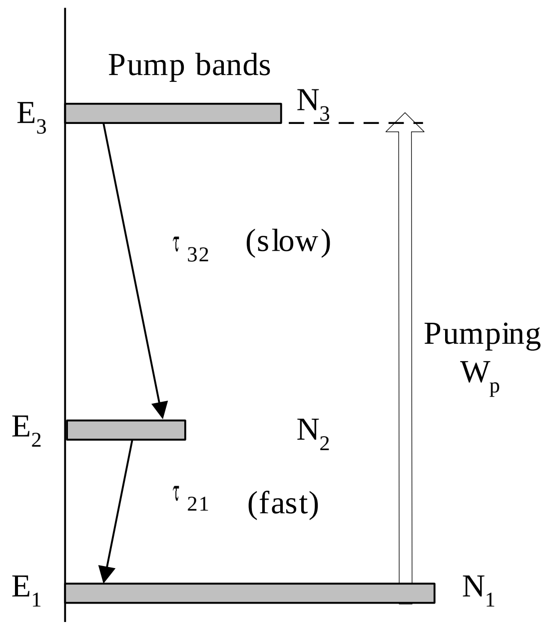


Figure 5.7 Three-level pumping scheme in which population inversion can be build-up between level 3 and level 2

The conservation of total number of atoms gives that

$$N_1 + N_2 + N_3 = N \qquad (5.f)$$

Therefore, we have

$$N = N_1 + \frac{\tau_{21}}{\tau_{32}} \left[\frac{W_p.\tau_3}{1 + W_p.\tau_3}\right] . N_1 + \left[\frac{W_p.\tau_3}{1 + W_p.\tau_3}\right] . N_1$$

$$N = \left[1 + \left(1 + \frac{\tau_{21}}{\tau_{32}}\right) \cdot \left(\frac{W_p.\tau_3}{1 + W_p.\tau_3}\right)\right] . N_1 \qquad (5.g)$$

Now from equations (5.e) and (5.g) we get

$$\frac{N_3 - N_2}{N} = \frac{\left(1 - \frac{\tau_{21}}{\tau_{32}}\right) \cdot \left[\frac{W_p . \tau_3}{1 + W_p . \tau_3}\right]}{1 + \left(1 + \frac{\tau_{21}}{\tau_{32}}\right) \cdot \left[\frac{W_p . \tau_3}{1 + W_p . \tau_3}\right]} \qquad (5.h)$$

let $\beta = \frac{\tau_{21}}{\tau_{32}}$ *, then above equation can be written as*

$$\frac{N_3 - N_2}{N} = \frac{(1 - \beta) . W_P . \tau_3}{(2 + \beta) . W_P . \tau_3 + 1} \qquad (5.j)$$

In this kind of three-level laser, the fluorescence efficiency, η*, can be defined as*

$\eta = \frac{\tau_3}{\tau_{rad}}$ *, the last equation can be written as*

$$\frac{N_3 - N_2}{N} = \frac{(1 - \beta) . W_P . \eta . \tau_{rad}}{(2 + \beta) . W_P . \eta . \tau_{rad} + 1} \qquad (5.k)$$

Population inversion in the given three-level laser system will be obtained if $\beta<1$ *and* $\eta\sim1$*. Therefore, substituting* $\beta \to 0$ *and* $\eta \to 1$ *in the above equation.*

$$\frac{N_3 - N_2}{N} = \frac{W_p . \tau_{rad}}{2 . W_p . \tau_{rad} + 1} \qquad (5.l)$$

The above equation shows that population inversion is 0 at zero pumping rate and inversion can be achieved with a small pumping rate, that is, lasing threshold is not as high as in the case of ruby laser, but unfortunately, no such laser exist.

Example 2. *Consider a five-level atomic system in which the optical frequency approximation applies between all levels. Assume that this system is pumped on the* $1 \to 5$ *transition with a pumping transition probability* $W_{15} = W_{51} = W_p$*, and that each upper level in the system relaxes only into the level immediately beneath it. Evaluate the population difference on the* $3 \to 2$ *transition as a function of* W_p*.*

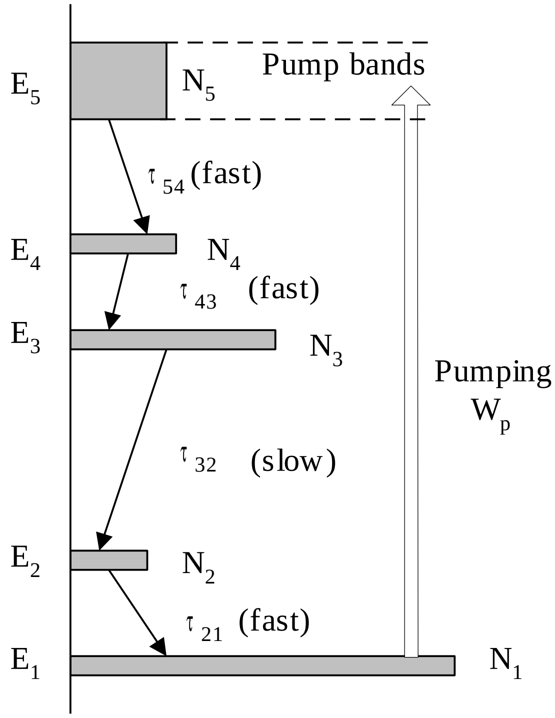


Figure 5.8 Five-level scheme of energy levels for pumping.

Solution: *Energy level diagram for the five-level atomic system is shown in the Figure 5.8. In the optical frequency approximation, the thermally stimulated terms in the relaxation rates are totally negligible compared to the spontaneous rates. This is due to the fact that Boltzmann ratio at optical frequencies is very small, on the order of* 10^{-36} *at room temperature.*

The rate equation for level 5 with pumping transition probability W_p*, at steady-state, is given by (it is given that all the levels can relax only into the level immediately beneath it, therefore, relaxation from* $5 \rightarrow 4$ *is only considered)*

$$\frac{dN_5}{dt} = W_p.N_1 - \frac{N_5}{\tau_{54}} = 0 \qquad (5.m)$$

$$\Rightarrow N_5 = W_p.\tau_{54}.N_1$$

similarly the steady-state population of levels 4, 3, and 2 are given by

$$\frac{dN_4}{dt} = \frac{N_5}{\tau_{54}} - \frac{N_4}{\tau_{43}} = 0 \quad , \qquad \text{(5.n)}$$
$$\Rightarrow N_4 = \frac{\tau_{43}}{\tau_{54}}.N_5 = W_p.\tau_{43}.N_1$$

$$\frac{dN_3}{dt} = \frac{N_4}{\tau_{43}} - \frac{N_3}{\tau_{32}} = 0 \qquad \text{(5.o)}$$
$$\Rightarrow N_3 = \frac{\tau_{32}}{\tau_{43}}.N_4 = W_p.\tau_{32}.N_1$$

and

$$\frac{dN_2}{dt} = \frac{N_3}{\tau_{32}} - \frac{N_2}{\tau_{21}} = 0 \qquad \text{(5.p)}$$
$$\Rightarrow N_2 = \frac{\tau_{21}}{\tau_{32}}.N_3 = W_p.\tau_{21}.N_1$$

From equations (5.o) and (5.p),

$$N_3 - N_2 = (\tau_{32} - \tau_{21}).W_p.N_1 \qquad \text{(5.q)}$$

The conservation condition for total number of atoms gives that

$$N = N_1 + N_2 + N_3 + N_4 + N_5$$

or

$$N = \left[1 + (\tau_{21} + \tau_{32} + \tau_{43} + \tau_{54})W_p\right].N_1 \qquad \text{(5.r)}$$

From (5.q) and (5.r), we have

$$\frac{N_3 - N_2}{N} = \frac{(\tau_{32} - \tau_{21}).W_p}{1 + (\tau_{21} + \tau_{32} + \tau_{43} + \tau_{54})W_p} \qquad \text{(5.s)}$$

The condition of population inversion depends mainly on the lifetimes of the level 3 and level 2, and to achieve population inversion, the lifetime of the level 3, τ_{32} should *be larger that τ_{21}. If all the lifetimes are very small as compared to the τ_{32}, then equation (5.s) can be reduced as*

$$\frac{N_3 - N_2}{N} = \frac{\tau_{32}.W_p}{1 + \tau_{32}.W_p} \tag{5.t}$$

in this case population inversion can easily be achieved.

5.4 Laser Gain Saturation

The steady-state gain is a decreasing function of the cavity photon flux. Physically, the decrease of the gain with increasing photon flux is due to the fact that a large cavity photon number, and therefore a large stimulated emission (and absorption) rate, tends to equalize the populations of upper and lower laser levels. In this case the gain is saturated. The gain saturation plays a very important role in determining the output power of a laser.

The gain saturation can be understood by considering the microscopic behavior. The spontaneous and stimulated emissions apply to an atom in the upper laser level and cause it to drop to the lower level and emit a photon. Once it is in the lower level, it can decay further or it can absorb a photon back from the field. The lower level's absorption rate is exactly equal to the upper level's stimulated emission rate. When the field is so strong that these rates are much greater than the levels decay rate, the atom jumps so rapidly between the upper and lower levels that it has effectively the same probability of being in one or the other of these levels. Then the atom is equally often an absorber and an emitter, and in this extreme limit the gain is zero. Thus in general the gain coefficient of the medium must be reduced as the cavity photon number is increased.

5.4.1 Laser Gain Saturation Analysis

In many real laser systems laser action takes place between two excited levels that are located high above the ground levels. The population density in these excited lasers levels always remains small compared to the total population density of atoms in the lowest energy level E_1. This is particularly true in gas lasers, where line widths are narrow, transitions are relatively strong, and only small inversion densities are necessary to give sufficient gain. It is not true for solid-state lasers, for example, ruby laser, where large fraction of the total atomic density may sometimes be pumped into the upper laser levels.

In this section we will use a simplified model to develop some rate equations analysis, showing how the laser gain itself saturates with increasing signal power in typical laser systems.

Figure 5.9 gives a simplified but a realistic model for many laser systems. Atoms are pumped by some pumping mechanism from the ground level E_1 into some upper level E_4. They then relax down into the upper laser level E_3, from where they relax or make stimulated laser transitions down to the lower laser level E_2, and hence back to the ground level. In this analysis we have included a laser signal, corresponding to laser amplification or oscillation, and represented transition probability W in the diagram.

Suppose the upper-level populations all remain small compared to the initial ground-state population. Then the pumping rate from the ground level E_1 into the upper atomic level E_4 caused by a pumping transition probability $W_{14}=W_{41}=W_p$ may be written as

$$\left(\frac{dN_4}{dt}\right)_{pump} = W_p(N_1 - N_4) \approx W_p N_1 \tag{5.25}$$

where $N_1 \approx N$ is nearly equal to the total density of laser atoms in the system.

In this situation there is essentially no back pumping from E_4 to E_1, since very few atoms accumulate in the upper levels and hence $N_4 \ll N_1$. It is then convenient to speak of a net pumping rate (atoms per second per unit volume) being lifted up out of the ground level, as given by $W_pN_1 \approx W_pN$.

This pumping rate W_pN_1 in a real laser system will be more or less directly proportional to the pump light intensity (in an optically pumped laser), or the discharge current density (in a discharge-pumped gas laser). Moreover, in many real lasers some fixed fraction η_p of the atoms pumped into an upper energy level will decay, often through some cascade process, down into the longer-lived upper laser level E_3. The number of atoms per second reaching the upper laser level is then given by an effective pumping rate, $R_p=\eta_p.W_p.N_1$, where η_p represents the quantum efficiency for pump excitation into this upper

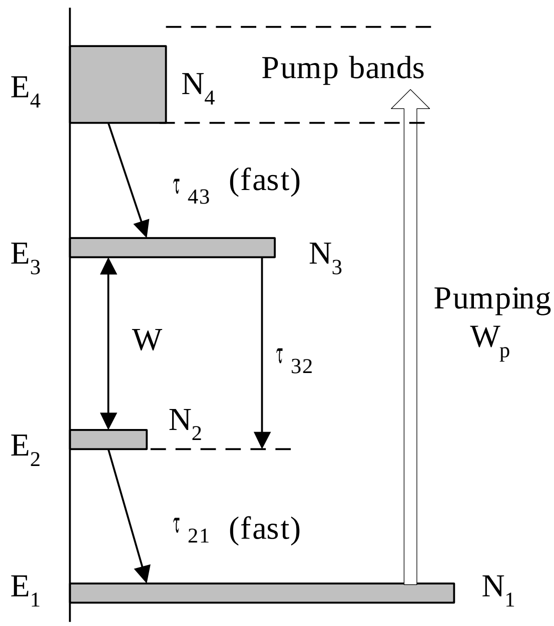


Figure 5.9 Simplified model for laser pumping and gain saturati between two high lying energy levels E_2 and E_3.

laser level. This pumping efficiency may be quite high, even approaching unity (for some lasers), for many solid state and organic dye laser, and may be very small for many typical discharge-pumped gas lasers.

With these generally valid assumptions, the rate equations for the excited laser levels E_3 and E_2, including stimulated transition probabilities $W_{23} = W_{32} = W$ on the laser transition, may be written as

$$\frac{dN_3}{dt} = R_p - W(N_3 - N_2) - \frac{N_3}{\tau_3} \tag{5.26}$$

and

$$\frac{dN_2}{dt} = W(N_3 - N_2) + \frac{N_3}{\tau_{32}} - \frac{N_2}{\tau_2} \tag{5.27}$$

where τ_3 is the total lifetime of the upper laser level and

$$\frac{1}{\tau_3} = \frac{1}{\tau_{32}} + \frac{1}{\tau_{31}}. \tag{5.28}$$

The steady-state solutions to these equations are given by

$$N_2 = \frac{W + 1/\tau_{32}}{W(1/\tau_2 + 1/\tau_{31}) + 1/\tau_2.\tau_3}.R_p \tag{5.29}$$

$$N_3 = \frac{W + 1/\tau_2}{W(1/\tau_2 + 1/\tau_{31}) + 1/\tau_2.\tau_3}.R_p \tag{5.30}$$

The steady-state population difference $\Delta N = N_3 - N_2$ on the laser transition is given by

$$\Delta N \equiv N_3 - N_2 = \left(1 - \tau_2/\tau_{32}\right)\tau_3 . \frac{R_p}{1 + \left(1 + \tau_2/\tau_{31}\right)\tau_3.W} \tag{5.31}$$

If the upper laser level E_3 relaxes primarily into the lower laser level E_2 and not directly down to any lower levels $E_{1,}$ i.e. $\tau_3 \sim t_{32}$, then the above expression for population reduces to simply

$$\Delta N = R_p.\left(\tau_3 - \tau_2\right)\frac{1}{1 + \tau_3.W} \tag{5.32}$$

The inverted population difference in this example varies with both pumping rate and signals intensity in the opposite manner. The second term in the equation represent saturation behavior. As we increase the pump rate, R_p, the signal intensity increases and hence the stimulated emission probability W. At one stage it is the dominant mechanism to bring down the upper laser level and hence responsible for decrease the population inversion, thus reducing the gain. Therefore, at the saturation level gain is not proportional to the pumping rate, R_p.

Problems

5.1 Describe briefly:

(a) Why optical pumping is preferred for solid-state lasers and not for gas lasers?

(b) What is optical frequency approximation?

(c) Role of helium in He-Ne laser.

(d) Laser gain saturation.

5.2 Consider a three-level atomic system similar to one shown in Figure 5.6 with the difference that level 3 can decay to level 2 as well as to level 1 with life times τ_{32} and τ_{31} respectively. Calculate total lifetime τ_3, if τ_{31} = 3 ms, and τ_{32} = 300 μs. Is τ_3 is less than 300 μs? Give reasons.

5.3 Discuss the rate equations of a four-level system and derive expression of population inversion in the following situations:

(a) $(N_4\text{-}N_3)/N_1$ for τ_4 is large and τ_3 and τ_2 are very short.

(b) $(N_2\text{-}N_1)/N_1$ for τ_4 & τ_3 are very short and τ_2 is large.

5.4 Consider a four-level system in which the optical frequency approximation applies between all levels. Assume that this system is pumped from level1 to level 4 with transition probability $W_{14} = W_{41} = W_p$ and each upper level in the system relaxes to all lowering levels. Evaluate population difference on $3 \rightarrow 2$ transition as a function of W_p and relaxation times. What will be effect on population inversion when there is significant decay from level 4 to level 2, i.e., τ_{42} is not very long?

5.5 Suppose a four-level system is pumped with two separate pumping transition probabilities $W_{13} = W_{31} = W_A$ and $W_{34} = W_{43} = W_B$. The optical frequency approximation applies for all levels. Solve for the population difference $N_4\text{-}N_2$ in this system as a function of the two pumping powers WA and WB. Discuss what conditions are needed for an inversion on the 4-2 transition and how this inversion depends on the two pumping powers?

5.6 For a three-level laser system (like the ruby laser discussed in the text) assume for simplicity that the only relaxation processes present are τ_{32}, which is very fast, and τ_{21}, which is slow and purely radiative. In addition to a pump transition probability (W_p) on the $1 \rightarrow 3$ transition, add a signal transition probability W_{sig} on the $2 \rightarrow 1$ transition. Analyze the steady-state population inversion on the $2 \rightarrow 1$ transition as a function of W_p and W_{sig}.

Books for further reading

A. E. Seigman, *Lasers,* (University Science Books, California, 1986).

O. Svelto, *Principles of Lasers*, 3rd ed. (Plenum Press, New York, 1989)

P. W. Milonni and J. H. Eberly, *Lasers,* (John Wiley & Sons, New York, 1991).

Unit-6

Optical Resonators and Laser Modes

Objective

The cavity resonators used for optical radiation usually consist of two flat or curved mirrors set up facing each other, so that an optical wave can bounce back and forth between the two mirrors. Optical resonators of this type have many features in common with lens waveguide in which light is transmitted through a repeated series of identical lenses mounted in a line. In this unit we will discuss the important features of optical resonators. The laser output beam has some frequency structure due to the optical resonator, which are called laser modes. After the resonator, laser modes have been discussed in this unit.

6.1 Introduction

The laser is more analogous to an oscillator than an amplifier. In an electronic oscillator an amplifier which is tuned to a particular frequency is provided with positive feedback. The amplified output is fed back to the input and amplified yet again and so on. A stable output is quickly reached, however, since the amplifier saturates at high input voltages, as it cannot produce a larger output than the supply voltage.

In the laser, positive feedback may be obtained by placing the gain medium between a pair of mirrors which form an optical cavity or resonator. The initial stimulus is provided by any spontaneous transitions between appropriate energy levels in which the emitted photon travels along the axis of the system. The signal is amplified by the process of stimulated emission, as it passes through the medium and 'fed back' by mirrors. Saturation is reached when the gain provided by the medium exactly matches the losses incurred during a complete round trip.

The gain per unit length of most active media is so small that very little amplification of a beam of light results from a single pass through the media. In the multiple passes, which a beam undergoes when the medium is placed within a cavity, the amplification may be substantial.

We have quietly assumed that the radiation within the cavity propagates to and fro between two plane-parallel mirrors in a well-collimated beam. Because of diffraction effects, however, this cannot be the case as a perfectly collimated beam cannot be maintained with mirrors of finite size, some radiation will spread out beyond the edges of the mirrors. Diffraction losses of this nature can be reduced by using concave mirrors. In practice a number of different mirror curvatures and configuration are used depending on the applications and types of laser being used.

The commonly used mirror configurations are:

6.1.1 Symmetric Resonators: In symmetric resonators, both of the cavity mirrors have exactly same curvatures.

Plain-parallel configuration*: This configuration makes maximum use of the laser medium (i.e., we say mode volume is large) but it is difficult to align correctly.*

Confocal configuration: In this configuration the radii of curvatures of the two mirrors is equal to the length of the resonator. The confocal arrangement is easy to align but the only a fraction of active medium is used for lasing action (i.e., mode volume is small). In gas lasers, if maximum output power is required, we use a large radius resonator.

Concentric configuration: In this configuration two identically curved spherical mirrors are separated by a distance equal to twice their radius of curvature, therefore, their centres of curvature coincide.

6.1.2 Hemispherical Resonators: This configuration has combination of plane and curved mirrors in the resonator. If uniphase operation, i.e., maximum beam coherence is required, this configuration is preferred.

6.1.3 Unstable Resonator: In unstable resonators light rays diverge away from the axis. There are many variations in unstable resonators. One simple example is a convex spherical mirror opposite a flat mirror. Others include concave mirrors of different diameters and radii of curvatures.

Optical resonators can be studied using stability diagram. The design of an optical cavity can allow slightly different frequencies to get amplification in a resonator, which form the basis of laser modes.

6.2 Optical Resonators

A typical optical resonator formed by two curved mirrors with radii of curvature R_1 and R_2 spaced a distance L apart is shown in Figure 6.1. For each mirrors R > 0 implies that the

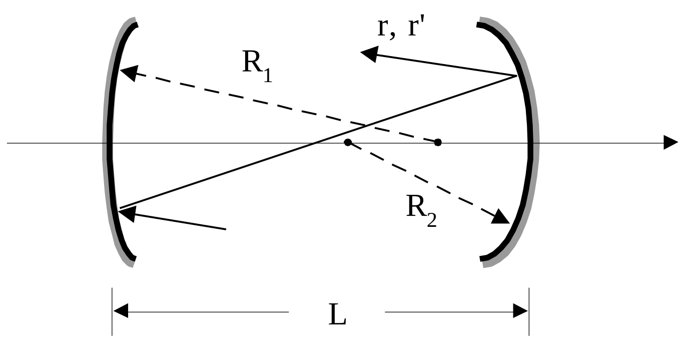


Figure 6.1 A typical optical resonator.

mirror is concave towards the resonator. In the discussion of ray matrices of optical components (Unit-3) we found that ray matrix of reflection from a spherical mirror is similar to the one for a lens. From this similarity (between a curved mirror and a thin lens) we can deduce that the behaviour of a ray upon repeated bounces back and forth between these two mirrors will be exactly the same as the behaviour of a ray passing through a series of lenses spaced at intervals L, with alternate focal lengths $f_1 = R_1/2$ and $f_2 = R_2/2$, as shown in Figure

6.2. The ray properties of the resonator in Figure 6.1 should be exactly same as the ray properties of the series of lenses in Figure 6.2.

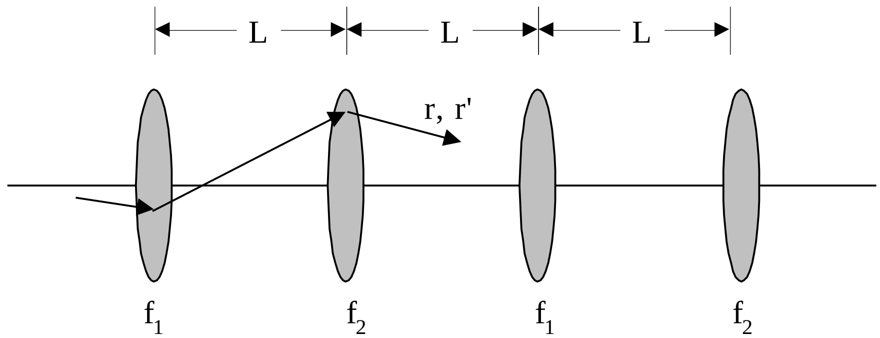


Figure 6.2 An equivalent peeriodic lens waveguide to an optical resonator.

6.2.1 Resonator 'g' Parameters

Let us define a new parameter 'g' which will be very convenient to describe the curvatures and spacing of a two mirror optical resonator, or the equivalent waveguide. The new parameter 'g' is defined as

$$g_1 \equiv 1 - \frac{L}{R_1}, \tag{6.1a}$$

$$g_2 \equiv 1 - \frac{L}{R_2}. \tag{6.1b}$$

The focal lengths of the two lenses f_1 and f_2 can be expressed in terms of the lens spacing L and the g parameters as:

$$f_1 = \frac{R_1}{2} = \frac{L}{2}\frac{1}{(1 - g_1)} \tag{6.2a}$$

$$f_2 = \frac{R_1}{2} = \frac{L}{2}\frac{1}{(1 - g_2)} \tag{6.2b}$$

Let us now analyse the behaviour of a ray upon repeated bounces in an optical resonator of Figure 6.1, or upon passing through a lens waveguide. We are interested to know whether after many bounces the ray will still be reasonably close to the axis of the system or whether it will diverge outward a large distance from the axis. If the ray remains close to the axis after multiple bounces, the resonator is said to be stable otherwise unstable.

We will first consider the net transformation of a ray in passing through one complete round trip inside the resonator or one full period of the series of periodic lens waveguide. It will simplify the analysis somewhat if we consider one full symmetric period of the system, such as the symmetric transformation from the mid-plane of one lens to the mid-plane of the next identical lens, as shown in Figure 6.3. When a ray passes through such a series of optical elements in cascade, the total or overall ray transformation of r and r\` can be computed by successive application of the individual (ABCD) matrices, that is, successive multiplication of the ray vector by the individual ray matrices. If the procedure is applied to the sequence of elements shown in Figure 6.3, the resulting overall transformation for one full period (length 2L) is

$$\begin{bmatrix} r_1 \\ r_1' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/2f_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1/f_2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & L \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -1/2f_1 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_0 \\ r_0' \end{bmatrix} \qquad (6.3)$$

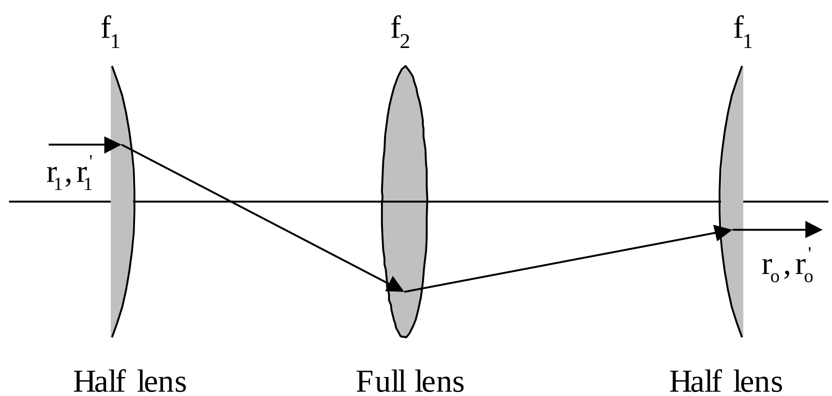


Figure 6.3 The basic unit for analysis, a single symmetric period of len wavegu extending from midplane of one lens to the midpoint of the next identical lens.

If the lens parameters are expressed in terms of g_1 and g_2, and the matrix multiplication in equation (6.3) are carried out, this expression for one complete bounce in the optical resonator, may be reduced to

$$\mathbf{r}_1 = \begin{bmatrix} r_1 \\ r_1' \end{bmatrix} = \begin{bmatrix} 2g_1g_2 - 1 & 2g_2 \\ 2g_1(g_1g_2 - 1) & 2g_1g_2 - 1 \end{bmatrix} \cdot \begin{bmatrix} r_0 \\ r_0' \end{bmatrix} = \mathbf{M}_{\text{total}}\, \mathbf{r}_0 \qquad (6.4)$$

We will try to find the eigenmodes of this system (say λ), i.e.; the output vector $\mathbf{r}_1$ may be obtained by multiplying the input vector $\mathbf{r}_0$ with λ. That is, we must determine input vector $\mathbf{r}_0$ and constant λ that will, together, exactly satisfy the requirement

$$\mathbf{r}_1 = \mathbf{M}_{total}\, \mathbf{r}_0 = \lambda \mathbf{r}_0$$

for this particular total ray matrix. This is equivalent to finding solutions to the matrix equation

$$\left(\begin{bmatrix} 2g_1g_2 - 1 & 2g_2 \\ 2g_1(g_1g_2 - 1) & 2g_1g_2 - 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) \cdot \begin{bmatrix} r_0 \\ r_0' \end{bmatrix} = 0 \tag{6.5}$$

The mathematical procedure for finding possible values of the constant λ, called the eigenvalues of the problem, consists of setting the determinant formed from the total matrix in equation (6.5) equal to zero,

$$\begin{vmatrix} (2g_1g_2 - 1) - \lambda & 2g_2 \\ 2g_1(g_1g_2 - 1) & (2g_1g_2 - 1) - \lambda \end{vmatrix} = 0 \tag{6.6}$$

which gives the quadratic equation for the eigenvalues,

$$\lambda^2 - 2(2g_1g_2\text{-}1)\lambda + 1 = 0.$$

The two eigenvalues or solutions of this equation are then

$$\lambda = \lambda_a, \lambda_b = (2g_1g_2 - 1) \pm \sqrt{4g_1g_2(g_1g_2 - 1)} \tag{6.7}$$

Along with these two eigenvalues λ_a and λ_b, there will be two eigenvectors $\mathbf{r}_a$ and $\mathbf{r}_b$, such that either will satisfy the basic eigenequation

$$\mathbf{M}_{total}\, \mathbf{r}_a = \lambda_a\, \mathbf{r}_a, \qquad \mathbf{M}_{total}\, \mathbf{r}_b = \lambda_b\, \mathbf{r}_b \tag{6.8}$$

It is an important property of the eigenvectors $\mathbf{r}_a$ and $\mathbf{r}_b$ that they form a mathematically complete set: that is, any arbitrary input vector $\mathbf{r}_0$ may be written as a sum of $\mathbf{r}_a$ and $\mathbf{r}_b$ components in the form

$$\mathbf{r}_0 = C_a\, \mathbf{r}_a + C_b\, \mathbf{r}_b \tag{6.9}$$

where C_a and C_b are expansion coefficients. Since passage through one section of the lens waveguide simply implies each eigenvector by its respective eigenvalue, therefore, the output ray after one section (or one complete round trip around the resonator) is given by

$$\mathbf{r}_1 = \lambda_a C_a \mathbf{r}_a + \lambda_b C_b \mathbf{r}_b \tag{6.10}$$

It is easy to calculate the output after an arbitrary number n of complete periods, since each eigenvector component is simply multiplied by its own eigenvalue to the nth power; i.e., the output is

$$\mathbf{r}_n = (\lambda_a)^n C_a \mathbf{r}_a + (\lambda_b)^n C_b \mathbf{r}_b \tag{6.11}$$

6.2.2 Stable Systems

If the g parameters satisfy the condition $0 \le g_1 g_2 \le 1$, the eigenvalues may be written as

$$\lambda_a, \lambda_b = (2g_1g_2 - 1) \pm j\sqrt{4g_1g_2(1 - g_1g_2)} \tag{6.12}$$

and this may also be written as

$$\lambda_a, \lambda_b = \cos\theta \pm j \cdot \sin\theta = e^{\pm j\theta} \tag{6.13}$$

where $j^2 = -1$ and $\theta = \cos^{-1}(2g_1g_2 - 1) \qquad 0 \le g_1g_2 \le 1$

The eigenvalues in this case are complex numbers with magnitude unity and phase angle $\pm\theta$. Since $\lambda^n = (e^{\pm j\theta})^n = e^{\pm jn\theta}$, propagation of an arbitrary initial ray through n periods of a lens waveguide, or through n bounces in an optical resonator, will lead to an output of the form

$$\mathbf{r}_n = C_a e^{jn\theta} \mathbf{r}_a + C_b e^{-jn\theta} \mathbf{r}_b = (C_a \mathbf{r}_a + C_b \mathbf{r}_b)\cos(n\theta) + j(C_a \mathbf{r}_a - C_b \mathbf{r}_b)\sin(n\theta) \tag{6.14}$$

The displacement of the ray after n periods, where n is an arbitrary integer, will be given by an expression of the form

$$\mathbf{r}_n = \mathbf{r}_0 \cos(n\theta) + \mathbf{S}_0 \sin(n\theta) \tag{6.15}$$

where S_0 is related to the initial ray slope. The displacement of a ray while passing through a lens waveguide is illustrated in Figure 6.4, the ray in this case oscillate back and forth in sinusoidal fashion about the system axis, with a maximum limit determined by the entrance condition $\mathbf{r}_0$ and r_0'. Thus the focussing system is stable in the sense that while a ray oscillates about the axis; it always remains within bounded maximum limits. Therefore, for g values satisfying the condition $0 \le g_1 g_2 \le 1$ the focusing system is stable.

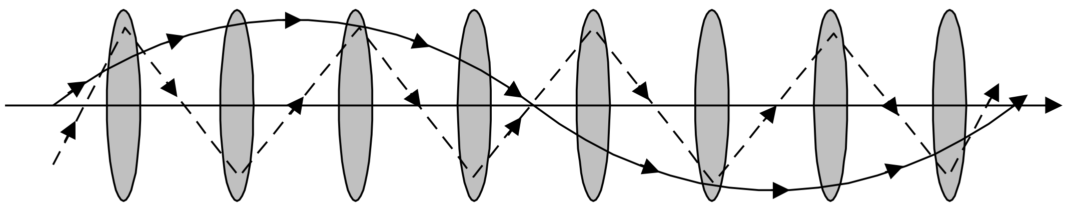


Figure 6.4 Ray propagation in a stable periodic-focussing system. The solid dashed lines show two different types of stable oscillatory behaviour depending the focussing properties of the lenses.

6.2.3 Unstable Systems

In optical resonators or lens waveguides with $g_1g_2<0$ or $g_1g_2>1$, the eigenvalues have the form

$$\lambda_a, \lambda_b = (2g_1g_2 - 1) \pm \sqrt{4g_1g_2(1 - g_1g_2)} \tag{6.16}$$

which may be written as

$$\lambda_a, \lambda_b = e^{+\alpha a}, e^{-\alpha b} \qquad g_1g_2<0 \text{ or } g_1g_2>1 \tag{6.17}$$

In contrast to the stable situation, the trajectory of an arbitrary ray in this case is shown in Figure 6.5. A ray in such a system will diverge exponentially with increasing number of sections or bounces and will eventually pass out of the system or intercept some limiting boundary of the system. The solid and dashed lines show two different forms of unstable behaviour.

6.2.4 Stability Diagram

The stability criterion developed from the ray-optics approach is basic and retains its validity even in more rigorous analysis. This stability criterion is

$$0 \le g_1 g_2 \le 1 \qquad \text{stable resonators} \tag{6.18}$$

$g_1g_2<0$ or $g_1g_2>1$ unstable resonator. (6.19)

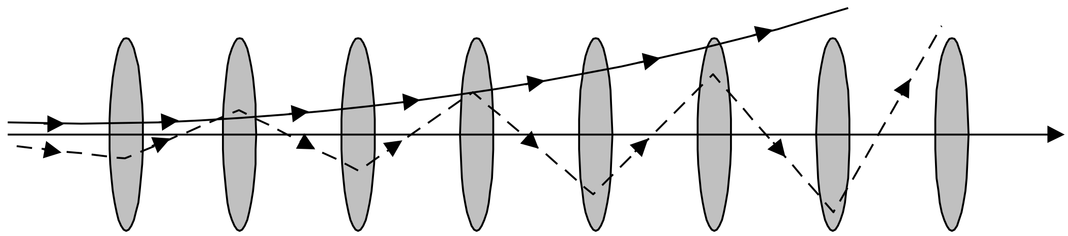


Figure 6.5 Ray propagation in an unstable system.

The criterion can be represented graphically by the stability diagram of Figure 6.6. Every two-mirror optical resonator can be characterised by the parameter g_1 and g_2, and hence represented by a point on the g_1g_2 plane. The resonator or focussing system will be stable only if this point falls within the shaded region of Figure 6.6. The planar, confocal, and

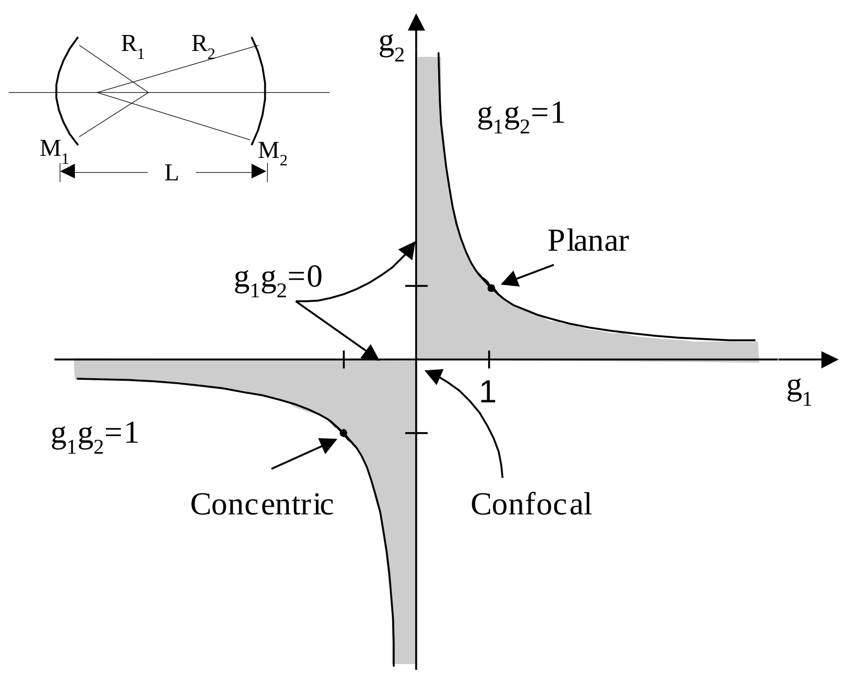


Figure 6.6 The stability diagram for optical resonators and lens waveguide; systems falling in between the axis and the $g_1g_2 = 1$ curve are stable.

concentric resonators are indicated specifically on the diagram. In Figure 6.7 the mirror curvatures appropriate to various regions have been overlaid on the stability diagram. It is interesting to note that a system with two convergent mirrors can still be unstable if the mirror convergence is too strong. Such a system is 'overfocused'. The diagram also shows that it is possible in some cases to have a stable resonator with one divergent mirror, if the other mirror has the proper convergence to make the overall system stable. It is important to note that the stability and instability depend only on the g parameters, and are independent of either the optical wavelength or the transverse size or dimensions of the resonator. In the following section we will examine the various types of resonators that occur in various regions of the stability diagram.

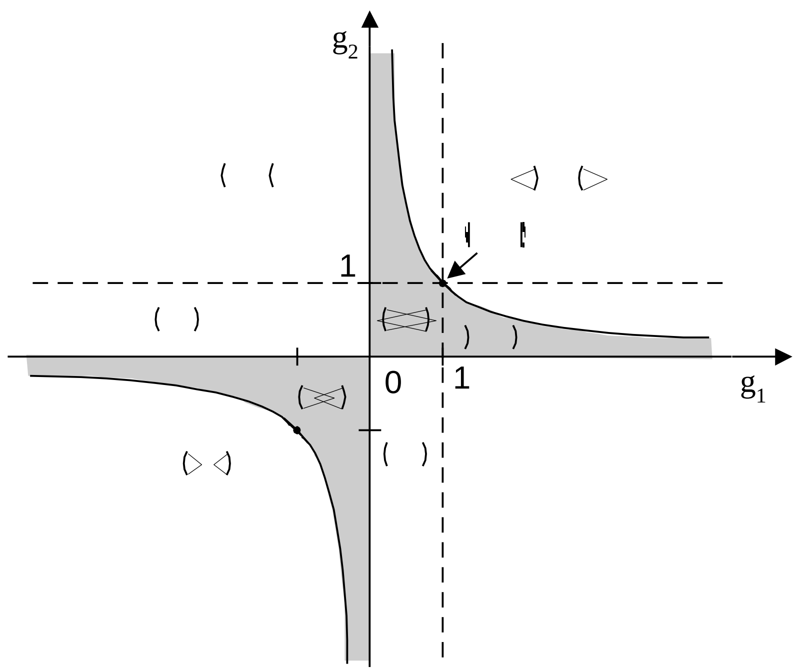


Figure 6.7 The examples of mirror configurations corresponding to differ regions of g_1g_2 plane.

6.3 Important Resonators Types

To gain more insight into the optical resonators, we will survey some resonators at various different points of interest in the stability diagram.

6.3.1 Symmetric Resonators

The simplest resonator configurations to analyse are symmetric resonators, which have mirror curvatures $R_1 = R_2 = R$, and hence g parameters $g_1 = g_2 = g = 1 - L/R$. These symmetric resonators lie along the +45° diagonal through the origin in the g-plane, as shown in Figure 6.8. Three important categories in this configuration are:

1. symmetric confocal resonator, i.e., g = 0
2. two-planar mirror resonator, i.e., g = 1, and
3. concentric or spherical resonator, i.e., g = -1.

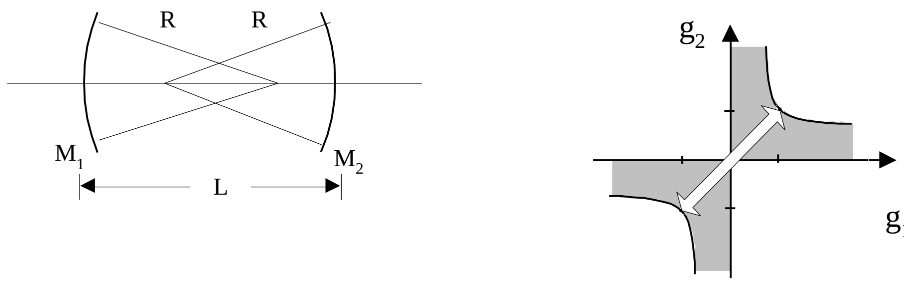


Figure 6.8 Symmetric stable resonators lie along the diagonal axis in the g_1g_2 plane.

6.3.1.1 Symmetric Confocal Resonator

The central point in the stability diagram and an important type of stable optical resonator is the symmetric confocal resonator. This resonator consists of two concave mirrors having radius of curvatures $R_1 = R_2 = L$ thus $g_1 = g_2 = 0$ (Figure 6.9). This is referred to as a confocal resonator because the focal points of the two end mirrors (which are located at the R/2 out from the mirror) coincide with each other at the centre of the resonator.

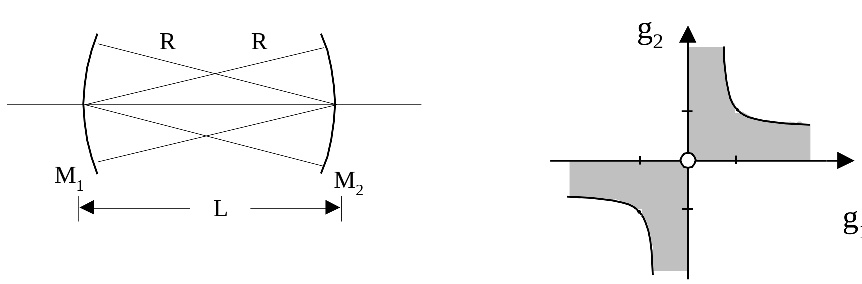


Figure 6.9 The symmetric confocal resonator is a special case, located exactl the centre of the stability diagram.

The confocal resonator is highly insensitive to misalignment of either mirror. Tilting of either mirror still leaves the centre of curvature located on the other mirror surface, and merely displaces the optic axis of the resonator by a small amount. The confocal can be very useful, for example, as a trial resonator design when we are first attempting to obtain laser oscillation from a laser medium whose gain is small or uncertain.

6.3.1.2 Long Radius (Near-Plane) Resonators

Another elementary resonator configuration, and one of that was used in many of the earliest laser devices, is the near-planar or long-radius stable resonator of Figure 6.10. A planar or flat-mirror resonator can be regarded as the limiting case of a long-radius stable resonator as the radii of curvature of the mirrors go to infinity. The resonator parameters then become $R_1 \approx R_2 \approx \infty$ and $g_1 \approx g_2 \approx 1$. The exactly planar resonator occurs right on the stability boundary, at $g_1 = g_2 = 1$.

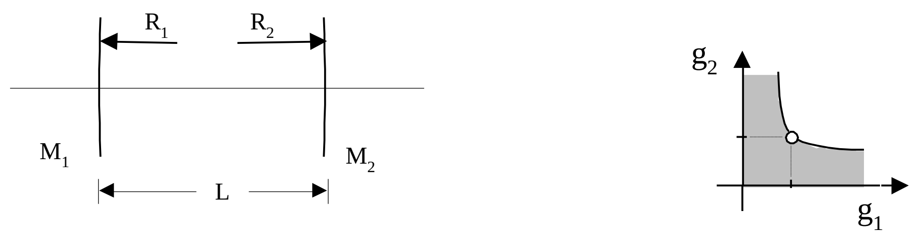


Figure 6.10 Long radius or near planer resonator can have larger mode volu but is very sensitive to mirror allignment.

Although this configuration makes maximum use of the laser medium, however, the long radius resonators are generally avoided in practical laser designs because of their serious alignment difficulties. Long-radius mirrors are also difficult to manufacture and to test.

6.3.1.3 Near-Concentric Resonators

The near-concentric resonator is another design, which is on the boundary of the stability region. In this resonator the cavity length is less than the sum of the two radii R_1+R_2 by the small amount ΔL as illustrated by Figure 6.11. The resonator parameters are given by

$$R_1 \approx R_2 \approx R = L/2 + \Delta L \text{ and } g_1 \approx g_2 = -1+\Delta L/R. \qquad (6.20)$$

The mirror radii are physically reasonable and they can be pulled slowly apart in order to bring the resonator closer to or even across the stability boundary. The central portion of the resonator is not very useful, at least for laser power extraction. This resonator is also very sensitive to very small mirror misalignments.

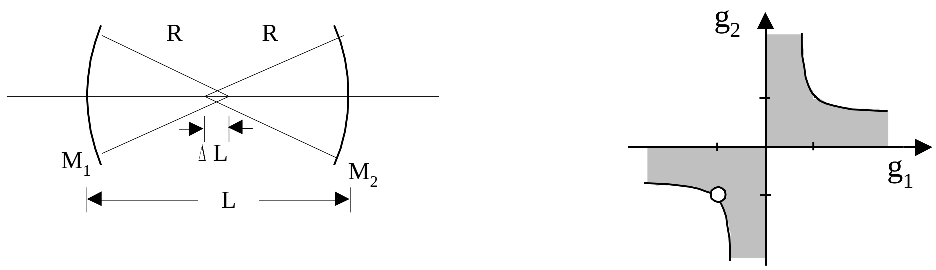


Figure 6.11 The near concentric resonator lies close to the $g_1 g_2 = -1$ point in the stability diagram.

6.3.2 Hemispherical Resonators

The resonator design that is commonly used in practical stable-resonator lasers (e.g., most medium and low-power gas lasers) is the near-hemispherical or half-concentric stable resonator, shown in Figure 6.12. The resonator parameters for this resonator are $R_1 = \infty$ and $R_2 = L + \Delta L$, and hence $g_1 = 1$ and $g_2 = \Delta L/L \approx 0$.

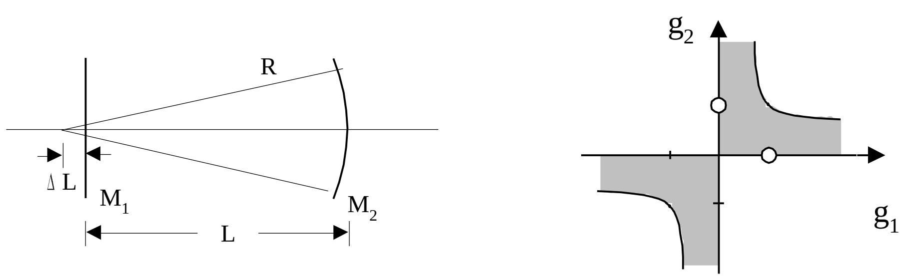


Figure 6.12 The near-hemispherical resonator is widely used in practical la oscillators.

The useful volume in a near-hemispherical resonator is essentially in the shape of a cone. Lasers with near-hemispherical resonators are usually designed with the cavity somewhat longer than the active laser volume. The laser tube is placed near the large-diameter end of the cavity. In typical helium neon lasers, the discharge region is usually stopped well short of the flat-mirror end of the laser. The main advantage of the hemispherical design is that alignment difficulties are largely eliminated.

6.3.3 Concave-Convex Resonators

Any design which operate close to the stability boundary can give large volume size but at the expense of high sensitivity to small fluctuation in the mirror curvature or spacing. By moving out into the regions of the stability diagram beyond $g_1 = 1$ or $g_2 = 1$. It is possible to have so called concave-convex stable resonator such as illustrated in Figure 6.13. Resonators of this configuration have found some practical use, but generally tend to require inconveniently long mirror radii and sensitive alignment procedures.

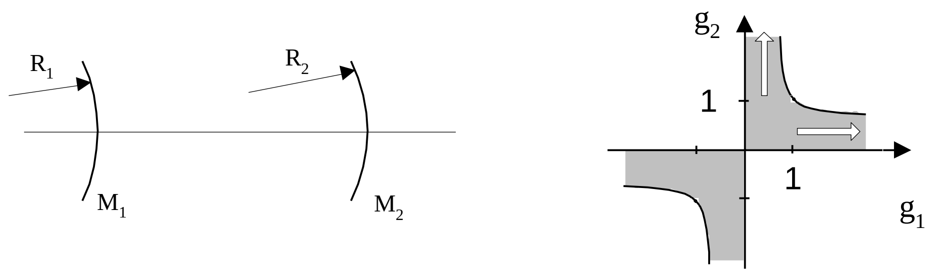


Figure 6.13 Concave-convex stable resonators can also provide large mc volumes, but are seldom used in practice.

6.3.4 Unstable Confocal Resonator

Finally, it is also possible to have resonators that are confocal but asymmetric i.e., resonator in which the two mirrors have different radii of curvature R_1 and R_2 but their focal points still coincide as in Figure 6.14. The spacing for a general asymmetrical confocal resonator is

$$R_1/2 + R_2/2 = L \qquad (6.21)$$

which can be translated into the condition

$$g_1 + g_2 = 2.g_1g_2. \qquad (6.22)$$

Examination of the stability diagram shows that this condition corresponds to a contour of locus which is unstable everywhere in the g_1, g_2 plane, except at the symmetric confocal point $g_1 = g_2 = 0$, and the planar symmetric point $g_1 = g_2 = 1$. Therefore, all asymmetric confocal resonators are unstable. The symmetric confocal resonator shown in Figure 6.9 is located at a kind of singular point in the stability diagram. A small deviation from this point in different directions can lead to either stable or unstable regions of the plane.

The emphasis on stable resonators does not imply that unstable resonators have no practical applications. On the contrary, unstable resonators enjoy certain advantages, and they are essential to the design of many high power lasers. In fact, stable resonators have some drawbacks at high power systems. A major disadvantage is that the modes of stable resonators tend to be concentrated in very thin regions within the resonator. Therefore they do not overlap a very large portion of the gain medium, and this presents a problem if high power extraction from the medium is desired.

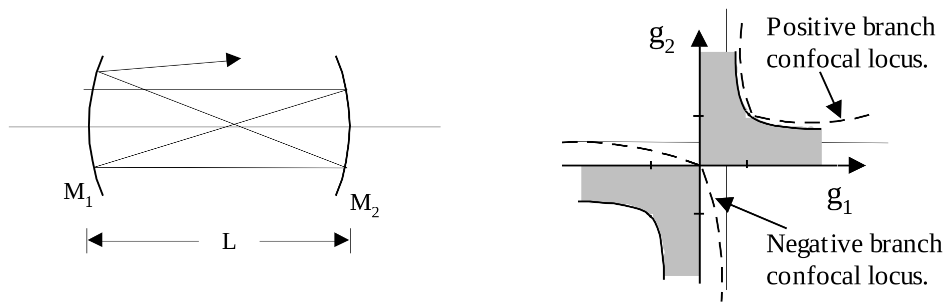


Figure 6.14 All asymmetric confocal resonators lie outside the shaded region a
are thus unstable.

Unstable resonators typically have much larger mode volumes, and can therefore make better use of the available gain region. Figure 6.15 shows an important practical example of an unstable confocal resonator (g_1g_2>1). The intra-cavity field fills a large portion of the cavity, and can be made larger simply by using larger mirrors. The output beam for resonator shown in Figure 6.15 is a collimated annular (doughnut-shaped) beam in the near field close to the resonator.

Unstable resonators offer other advantages in addition to their large mode volume. They can give higher output powers when operating on the lowest-loss transverse mode rather than on several modes. This is an important property and in stable resonators the situation is completely reverse. Another advantage is that unstable resonator lasers use all-reflective optics. That is, the output does not pass through any mirrors but simply spills around the edges. Therefore, at high power operation water cooling of the optics is easier as compare to transmission optics. Examples of a few configurations of unstable resonator are summarised in Figure 6.16.

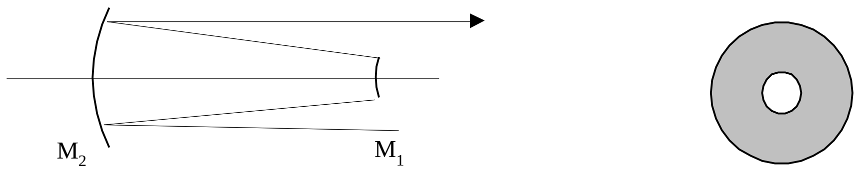


Figure 6.15 A positive branch (g_1g_2 >1) confocal redonator.

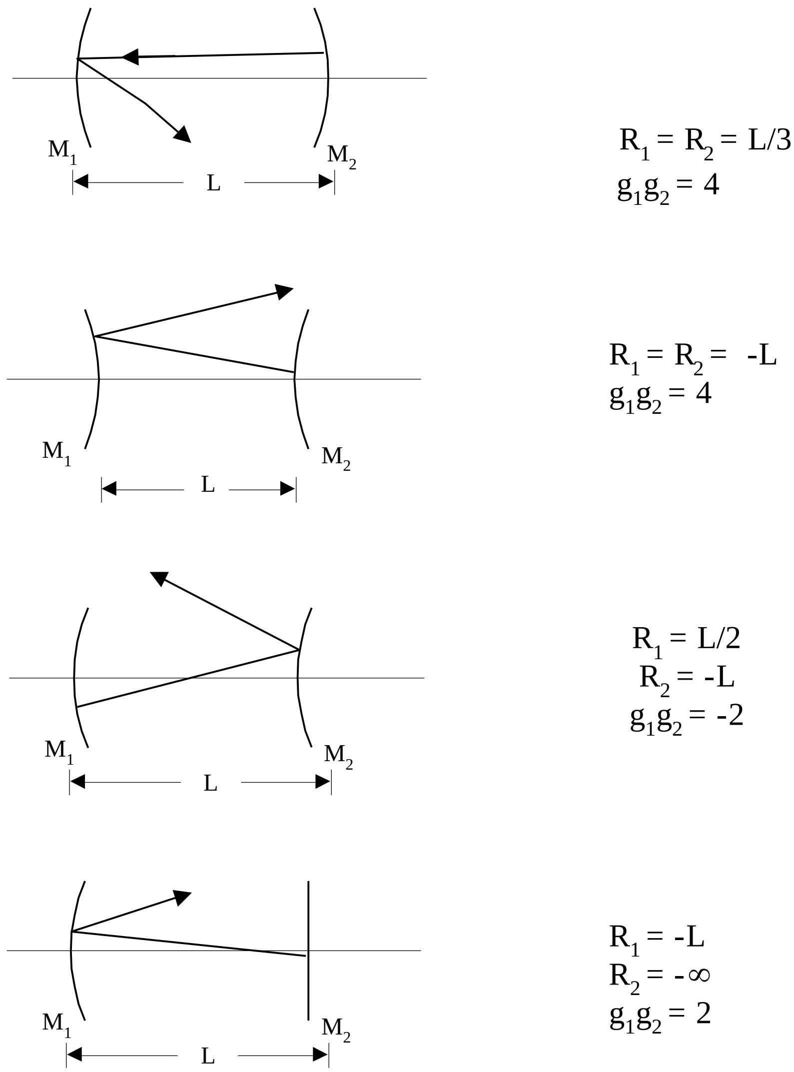


Figure 6.16 Examples of unstable resonators.

6.4 Laser Modes

Examination of a laser output with a spectrometer of very high resolving power, such as the scanning Fabry-Perot interferometer, can show that it consists of a number of discrete frequency components. To understand the reason of these discrete lines and its relationship with the laser transition lineshape we have to examine the effects of mirrors on the light within the laser cavity and the quality factor Q of the resonator. The factor Q can be defined in general by

Q=2π x the energy stored in the resonator / energy dissipated per cycles

or in lasers

$$Q = \text{resonant frequency / linewidth} = \frac{\nu}{\Delta\nu} \tag{6.23}$$

For an electrical oscillator Q may be approximately 100, whereas for a laser Q may be $\sim 10^8$. In lasers the active medium is actually supplying energy to the oscillating modes so that in theory the energy dissipation can be zero and Q infinite. In practice there are always losses, which prevent this happening.

6.4.1 Longitudinal Modes

The optical cavity of a laser is a resonator with extremely high Q and low losses. If these losses are smaller than the gain of the active medium, threshold is achieved and lasing occurs. But the high Q condition does not hold for all frequencies within the laser emission linewidth; only certain frequencies fulfil the resonance conditions. Thus the laser output spectrum does not resemble the spontaneous emission lineshape, but rather consists of a series of narrower lines corresponding to the high-Q frequencies of the laser cavity.

To determine the conditions for high Q in a laser, we start with a plane wave of light propagating along a line in between two parallel mirrors. The round trip distance for a wave undergoing reflection at the mirrors is 2L, twice the distance between mirrors. The total phase change, $\Delta\phi$, of the wave in travelling a full round trip is equal to the 2π times the number of waves in length 2L (i.e., $2L/\lambda$) is:

$$\Delta\phi = 2\pi\frac{2L}{\lambda} = \frac{4\pi L}{\lambda} \tag{6.24}$$

If the reflected wave is 180^o out of phase with the original wave and of equal magnitude then there is no net field and therefore no net energy output from the resonator. This is due to the fact that the wave has not replicated itself upon reflection. Only at such a frequency that the wave and its reflections are in phase ($\Delta\phi = 2\pi q$, q is an integer) does the wave replicate itself. With replication, the electric fields add in phase. The resultant energy density is sufficient to induce substantial stimulated emission at that frequency. In other words, the mirrors form a resonant cavity in which light energy may be stored by multiple reflections between them. If the waves are replicated in the cavity, then the mirror cavity has a high Q. The condition for a self-repeating field (setting $\Delta\phi = 2\pi q$ in the equation 6.24) is that the length of the cavity be equal to an integral number of half-wavelengths, or $L = q(\lambda/2)$, q an integer. Only at those wavelengths is the cavity resonant. The integer q in most cases quite large. For example, if the central wavelength is 500 nm and the mirror separation is 50 cm, q has a value of $2\text{x}10^6$. Since q can be any integer, there are many possible wavelengths within the laser transition lineshape for which the field is self-replicating. We refer to such self-replicating field pattern as longitudinal mode or axial mode of the cavity. It is easier to refer to these axial modes by their frequency than by wavelength. Using the condition for self-replicating field stated above, we have

$$\nu = \frac{c}{\lambda} = \frac{c}{(2L/q)} = q\frac{c}{2L} \tag{6.25}$$

Each mode frequency can be labeled with its corresponding integer q, with the result

$$\nu_q = q\left(\frac{c}{2L}\right) \tag{6.26}$$

It is at these frequencies that the laser cavity is resonant.

By subtracting the frequency of one cavity mode from its nearest neighbour, we find that the separation between mode frequencies is

$$\begin{aligned}\Delta\nu = \nu_{q+1} - \nu_q &= (q+1)\frac{c}{2L} - q\frac{c}{2L} \\ &= \frac{c}{2L}\end{aligned} \tag{6.27}$$

The separation between longitudinal mode frequencies only depends on the mirror separation or cavity length, L, and is independent of q. If we use values from the previous example, the separation between the neighbouring resonance frequencies for a typical laser (50 cm long) is calculated to be

$$\Delta\nu = \frac{3\times 10^{8}\,\mathrm{m/sec}}{2\times 50\times 10^{-2}\,\mathrm{m}} = 3\times 10^{8}\,\mathrm{sec}^{-1} = 300\,\mathrm{MHz}. \tag{6.28}$$

Many laser transition lines are much broader than 300 MHz, and thus there can be many axial modes (…q-2, q-1, q, q+1, q+2, …) within the broadened linewidth. Since sustained laser action can occur only at those frequencies within the lasing transition for which the cavity is resonant, the output of laser contains a number of discrete frequencies, separated by (c/2L), as shown in Figure 6.17. These frequencies are called the axial mode frequencies of the laser.

It is interesting to know that while all the integers q give possible axial cavity modes only those which lie within the gain curve of the laser transition line will actually oscillate. The linewidth of laser transition (Figure 6.17), for the 632.8 nm wavelength, emitted by neon is about $1.5\mathrm{x}10^{9}$ Hz wide. Therefore, with the 0.5 meter long cavity in the above example we would expect four or five modes to be present as illustrated in Figure 6.17c.

6.4.2 Longitudinal-Transverse Modes

In the previous section, we have restricted our attention to resonance conditions for plane waves travelling along a line joining the centres of the mirrors, i.e. along the optic axis of the cavity. For any real laser cavity, the complete wave pattern consists of a superposition of a large number of waves, each wave travelling in a slightly different direction, only nominally along the optic axis. Under these circumstances, the resonance condition is more complicated. The basic requirement is that the electromagnetic field distribution in the cavity replicate itself upon round-trip reflection by the mirrors. Because of their three dimensional

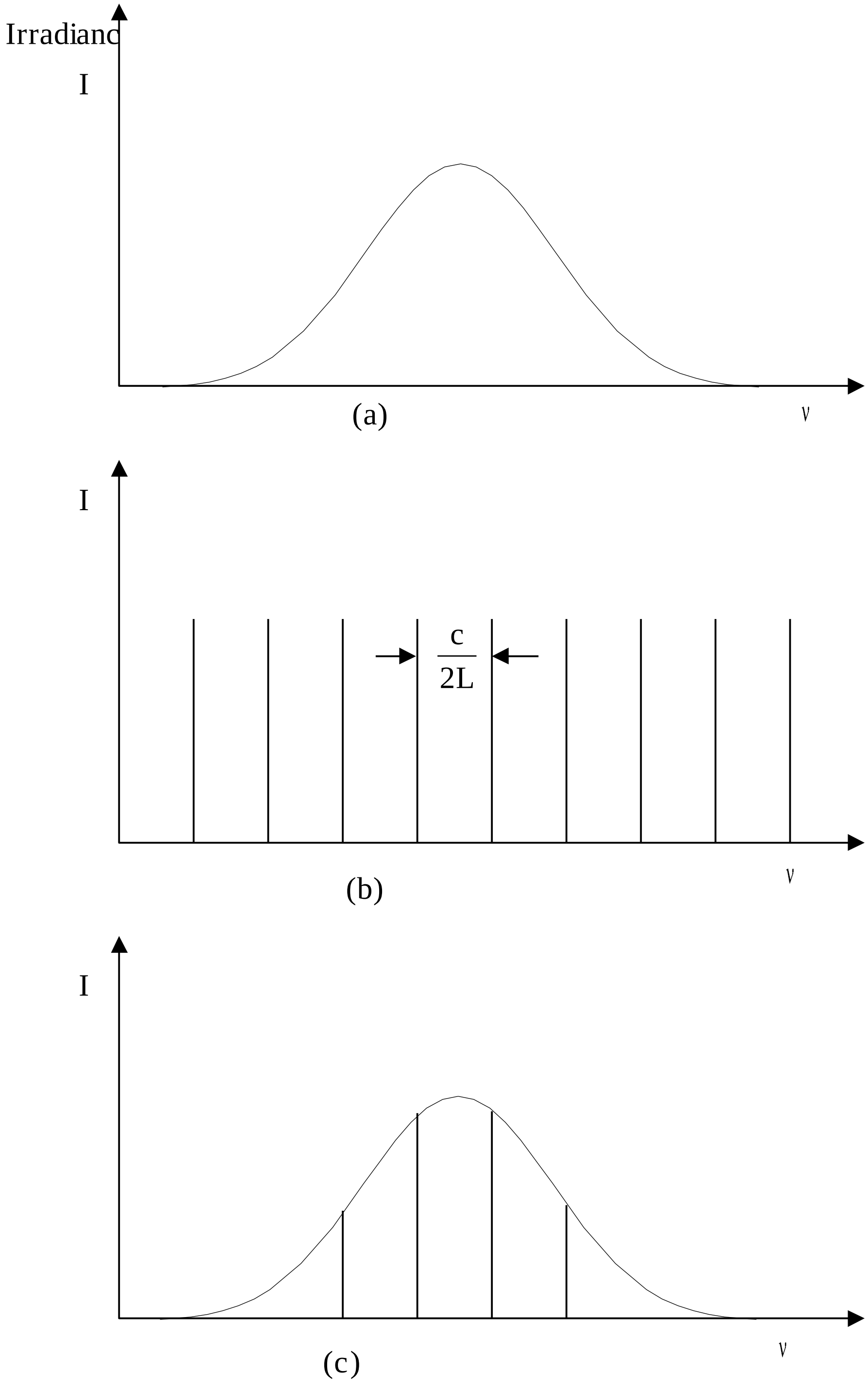


Figure 6.17. (a) The broadened laser transition, (b) cavity modes and (c) axial modes in the laser output.

nature, these self-replicating field pattern can be called as the longitudinal-transverse resonance modes of the laser cavity.

After considering the transverse character of the mode, it is no longer possible to characterize a cavity mode by a single number q. With the application of diffraction theory, it is possible to determine those field distributions that exhibit this self-replicating property in a laser cavity and to determine the frequencies of the corresponding modes. We simply quote the result using g parameters for optical cavities, the frequencies satisfying the full three-dimensional resonance condition can be expressed by the relationship

$$\nu_{mnq} = \left(q + (m+n+1)\frac{\cos^{-1}\sqrt{g_1 g_2}}{\pi} \right) \frac{c}{2L} \tag{6.29}$$

where m, n, and q are integers. This expression can be simplified for the common case of identical near-planar mirrors. Under these circumstances, $g_1=g_2$ and $L/R<<1$ (here L is the length of the cavity and R is the radius of curvature of the end mirrors), we obtained the simpler expression,

$$\nu_{mnq} \cong \left(q + (m+n+1)\sqrt{\frac{2L}{R}} \right) \frac{c}{2L} \tag{6.30}$$

In both equations, the number q is associated with the axial character of the mode. On the other hand, m and n relate to the transverse mode number. Three mode numbers thus characterise a mode of a laser. In practice, the term “axial mode” is often used to differentiate between cavity modes with different values of q. In the same way; the term “transverse mode” is used to differentiate between modes of different m and n values. The different modes are designated by the notation TEM_{mnq}, where TEM denotes Transverse Electro-Magnetic, the light waves consisting of electromagnetic fields that are transverse to the direction of propagation. The value of q is quite large for practical laser dimensions. The values of m and n are usually quite small and often determined by an inspection of the laser output. As a consequence, in the labeling of modes with specific values of m, n, and q, the q is generally suppressed, with the resultant designation having the form TEM_{mn}. It is important to remember that although the mode designation does not contain q, each mode still retains a longitudinal character corresponding to some specific value of q. The important consideration is generally how many longitudinal modes (i.e., different values of q) are present in a laser output rather than the specific q-value, since the number of modes determine the total spread in the laser output spectrum.

Sophisticated equipment (e.g. Fabry-Perot interferometer) is required to observe the longitudinal character or mode structure of the laser output. But the transverse character can be easily identified by placing an inexpensive lens in the beam and observe the expanded beam on a screen. Typical pattern is illustrated in Figure 6.18. These transverse mode patterns depend only on m and n, not on q. The pattern associated with each mode are different and easily distinguished. The TEM^*_{10} mode pattern is a combination of the TEM_{01} and TEM_{10} patterns and may occur when there is a small lossy region or obstruction on the optic axis. It is also given a descriptive name as “doughnut mode” A self-replicating ray for TEM_{01} mode is shown in Figure 6.19. In lasers the most common mode is known as TEM_{00} mode, which has a Gaussian irradiance pattern and the smallest beam divergence of any of the modes.

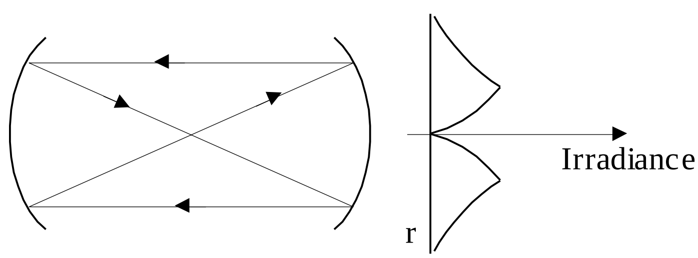


Figure 6.18. Formation of the TEM_{01} mode.

It is possible for a laser to operate in more than one transverse mode, just as it is possible for a laser output to contain more than one axial mode frequency. The expanded beam patterns are then a combination of the contributing separate patterns. It is possible that two modes may have the same q-value (i.e., identified with the same longitudinal mode), the frequencies can still be different.

How many lines can there be beneath the lasing transition? The number of longitudinal modes is determined by the linewidth of the transition line and by the length of the laser. The longer the laser, the smaller the separation between the modes of different q-values, and thus

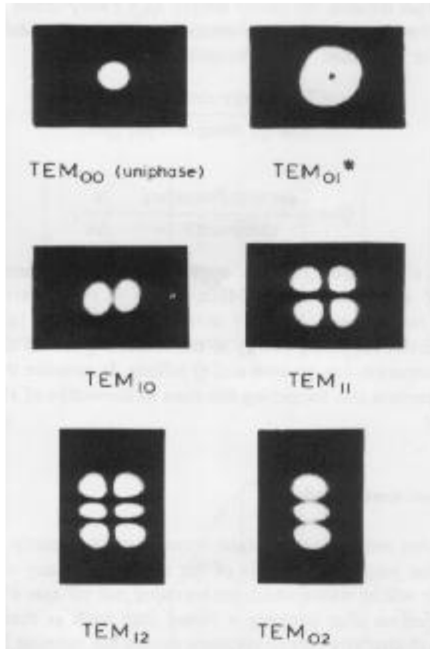


Figure 6.19 Modes

greater the number of modes present within the laser transition linewidth. The number of transverse modes will depend on mirror shape, size, and other aspects of laser construction. When there are a number of modes in the laser output, we refer to the laser as operating multimode.

6.4.3 Single Mode Operation

In many applications including chemical and physical investigation it is desirable to have the greatest possible spectral purity. We can achieve this by operating a cw laser in a single longitudinal and transverse mode. Since an inhomogeneously broadened laser can support several longitudinal and transverse modes simultaneously. Single mode operation can be achieved only by arranging for one mode to have a higher gain than other modes. We can ensure that the cavity will support a single transverse mode only, the TEM_{00} mode, by placing an aperture within the cavity. As the higher order TEM modes spread out more than the TEM_{00} mode an aperture of suitable diameter will transmit the TEM_{00} mode while eliminating the others. All but one of the longitudinal modes can be rejected by reducing the length L of the laser cavity until the frequency separation between the adjacent modes, that is

$\Delta\nu = c/2L$, is greater than the linewidth of the laser transition. Figure 6.17 then shows that the single mode which falls within the transition linewidth is the only one that can oscillate. The disadvantage of this system is that the active length of the laser cavity may become so small as to severely limit the power output. This shortcoming of lower power in single longitudinal mode can be overcome by using Fabry-Perot etalon inside the resonating laser cavity, details of the technique can be found in a number of text books e.g. books given at the end of the Unit.

Problems

6.1 Why plane mirror resonator is called marginally stable?

6.2 Discuss the merits and demerits of unstable resonators versus stable resonators.

6.3 Make a series of sketches showing two mirrors facing each other, with the mirror spacing fixed. Assuming that the centre of curvature C_1 of the right hand mirror is located in the following positions:

(i) C_1 located at the centre of the two mirrors.

(ii) C_1 located to the left of the left-hand mirror.

Indicate by cross hatching those sections of the axis within which the centre of curvature C_2 of the left-hand mirror must be located in order to have a stable resonator.

6.4 An optical cavity of length L is formed by two end mirrors. Set-up the necessary analysis and find out whether the resonator will be stable or unstable, given that

(i) Length of the cavity = $L = R_1 = R_2 = 10$ cm

(ii) $L = 12$ cm, $R_1 = R_2 = 10$ cm

(iii) $L = 12$ cm, $R_1 = 10$ cm, $R_2 = 15$cm

(iv) For all the above parts, draw the stability diagram and show the stability region (if any)

6.5 A resonator is formed by a convex mirror of radius $R_1 = -1$ m and a concave mirror of radius$R_2 = 1.5$ m. What is the maximum possible mirror separation if this is to remain a stable resonator?.

6.6 A resonator is made up of two plane mirrors with a positive lens inserted between the two mirrors. If the focal length of the lens is f and L1 and L2 are the distances of the lens from the two mirrors. Calculate the condition under which the cavity is stable.

6.7 What is the mode separation; how many longitudinal modes could possible if the width of the gain curve is 1.5×10^9 Hz? Take the mirror separation to be 2 m and $\lambda = 588$ nm.

6.8 Consider a confocal resonator of length L = 1 m used for a He-Ne laser at a wavelength, $\lambda = 0.6328$ μm. Calculate the frequency difference between adjacent longitudinal modes.

6.9 Calculate the cavity length for a He-Ne laser, which would sustain only two longitudinal modes. The bandwidth of 632.8 nm He-Ne laser is about 1.5×10^9 Hz.

Books for further reading:

A. E. Seigman, *Lasers*, (University Science Books, California, 1986).

O. Svelto, *Principles of Lasers*, 3rd ed. (Plenum Press, New York, 1989).

P. W. Milonni and J. H. Eberly, *Lasers*, (John Wiley & Sons, New York, 1991).

A. E. Seigman, *Lasers and Masers*, (McGraw-Hill, New York, 1971)

Unit-7

The Laser Output

Objective

In the previous units we have discussed that in general lasers systems have three basic requirements; (i) active medium, (ii) population inversion, and (iii) optical resonator. In this unit we will investigate how these requirements determine the characteristics of the laser output. In the course of this investigation we shall discuss the atomic lineshape associated with a laser transition.

7.1 Introduction

This unit deals with the discussion of the characteristics of the laser beams. Single mode operation of a laser beam makes it useful in many applications where low frequency spread is necessary e.g. high resolution spectroscopy, interferometry. In pulsed lasers, narrowing the pulse duration increases the peak power; this can be achieved by Q-switching. In addition to these, some common properties of laser radiation are discussed. The objective is to make the student familiar with these characteristics of laser beams.

7.2 Lineshape Function

In deriving the expression for small signal gain (unit-4) we assumed that all the atoms in either the upper or lower levels would be able to interact with the perfectly monochromatic beam. In fact this is not so; spectral lines have a finite wavelength (or frequency) spread, that is, they have a spectral width. This can be seen in both emission and absorption. For example, if we measure the absorption as a function of frequency for the transition between the two energy levels E_1 and E_2, we will obtain a bell-shape curve as shown in Figure 7.1a. The emission curve will be the inverse of this Figure 7.1b. The shape of these curves is

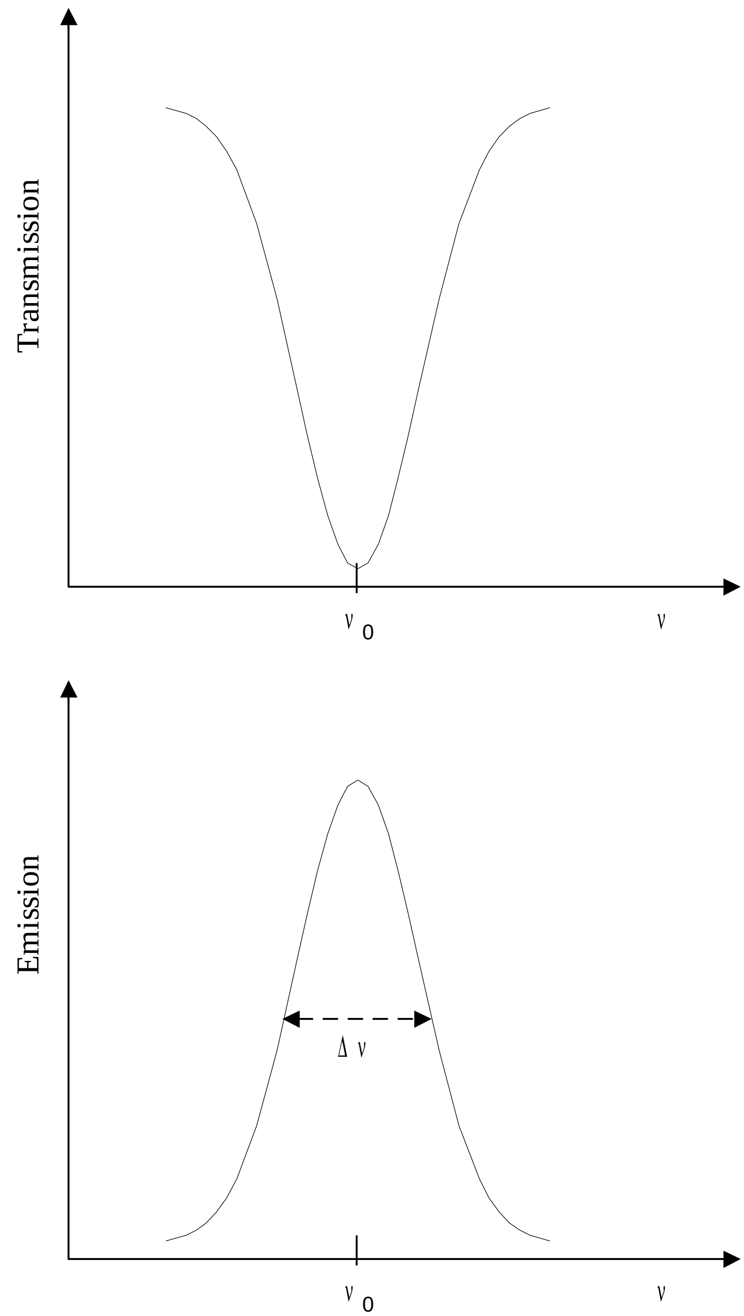


Figure 7.1 (a) The transmission curve for transitions between energy levels E_1 and E_2 and (b) the emission curve for transitions between E_2 and E_1. The precise form of these curves (the lineshape) depends on the spectral broadening mechanisms

described by the lineshape function g(ν). Thus we may define g(ν)dν as the probability that a given transition between the two energy levels will result in emission (or absorption) of a photon whose frequency lies between ν and ν+dν. g(ν) is normalized such that

$$\int_{-\infty}^{\infty} g(\nu).d\nu = 1. \tag{7.1}$$

Therefore, we see that a photon of energy hν may not necessarily stimulate another photon of energy hν. We then take g(ν)dν as the probability that the stimulated photon will have an energy between hν and h(ν+dν).

When a monochromatic beam of frequency ν_s interacts with a group of atoms having a lineshape function g(ν), the small signal gain coefficient may be written as (after equation 4.21):

$$k(\nu_s) = (N_2 - N_1)\frac{B_{21}.h\nu_s.ng(\nu_s)}{c} \tag{7.2}$$

The form of the lineshape function g(ν) depends on the particular mechanism responsible for the spectral broadening in a given transition. The three most important mechanisms are Doppler broadening, collision (or pressure) broadening and natural (or lifetime) broadening.
In the subsequent discussions the broadening mechanisms are described briefly.

7.2.1 Natural Broadening

The natural linewidth of an atomic or molecular transition is due to the spontaneous emission from the excited state. The lifetime (τ) and energy spread (ΔE) of the excited state is related by the uncertainty principle (assuming atom or molecule is free from all other interactions):

$$\tau.\Delta E \approx \hbar = \frac{h}{2\pi} \tag{7.3}$$

The corresponding spread in frequency (half width) is:

$$(\Delta\nu)_{nat} = \frac{\Delta E}{h} \approx \frac{1}{2\pi.\tau} \tag{7.4}$$

In natural broadening, the probability that an atom or molecule will emit a photon within this linewidth is same. Therefore, this is also known as homogeneous broadening.

7.2.2 Doppler Broadening

The Doppler effect occurs because of the relative motion of a source and observer. The frequency as measured by the observer increases if the source and observer approach one another and decreases as they move away. This effect applies to a collection of atoms emitting at an optical frequency ν_{12} so that the observed frequency is given by

$$\nu'_{12} = \nu_{12}.\left(1 \pm \frac{v_x}{c}\right), \qquad (7.5)$$

where v_x is the velocity of the atom along the direction of observation (we assume $v_x << c$). Since the atoms are in random motion, an observer would measure a range of frequencies depending on the magnitude and direction of v_x. That is, as for the observer is concerned, the collection of atoms would be emitting at a range of different resonant frequencies resulting in a broadening of the emission lineshape. The individual Doppler shifted resonant frequencies contribute to a smooth Doppler broadened lineshape.

The mean squared velocity components v_x depend on the temperature since

$$\tfrac{1}{2} m v_x^2 = \tfrac{1}{2} kT \qquad (7.6)$$

where m is the atomic mass, k is the Boltzmann constant, and T is absolute temperature. In thermal equilibrium the gas molecules have Maxwell-Boltzman distribution of velocities and give rise to a line shape function

$$g_{Dopp}(\nu, \nu_o) = \text{Cons}\tan\text{t.}\exp\left[-\frac{mc^2}{2kT}\left(\frac{\nu - \nu_o}{\nu_o}\right)^2\right] \qquad (7.7)$$

The Doppler linewidth (full width of the curve at half the maximum intensity of emission) of the curve is proportional to the square root of the temperature T and is given by

$$\begin{aligned}(\Delta\nu)_{Dopp} &= \frac{2.\nu_o}{c}\left(\frac{2.kT.\ln 2}{M}\right)^{1/2} \qquad (7.8)\\ &= 7.15 x 10^{-7} \nu_o \sqrt{T/M}\end{aligned}$$

where M is atomic weight in atomic mass unit (amu), T is absolute temperature in K, ν_0 and $(\Delta\nu)_{Dopp}$ in Hz. The above equation shows that the Doppler width is small in heavier molecules and can be reduced by some extent at low temperature. Doppler broadening is the predominant mechanism in most gas lasers emitting in the visible. The temperature is elevated, the resonance frequency is high, and the atomic mass is relatively low; all these conditions contribute to Doppler broadening.

Example 7.1: *Calculate the Doppler broadened linewidth for the CO_2 laser transition (λ=10.6μm) and the He-Ne laser transition (λ=632.8nm) assuming a gas discharge temperature of about 400 K. Take the relative atomic masses of carbon, oxygen and neon to be 12, 16 and 20.2 respectively.*

Solution*: The Doppler width for center wavelength λ_0 is given as*

$$(\Delta\nu)_{Dopp} = 7.15\times 10^{-7}.\frac{c}{\lambda_0}\sqrt{T/M}$$

(a) For CO_2 laser: M = 44 amu, λ_0=10.6x10^{-6} m, T=400 K, c=3x10^8 m/s

$(\Delta\nu)_{Dopp}$ *= 61 MHz*

(b) For He-Ne laser: M = 20.2 amu, λ_0=632.8x10^{-9} m, T=400 K, c=3x10^8 m/s

$(\Delta\nu)_{Dopp}$ *= 1508 MHz*

The Doppler width of He-Ne laser is much higher than the CO_2 laser.

7.2.3 Collision Broadening

The collision broadening arises from the collisions between molecules. If the collision is considered as abruptly terminating the life in a particular energy level and τ is the mean time between collisions, which ends the life in the state, the linewidth is

$$(\Delta\nu)_p = \frac{1}{2.\pi.\tau} \tag{7.9}$$

Clearly, the higher the pressure of the gas the more frequently will atoms suffer collisions and the greater will be the spectral broadening.

The Doppler linewidth of molecular lasers such as the CO_2 laser is relatively small because of their low resonant frequencies (in the infrared) and comparatively large molecular masses. In such lasers collision broadening becomes important. Collision broadening also occurs in doped insulator lasers. In these lasers the ions of the active medium may suffer collisions with phonons, which are quantized lattice vibrations.

7.2.4 Homogeneous and Inhomogeneous Broadening

Broadening mechanisms can be classified into homogeneous and inhomogeneous broadening. If all of the atoms of the collection have the same transition center frequency and the same resonance lineshape then the broadening in termed as homogeneous; such as collision and natural broadening. On the other hand, in some situations each atom has a slightly different resonance frequency or lineshape for the same transition. The observed lineshape is then the average of the individual ones, such as Doppler broadening, and the mechanism is termed as inhomogeneous. Local variations of temperature, pressure, and applied magnetic field as well as local variations due to crystal imperfections also lead to inhomogeneous broadening of the emission or absorption lineshapes.

Homogeneous broadening mechanisms leads to a Lorentzian lineshape, which may be written as:

$$g(\nu)_L = \frac{\Delta\nu}{2\pi}\left[(\nu-\nu_0)^2 + \left(\frac{\Delta\nu}{2}\right)^2\right]^{-1} \tag{7.10}$$

where $\Delta\nu$ is the linewidth, that is the separation between the two points on the frequency curve where the function falls to half of its peak value which occurs at frequency ν_0. Putting $\nu=\nu_0$ gives

$$g(\nu)_L = \frac{2}{\pi.\Delta\nu} \tag{7.11}$$

Inhomogeneous broadening mechanisms, on the other hand, lead to a Gaussian frequency distribution, given by:

$$g(\nu)_G = \frac{2}{\Delta\nu}\left(\frac{\ln 2}{\pi}\right)^{\frac{1}{2}} \exp\left[-(\ln 2)\left(\frac{\nu-\nu_0}{\Delta\nu/2}\right)^2\right] \tag{7.12}$$

and putting $\nu=\nu_0$ gives

$$g(\nu)_G = \frac{2}{\Delta\nu}\left(\frac{\ln 2}{\pi}\right)^{\frac{1}{2}} \tag{7.13}$$

Because of these various broadening mechanisms we can no longer treat a group of atoms as though they all radiate at the same frequency. Instead, we must consider a small spread of frequencies about some central value. It might then be expected that the output of the laser would contain the same distribution of frequencies as the broadened transitions of the atoms in the medium. This is; in fact, not the case of as the spectral character of the laser output is different from that of spontaneous emission in the same medium. Two factors account for this difference: namely, the effects of optical resonator and the effect of the amplification process on the irradiance.

7.3 Q-Switching

Q-switching is a way of obtaining short, powerful pulses of laser radiation. Here Q is the quality factor of the laser resonator and the term Q-switching refers to an abrupt change in the cavity loss. Specifically, it is a sudden switching of the cavity Q from a low value to a high value, i.e., a sudden lowering of the cavity loss. A high–Q cavity is one with low loss, whereas a lossy cavity will have a low Q.

Single high power pulses can be obtained by introducing time or intensity-dependent losses into the cavity. If there is initially a very high loss in the laser cavity, the gain due to population inversion can reach a very high value without laser oscillations occurring. The high loss prevents laser action while energy is being pumped into the excited state of the medium. When a large population inversion has been achieved, the cavity loss is suddenly reduced and laser oscillation will suddenly commence. On Q-switching, the threshold gain decreases immediately while the actual gain remains high because of the large population inversion. Due to the large difference between the actual and threshold gain the laser oscillations within the cavity build up very rapidly and all of the available energy emitted in a single, large pulse. This quickly depopulates the upper lasing level to such an extent that the gain is reduced below threshold and the lasing action stops. The time variation of some parameters during Q-switching is shown schematically in Figure 7.2. Q-switching dramatically increases the peak power obtainable from lasers. When the laser is Q-switched, the result is a single spike of great power, typically in the megawatt range, with a duration of a few nanoseconds. It is important to note that, although there is a vast increase in the peak power of a Q-switched laser, the total energy emitted is less than in non-Q-switched operation due to losses associated with the Q-switching mechanism.

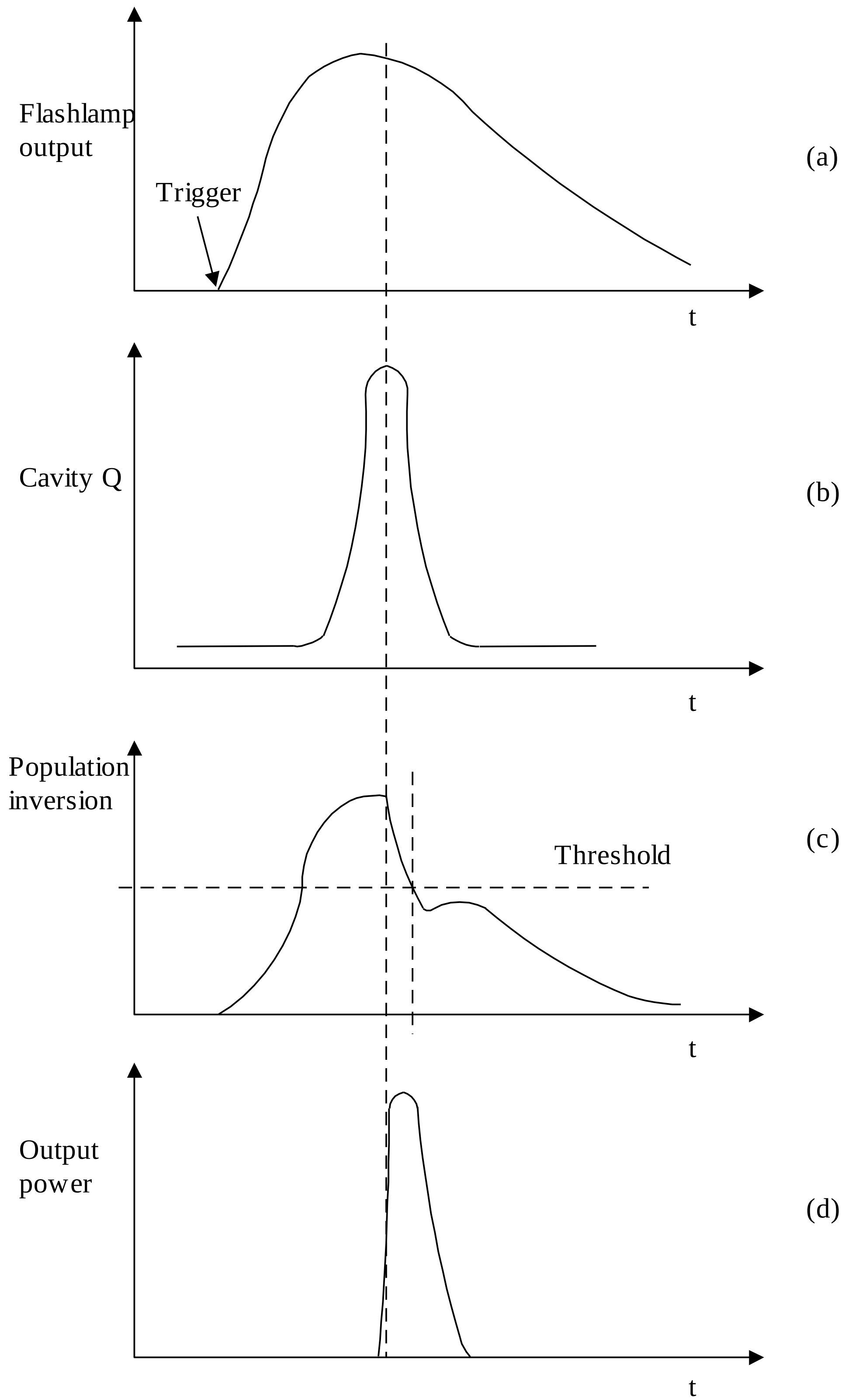


Figure 7.2 Schamatic representation of the variation of the parameters flashlamp output, (b) cavity Q, (c) population inversion and (d) output power a function of time during the formation of a Q-switched laser pulse.

Q-switching is carried out by placing a closed shutter, that is the Q-switch within the cavity, effectively isolating the cavity from the laser medium. After the laser has been pumped the shutter is opened so restoring the Q of the cavity. There are two important requirements for effective Q-switching. These are:

(i) the rate of pumping must be faster than the spontaneous decay rate of the upper lasing level otherwise the upper level will empty more quickly than it can be filled, so that a sufficiently large population inversion will not be achieved; and

(ii) the Q-switch must switch rapidly in comparison to the build up of the laser oscillations otherwise the latter will build up gradually and a longer pulse will be obtained so reducing the peak power. In practice the Q-switch should operate in a time less than 1 ns.

There are different methods of achieving Q-switching.

7.3.1 Rotating Mirror Q-Switch

This is the first method used for Q-switching of solid state laser. This method involves rotating one of the mirrors at very high angular velocity (Figure 7.3) so that optical losses are high except for the brief interval in each rotation cycle when the mirrors are parallel. Just before this point is reached a trigger mechanism initiates the flashlamp discharge to pump the laser. As the mirrors are not yet parallel the population inversion can build up without lasing. When the mirrors become parallel, Q-switching occurs allowing the Q-switched pulse to develop as illustrated in the Figure 7.2.

Although rotating mirror type Q-switch is cheap, reliable and rugged the method suffers from the major disadvantage of being slow. This results in an efficient production of Q-switched pulses with low peak power than can be produced by other methods.

7.3.2 Electro-Optic Q-Switch

Some crystals (and liquids) have the ability to rotate the plane of polarization of light passing through them, they are called optically active. Similar effect can be seen in some crystals by applying the electric field across them, the phenomenon is known as electro-optic effect. The

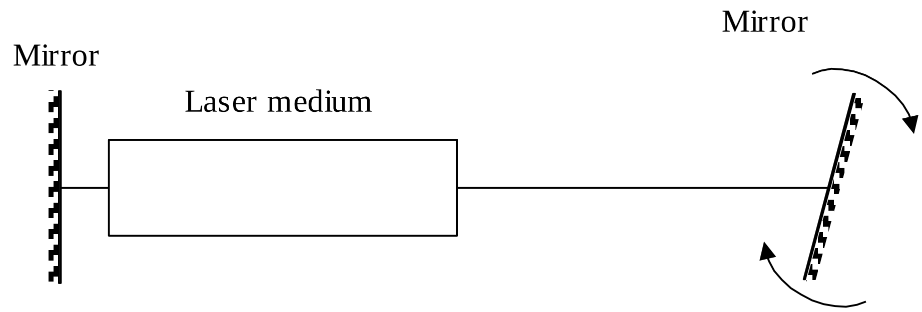


Figure 7.3 A laser cavity with a rotating mirror for Q switching.

electro-optic modulator can be used as fast Q-switch. When a Pockels cell is used and the laser output is not polarized then a polarizer must be placed in the cavity along with the electro-optic cell as shown in Figure 7.4.

A voltage is applied to the cell to produce a quarter wave-plate, which converts the linearly polarized light incident on it into circularly polarized light. The laser mirror reflects this light and in so doing reverses its direction of rotation. On passing again through the electro-optic cell it emerges as plane polarized light, but at 90^{o} to its original direction of polarization. The polarizer does therefore not transmit this light and the cavity is 'switched off'. When the voltage is reduced to zero, there is no rotation of the plane of polarization and Q-switching occurs. The change of voltage, which is synchronized with the pumping mechanism, can be accomplished in less than 10 ns and very effective Q-switching occurs.

7.3.3 Passive Q-Switch

The electronic circuits can be avoided in Q-switching of lasers using absorption characteristics of certain dyes. This passive Q-switching can be obtained by placing in the laser cavity a material that exhibits an absorptivity that decreases with increasing irradiance, as shown in Figure 7.5. An example of such a material is a saturable dye that possesses an absorption band at the lasing transition. At the beginning of the excitation flash, the dye is opaque due to the large number of unexcited molecules that can absorb the light. As in the other Q-switching mechanism, the low cavity Q prevents lasing action and allows a larger population inversion to be achieved than would otherwise occur. As the light irradiance in the cavity increases, more excited states of the dye are populated, until all possible excited

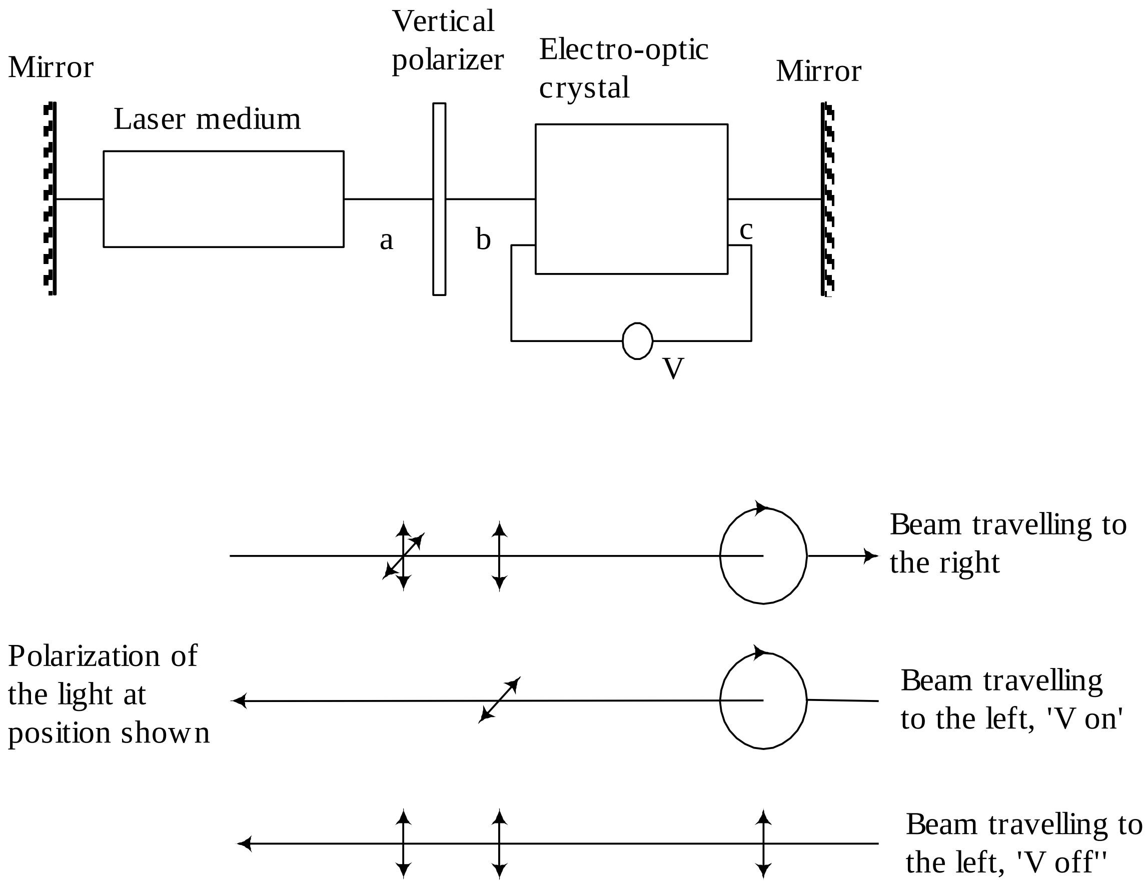


Figure 7.4 Electro-optic crystal used as Q-switch. With the voltage V on, the electro-o crystal act as a quarter wave plate and converts the vertically polarized light at b i circularly polarized light at c. The reflected light is converted to horizontally polariz light and eliminated by the polarizer so that cavity Q is low. With V off, the crysta ineffictive and the cavity Q is high.

states of the dye are filled. At this point, the dye can no longer absorb at the laser wavelength and is said to be *bleached*. The abrupt reduction in cavity losses causes Q-switching to occur. Passive Q-switching has the great advantage of being extremely simple to implement. The only equipment necessary is a small dye cell (which is also available in the form of a thin dye sheet) inserted into the cavity between the lasing medium and one of the ends mirrors.

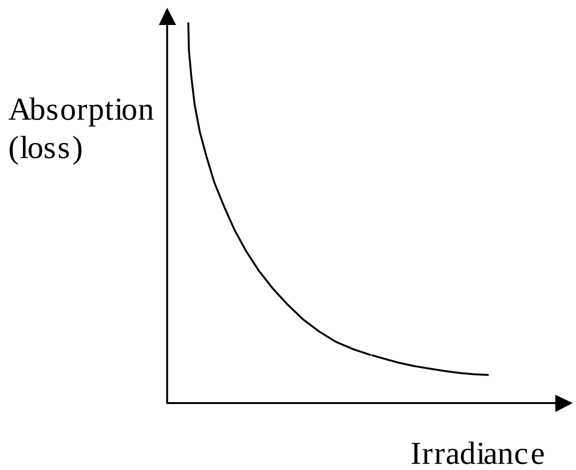


Figure 7.5 Absorption as a function of incidence light irradiance for a satura absorber.

7.4 Laser Beam Properties

The output beam from a laser system is basically consists of electromagnetic radiation, or light. There are some important and fundamental differences of laser light and light emitted from any other source of electromagnetic radiation. Laser beams are often described as being different from ordinary light sources in coherence, spectral purity, divergence, etc. These phrases refer to some characteristics of laser beams that we will review briefly in this section.

7.4.1 Monochromaticity

The word monochromatic is derived from Greek language, meaning single color. In the scientific world it is used for electromagnetic waves of single frequency. In fact no light source, including laser, is capable of producing absolutely monochromatic light, we can only make better and better approximations to the ideal. We might begin with a white light source that produces light of all colors of the spectrum, and filter it with a piece of colored glass. The monochromaticity of the filtered light is now as good as the filter. If the radiation from a gas discharge source, such as neon sign or a sodium vapor lamp, is directed through a prism, a series of lines of different colors is seen on a screen.

The degree of monochromaticity of light from some sources can quantitatively be determined by characterizing the spread in frequency of a line by $\Delta\nu$, the linewidth of the source. This spread can also be represented in term of wavelength, i.e., $\Delta\lambda$. The two spread are related as

$\Delta\nu = -(c/\lambda^2).\Delta\lambda$. This frequency spread depends upon the light source and the level of excitation it can range from the broad, $\Delta\lambda$ ~ 300 nm, (in the case of white light source) to the narrow, $\Delta\lambda$ ~ 0.01 nm (for gas discharge lines). By isolating one of these narrow lines with a suitable filter, we can achieve monochromaticity as good as the width of a single emission line.

If the gas in the above discharge can be made to undergo laser action, this particular emission line is replaced by a series of even narrower lines, which represent different modes. By suppressing all but one of these modes, further monochromatic source of light can be obtained. But even this exceedingly narrow line contains a small spread of different frequencies. If the light consisted of radiation oscillating at a single frequency, the line would be infinitely narrow ($\Delta\nu = 0$) and the light would be absolutely monochromatic. Absolute monochromaticity is an unattainable goal that can only approach by refining our light sources.

The short term spectral purity of a single mode laser can range from a few tens of MHz down to only a few Hz in a highly stabilized system. In fact, it is the laser cavity and not the laser transition that is primarily responsible for these spectral properties. The short term frequency jitters and the long term frequency drift of laser oscillator usually result primarily from mechanical vibrations and noise, thermal expansion, and other effects that tend to change the length L of the laser cavity. Very highly stabilized laser oscillations can nonetheless have long term absolute frequency stability better than 1 part in 10^{10}, and short term spectral purity as high as 1 part in 10^{13}, making them equal to or better than the best atomic clocks available in any frequency range. The ultimate limit on laser spectral purity is finally set by quantum noise fluctuations caused by the spontaneous emission from the atoms inside the laser cavity.

7.4.2 Coherence

Coherence probably is the best-known property of laser light. Light waves are coherent if they are in phase with each other, i.e., if their peaks and valleys are lined up at the same point, as shown in Figure 7.6. Two things are necessary for light waves to be coherent. First, the light waves must start out having the same phase at the same position. Second, their wavelengths must be the same, or they will drift out of phase because the peaks of the peaks of the longer wave.

The laser light is coherent because stimulated emission is coherent with the light wave and photons that stimulates it. The stimulated wave has the same phase and wavelength. It, in turn, can stimulate the emission of other photons, which are in phase and have the same wavelength both with it and the original wave.

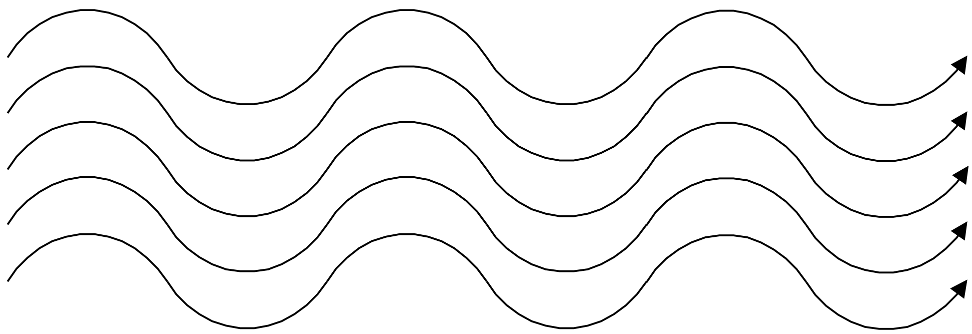


Figure 7.6 Coherence of laser light.

The preceding statement gives the impression that all laser light is perfectly in phase, but this is not the case. Not all photons in a laser beam are descended from the same original photon by stimulated emission, so they don't start out in phase. Even if they were, the uncertainty principle leads to tiny variations in the wavelengths of the emitted light, which accumulate and become significant after the light goes a long distance. In addition, tiny fluctuations within the laser, such as thermal gradients or vibrations, can affect one light wave in a different way than others, degrading coherence. Thus, there is no such thing as "perfect" coherence.

Not all laser light is equally coherent. In fact, the light from some laser is almost incoherent. The reasons lie inside the lasers themselves. Some lasers emit a broader range of wavelengths than others do, and we saw earlier, light waves must have the same wavelength to be coherent. Thus a laser should oscillate in only one frequency, more than one frequencies oscillations reduce the degree of coherence. Lasers with high gain tend to emit a broad range of wavelengths because a range of wavelengths (generated by spontaneous emission) can stimulate emission. In other words, high gain lasers are likely to amplify many different spontaneously emitted light waves. Because of the uncertainty principle, the range of wavelengths in a pulse increase as the length of a pulse decreases, so the shortest pulses tend to have the broadest wavelength ranges. On the other hand, continuous-wave lasers can have the narrowest wavelength range. Add this all together and one finds that the most coherent beams come from continuous-wave, low-gain lasers operating in the low order mode.

If we look closely, there actually are two kinds of coherence: temporal and spatial. Temporal coherence measure how long light waves remain in phase as they travel (the term “temporal” comes from the degree of coherence is compared at different times). Light wave become incoherent as differences in their paths or wavelengths make them drift out of phase. All light has some temporal coherence, but only over a characteristic “coherence length,” which is close to zero for ordinary light bulbs but can be many meters for lasers. In fact the degree of monochromaticity is related to the coherence length by:

$$\text{Coherence length, } L = \frac{c}{\Delta\nu}$$

where $\Delta\nu$ is the linewidth of the source radiation. The coherence time of the radiation is related to $\Delta\nu$ through the relation

$$\tau_c = \frac{1}{\Delta\nu}$$

Example 7.2: *Find coherence length for the following:*

(a) Light bulbs emit light from visible to infrared, i.e., 400 nm to 1000 nm

(b) An ordinary semiconductor laser operating at 800 nm with wavelength range of 1 nm.

(c) An ordinary helium neon laser with frequency bandwidth of 1500 MHz.

(d) A frequency stabilised helium neon laser with frequency bandwidth of 1 MHz.

Solution: *The coherence length* $L = c/\Delta\nu$ *, or in the units of wavelengths* $L = \lambda^2/\Delta\lambda$

(a) Average emission wavelength from a bulb is 700 nm, and wavelength spread $\Delta\lambda$*=600 nm, therefore,* $L= 8.2x10^{-7}$ *m.*

Hence ordinary bulb has a very short coherence length of 0.82 micrometer.

(b) Since λ=800 nm, and $\Delta\lambda$=1 nm, therefore, coherence length L = $6.4x10^{-4}$ m

The coherence length of an ordinary semiconductor laser is also not very large and is only 0.64 mm.

(c) Helium neon laser with a frequency bandwidth, $\Delta\nu$ = 1500 MHz, the coherence length, L, becomes 0.2 m.

The ordinary helium neon laser has a coherence length of 20 cm that is sufficient for some interferometric applications and holography.

(d) A frequency stabilised helium neon laser with frequency bandwidth, $\Delta\nu$=1MHz, L= 300 m.

which is sufficiently large. It has applications in interferometry, high-resolution spectroscopy, etc.

Spatial coherence, on the other hand, measures the area over which light is coherent. If a laser emits a single transverse mode (i.e. laser beam has Gaussian intensity profile in the transverse direction), its emission is spatially coherent across the diameter of the beam, at least over reasonable propagation distances.

7.5.3 Directionality

We think of laser beams as tightly focused and straight, but once they go for enough from the laser, they actually spread out slightly with distance. The spreading is called beam divergence. Laser beams, in general, have very low divergence. The light in a laser is contained between two highly reflective mirrors. As a first approximation, one can think these mirrors as collimating apertures. The mirrors have such a high reflectivity that a wave is reflected many times with only small portion of it being transmitted by the mirrors. The multiple reflections increase the distance the light travels, while confining it within a very small region between the mirrors. Thus, because of the long distance travelled, the curvature of the waves is very small and the light waves emerging from the laser are nearly planar. If we consider a single point source located between the two mirrors, we can get some idea of the collimating property of multiple reflections using only geometry. Each time a light wave reflects off a mirror, the distance between it and the point source that generated it increases. Since the output mirror transmits a smaller region of that wave front after each reflection, it serves as a collimating aperture for the wave. This collimating property can be visualised on a barber or beautician's chair between mirrors on opposite walls. In many lasers the number of reflections is over 50 and as high as several hundred (this number can be calculated from the mirror transmittance). The beam transmitted from the tube looks as though it comes from a point some 100 cavity lengths distant with 100 apertures collimating the beam.

The low divergence of the laser beam implies that the energy carried by the laser beam can be collected easily and focused into a small area. For conventional sources, where the radiation spreads out into a solid angle of 4π steradian, efficient collection is almost impossible. While for laser beam divergence angle is so small that efficient collection is possible even at large distances from the laser.

A single transverse-mode laser can produce an output beam that has almost uniform amplitude and phase across its full output aperture of diameter d. Such a beam can propagate for a sizeable distance with very little spread. This beam has a very small far-field angle at large distances and can be focused into a spot only a few wavelengths in diameter.

The directionality of the laser beam is expressed in terms of the full angle beam divergence, which is twice the angle that the outer edge of the beam makes with the center of the beam (Figure 7.7). The divergence tells us how rapidly the beam spreads when it is emitted from the laser. Although the divergence angle can be given in fractions of degrees, minutes, or seconds, it is common to specify the beam divergence in radians, where 2π radians equal 360^o. For a typical helium neon laser, the beam divergence is about 1 milliradian (10^{-3} radians). Using the small angle approximation ($\tan\theta \sim \theta$) one can easily show that this typical laser beam increases in size about 1 mm for every meter of beam travel.

Example 7.3: *Calculate diameter of the helium neon laser with divergence of 1 mrad at a distance of 2 km.*

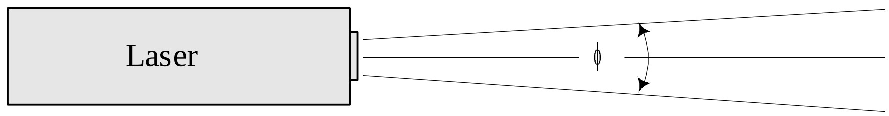


Figure 7.7 Full angle beam divergence of a laser is twice the angle of a laser be
and the beam center.

Solution: *The small angle approximation is valid for such a small angle, i.e., tanθ~θ, we can find the diameter of the laser beam at 2 km as:*

Diameter, d = distance x divergence angle

d = 2000x0.001=2 meters.

Therefore the beam diameter of helium neon laser at 2 km is 2 meter for beam divergence of 1 mrad.

7.4.4 Focusing Properties of Laser Radiation

The minimum spot size to which a laser beam can be focused is determined by diffraction. A single mode beam can be focused into a spot, which has dimensions of the order of the wavelength of light. The imperfections in the optical system may mean that we cannot achieve this in practice. A useful relation for estimating the spot size is that the radius r, at the focal plane of a lens of focal length f is given by

$$r = f.\theta \tag{7.14}$$

where θ is the beam divergence angle in radians. Diffraction provides the lower limit to divergence of a laser beam. In this limit θ is given approximately by λ/D, where D is the limiting aperture diameter, therefore, we have

$$r \approx f\,\frac{\lambda}{D} \approx \lambda.F \tag{7.15}$$

where F is the F-number of the lens. It is impractical to work with F-numbers much smaller than unity, so that r is of the order of λ.

Thus, for example, if we have a 10 mW He-Ne laser with a beam divergence of 0.1 mrad, then an F:1 lens will produce a focused spot with an area of about 10^{-12} m^2 and the power per unit area near the center of the spot will be around 10^{10} W m^{-2}.

If the beam divergence is large the power density is reduced. The insulating crystal lasers generates very high peak powers, they can easily produce high irradiance. A focal area of 10^{-7} m^2 is typical for such lasers giving rise to typical average irradiance of 10^5 W m^{-2} and peak irradiance of 10^8 Wm^{-2}.

Such high irradiance leads to the use of lasers in the drilling, cutting, welding and heat treatment of large number of different materials. The focussing properties of laser radiation are also important for low power applications, e.g., preparation and readout of some home videodisc systems. The information is imprinted on the videodisc in digital form by forming small pits in the surface of the disc with laser. A low power laser to provide a video signal for playback on a television set subsequently reads these pits.

7.5.5 Brightness

The primary characteristic of laser radiation is that lasers have a higher brightness than any other light sources. We define brightness as power emitted per unit area per unit solid angle. The relevant solid angle is that defined by the cone into which the beam spreads. Hence, as lasers can produce high levels of power in well-collimated beams, they represent sources of great brightness.

The brightness is affected by the presence of additional modes, for often, as laser power is increased, the number of mode increases but the brightness remains almost constant. Typical values of brightness from different sources are:

Sun	$\sim 1.3\text{x}10^6$ W m^{-2} sr^{-1}
He-Ne laser	$\sim 10^{10}$ W m^{-2} sr^{-1}
Q-switched ruby laser	$\sim 10^{16}$ W m^{-2} sr^{-1}
High power Nd:glass laser	$\sim 10^{21}$ W m^{-2} sr^{-1}

High brightness is essential for a large number of applications, including material processing, medical application, defense application, fusion, etc.

Problems

7.1 Reduction in linewidth increases the coherence length of the laser beam. Give physical reasons for this fact.

7.2 What are the merits and demerits of Q-switching in a laser system? Give a comparison of different Q-switching techniques.

7.3 A certain laser has frequency bandwidth of 1500 MHz. Represent it in terms of wavelength range if (i) λ = 632.8 nm; (ii) λ = 3.39 μm.

7.4 A coherence length of 50 km is desired for a laser beam. What should be maximum wavelength range for it.

7.5 Calculate the divergence angle of laser beam, given that its radius becomes 2 meters at the distance of 4 km.

7.6 The half-width of the 10.6 μm transition of a low-pressure CO_2 laser is 60 MHz, calculate the coherence length of the laser. If the cavity length is 1 meter show that not more than one mode will oscillate.

Books for further reading:

D. C. O'shea, W. R. Callen, and W. T. Rhode, *An introduction to lasers and their applications*, (Addison-Wesely, California, 1978).

A. E. Seigman, *Lasers*, (University Science Books, California, 1986).

O. Svelto, *Principles of Lasers*, 3rd ed. (Plenum Press, New York, 1989).

P. W. Milonni and J. H. Eberly, *Lasers*, (John Wiley & Sons, New York, 1991).

A. E. Seigman, *Lasers and Masers*, (1971).

Unit-8

Laser Systems

Objective

In this unit some typical laser systems are described. After studying the basic principles of laser, understanding of a few represented systems may clarify the ideas developed in the previous units.

8.1 Classes of Lasers

The common laser systems can be classified into four groups: gas, liquid, solid-state and semiconductor lasers. Before discussing laser systems, it might be useful to remind some of the basic requirements which must be satisfied for laser operation.

Firstly, there must be an active medium, which emits radiation in the required region of the electromagnetic spectrum. Secondly, a population inversion must be created within the medium; this requires the existence of suitable energy levels associated with the lasing transition for pumping.

Thirdly, for true laser oscillation there must be optical feedback at the ends of the medium to form a resonant cavity. The first two conditions can provide light amplification but not the highly collimated, monochromatic beam of light which makes lasers useful.

In the following, we discuss different types of lasers.

8.2 Gas Lasers

Gas lasers are most widely used type of lasers; they range from the low power helium-neon laser (commonly found in teaching laboratories) to the very high power carbon dioxide lasers which have many industrial applications. Basically there are three different classes of gas lasers according to whether the transitions are between the electronic energy levels of atoms or ions, or between vibrational/rotational levels of molecules. These types are called neutral atom gas lasers, ion gas lasers and molecular gas lasers. In general, the energy involved in the lasing processes is well defined and an absence of broad bands effectively eliminates the possibility of optical pumping. Though other methods can be used, most gas lasers are excited by electron collisions in a gas discharge.

Various types of gas lasers are available commercially, e.g. Helium-Neon Laser, Argon and Krypton Lasers, Copper Vapor Laser, Nitrogen Laser, Carbon Dioxide Laser, Excimer Laser etc. In order to have the knowledge about the inside picture of a gas laser, we will discuss the most common type of gas lasers, i.e., the He-Ne laser.

8.2.1 Helium Neon Laser

Helium-neon laser is commonly used neutral atom laser, because it is compact, portable, the low divergent, high coherence length, easily usable, and relatively inexpensive source of laser light. Laser action is obtained from the transitions of the neon atoms (active medium), while helium is added to the gas mixture to greatly enhance the efficiency of the pumping process. Although it has been found that the neon gas alone can provide laser action but the output is enhanced about 200 times when helium is mixed in neon in proportions of 85% helium and 15% neon.

The helium-neon laser oscillates on many wavelengths, which include 632.8 nm (red), 543 nm (green) and the infrared at 1.15 μm & 3.39 μm and many other wavelengths. The He-Ne laser oscillating on λ=1.15 μm transitions was the first gas laser to be operated and also gave the first demonstration of continuous wave (cw) operation in the laser world. But λ = 632.8 nm (red) has become the most popular due to its visibility to human eye.

The output of He-Ne laser is continuous and upto 10 mw(red) are commonly available. A few special-purpose Helium-Neon lasers, relatively larger in size, can produce upto 60 mw of red light.

He-Ne lasers can be used for many applications where a low power is needed e.g. for character reading, meteorology, holography, videodisk memories. They have many uses for educational, demonstration and alignment purposes.

8.2.1.1 Pumping Mechanism

The He-Ne laser is pumped electrically. A longitudinal DC electrical discharge is maintained in the narrow tube containing the gas at a pressure of about 10 torr. In order to provide the current through the tube, one must first provide a sufficiently high voltage to break down the gases in the tube. In a typical small helium-neon laser, the breakdown voltage may be around 3400V. After the breakdown voltage is applied, a plasma is created in the tube. This results in a large increase of current. After the breakdown, the voltage required to maintain the discharge with currents of the order of 10 to 20 milli amperes, is considerably reduced. A typical operating voltage might be around 1350 volts. In order to restrict the current flow to a safe value, a resistor is placed in series with the laser tube.

The pumping process can be described as follows. The first step is the excitation of helium atoms by electron collision to one of the two metastable states designated 2^1S and 2^3S; this is represented by

$$e_1 + He = He^* + e_2$$

where e_1 and e_2 are the electron energies before and after the collision. While in one of the excited states (He*), the helium atoms can transfer their energy to ground state neon atoms with which they may collide. The probability of this resonant transfer of energy is proportional to $\exp(-\Delta E/kT)$ where ΔE is the energy difference between the excited states of the two atoms involved, k is Boltzmann's constant and T is the absolute temperature. The energy level diagram for helium and neon is shown in Figure 8.1. This diagram shows that there is a group of neon levels at almost the same energies as each of the two excited helium states and resonant transfer thus occurs quite readily. The energy transfer is represented by

$$He^* + Ne = Ne^* + He$$

A population inversion is thus created between the 3s and 2p, between 3s and 3p, and between 2s and 2p states of neon. Transitions between the 3s and 2s levels and 3p and 2p

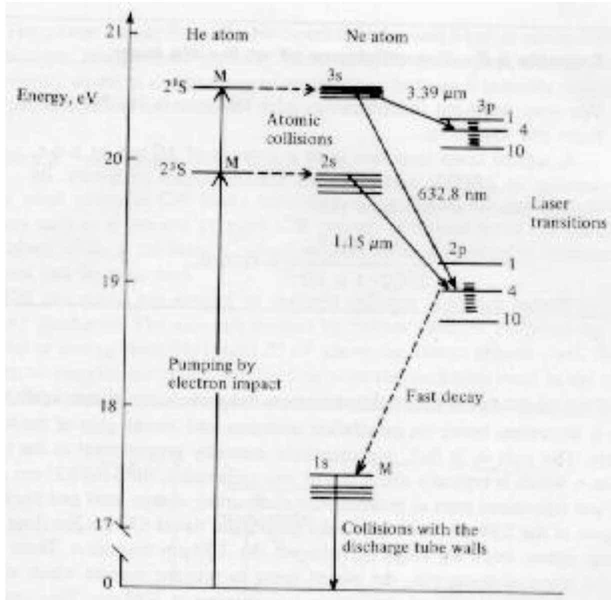


Figure 8.1 Energy levels relevant to the operation of He-Ne laser. M indicates a metastable state.

levels are forbidden by quantum mechanical selection rules. In addition, the decay time of the s states ($\tau_S \cong$ 100 ns) is an order of magnitude longer than the decay time of the p states ($\tau_p \cong$ 10 ns). The p states are rapidly depopulated by spontaneous emission to lower levels. From the ground state they can be again excited to the upper laser level by energy transfer from helium atoms. These parameters satisfy the conditions for operation as a cw laser. The 3s $\rightarrow$ 2p, 2s $\rightarrow$ 2p, and 3s $\rightarrow$ 3p transitions are optically allowed and hence can give rise to various laser wavelengths including 632.8 nm (red), 543 nm (green), 1.523 μm (IR) 3.39 μm (IR), respectively as shown in Figure 8.1.

8.2.1.2 Suppression of the Oscillations other than 632.8 nm

Helium-neon laser can oscillate on many wavelengths, but red laser (λ = 632.8 nm) is widely used. Because of the desirability of visible operation, most He-Ne lasers today are constructed to operate at λ = 632.8 nm (red). We see from the energy level diagram for neon that there is a strong competition between the 0.6328 μm line and the other lines.

When the 1.15 μm line is lasing, it tends to fill the lower laser level for the 0.6328 μm line which will effect the population inversion for 0.6328 μm. Infrared radiation of 3.39 μm wavelength frequently accompanies the emission of the visible line. This transition originates at the $3s_2$ level, the starting level of the visible radiation. So the operation of 3.39 μm laser depletes the $3s_2$ level and hence effects the population inversion required for the visible laser, i.e. this transition uses up atoms that would otherwise contribute to the 0.6328 μm line.

Thus in many practical cases one tries to suppress oscillations of the infrared lines. There are several ways of doing this; the easiest one is to prepare mirrors in such a way that they have high reflectivities at 633 nm and high transmission at 3.39 μm. This can be achieved quite easily by using the multi-layer coated mirrors, which have a wavelength-dependent reflectance. The very low absorption loss of such mirrors is an essential feature as the gain in the He-Ne medium is quite small.

8.2.1.3 Selection of Different Visible Wavelengths

He-Ne laser not only emits 632.8 nm in the visible but it also emits other visible wavelengths, which have very low gain as compared to 632.8 nm. So a He-Ne laser may be operated in the visible region other than 632.8 nm. It can be done by insertion of a prism in the laser cavity. The operation proceeds from the same upper level as the 0.6328 μm line but to different lower sublevels. The dispersion of the prism causes the different wavelengths passing through it to travel in slightly different directions. The prism is rotated so that only one of these wavelengths will strike the mirror normally and be reflected directly back. In this way any one of the several transmission can be favored over the others. Recently laser manufactures have begun offering the He-Ne laser that emit at weaker visible lines particularly the 543 nm (green light).

8.2.1.4 Output Power versus Current

The output power as a function of tube current is shown in Figure 8.2. For low tube currents, the gas discharge is unstable. The plasma flickers on and off and there is no output. For current greater than few milli amperes, a stable plasma is produced and the laser output increases with increasing current. At still higher tube currents, the output power saturates and begins to decrease. This is probably the result of heating of the gas and a decrease in the efficiency of excitation at increased current. In addition, operation at excessive values of currents may decrease the tube life. Thus it is essential to add the ballast resistor in order to keep the tube current within the desired limit.

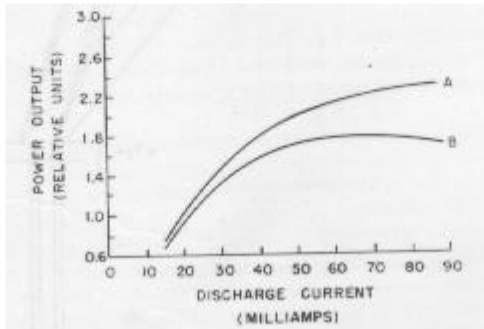


Figure 8.2 Output of He-Ne laser as a function of discharge current. Curve A is for a gas mixture of 0.5 torr of ^{3}He and 0.1 torr of Ne. Curve B is for 0.5 torr of He and 0.1 torr of Ne.

8.2.1.5 Output Power versus Output Mirror Transmission

The gain of the He-Ne laser at 0.6328 μm is relatively low, so that only very small losses can be tolerated in the laser cavity. Thus the mirrors must be of high quality with low scattering losses. The rear cavity mirror is totally reflective. The output mirror typically has transmission around 1 or 2 %. If the mirrors are external to the gas-filled laser tube, Brewster angle windows are necessary on the tube. The output of the helium-neon laser as a function of mirror transmission is shown in Figure 8.3. At zero transmission, the output is of course zero. With increasing transmission, the output increases upto a maximum value. Above the maximum value, the total losses in the laser cavity including the loss due to the output, become too great and the output power decreases. At values of the transmission above 2.5 %, the total losses exceed the gain which in helium neon laser is small, so that the laser operation ceases. Because of the small gain in helium-neon laser, the optimum output coupling is low, less than 1%. It, therefore, follows that the intracavity beam is upto 200 times more intense than the output beam.

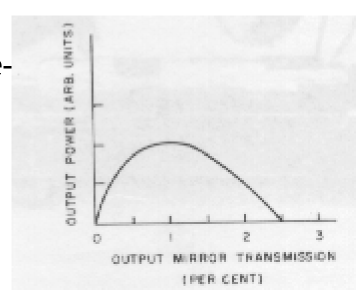


Figure 8.3 Output of He- rror transmission.

8.2.1.6 Mirror Arrangements

The mirrors forming the resonant cavity are sometimes cemented to the ends of the discharge tube. Alternatively the mirrors can be external to the tube which is then sealed with glass windows, which are orientated at the Brewster angle to the axis of the tube. This arrangement allows 100% transmission for the radiation with its electric vector vibrating parallel to the plane of incidence. The arrangement ensures the maximum possible gain (minimum losses) in each round trip. Due to the Brewster windows, output is plane polarized. Although this arrangement is slightly more complicated than the earlier one, but it enables us to insert frequency stabilizing, mode selecting and other devices into the cavity. The mirrors can also be changed to allow operation with other output characteristics and other wavelengths.

8.2.1.7 Gain Diameter Relationship

In the helium-neon laser, the gain is an inverse function of the diameter of the gas tube. This probably occurs because the lower level of the laser transition is depopulated by the collisions of the neon atoms with the walls of the tube. Thus to maintain the population inversion and to keep the laser operation, the neon atoms must be able to collide readily with the walls of the tube. In order that the collision rate of the neon atoms with the walls be high, the diameter of the tube must be small. But it limits the helium-neon laser output power because the volume of the gas can't be increased beyond a certain limit for smaller diameter of the tube.

8.2.1.8 Longitudinal Modes

In the gas laser tube, due to the rapid motion of the atoms or molecules, laser transition does not consist of a sharp line. The Doppler effect causes it to be broadened to a smooth Gaussian profile. The typical Doppler broadened linewidth (FWHM) of a red helium neon laser is 1.5 GHz which corresponds to coherence length of 20 cm. Superimposed on the Doppler broadened, gain curve is the resonant cavity function, with the mode spacing given by:

$$\Delta\nu = c/2L, \tag{8.1}$$

where c is the velocity of light and L is the length of the cavity.

For a 0.5 m long cavity, the mode spacing is 300 MHz. For such length of a cavity, laser output therefore consists of four to five longitudinal modes separated by 300 MHz. The sharpness of these modes depends on the homogeneous broadened mechanisms and is typically 1 MHz. The relative intensities of the cavity modes is defined by the Doppler broadened gain curve.

If one needs a longer coherence length or a narrow bandwidth, the length of the helium neon laser cavity is adjusted such that the output will exhibit either one mode or two, depending on the precise length of the cavity. In fact this implies $L < 10$ cm.

8.2.1.9 Transverse Modes

Most commercial helium-neon lasers are now constructed so as to operate in the TEM_{00} mode. This is done at least partially by ensuring that the ratio of the length of the tube to its diameter is such that the TEM_{00} mode is favored. Other higher order modes have larger diameters and are cutoff by the narrow aperture defined by the plasma tube. So helium-neon lasers typically emit TEM_{00} beams, with diameter about a millimeter and divergence about a milliradian.

8.3 Liquid Dye Lasers

Liquids have useful advantages in relation to both solid and gas laser media. Solids are **very** difficult to prepare with the requisite degree of optical homogeneity and they may suffer permanent damage if overheated. Gases do not suffer from these difficulties but have a much smaller density of active atoms. Several different liquid lasers have been developed but the most important is the dye laser. The important feature that dye lasers offer is tunability, i.e. it has the advantage that it can be tuned over a significant wavelength range, from near I.R. to near U.V. This is extremely useful in many applications such as spectroscopy and the study of chemical reactions.

The active medium is an organic dye dissolved in a solvent. The dye materials are relatively complex organic molecules, with molecular weights of several hundred. For example, Rhodamine 6G is one of the important dye material to be used for lasing action. It contains several benzene rings and has the chemical formula $C_{26}H_{27}N_2O_3Cl$ with a molecular weight of around 450. The dye materials are dissolved in organic solvents, commonly methyl alcohol. Thus the active martial for dye lasers is a liquid.

8.3.1 Energy Level Diagram

Optical pumping is used for dye materials. When the dye is excited by short wavelength light, it emits radiation at a longer wavelength i.e., it fluoresces. The energy difference between the absorbed and emitted photons ultimately appears as heat. Typical absorption and emission spectra are shown in Figure 8.4**.** The broad fluorescence spectrum can be explained by the energy level diagram of a typical dye molecule.

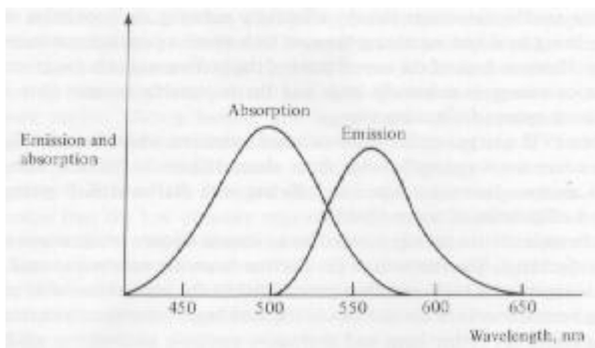


Figure 8.4 Absorption and emission spectra of a typical dye laser.

A typical energy level diagram of a dye material is shown in Figure 8.5. This figure shows

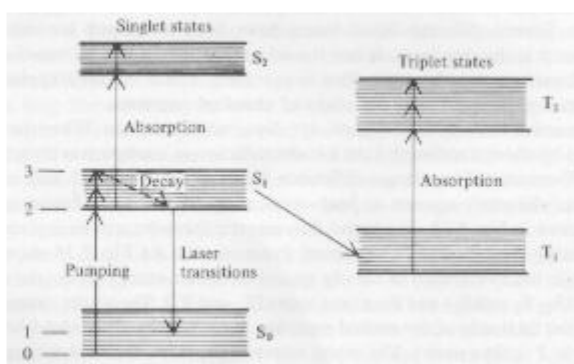


Figure 8.5 Energy level scheme for a dye molecule.

that the dye molecules have two groups of closely spaced electronic energy levels, the singlet states (S_0, S_1 and S_2) and the triplet states (T_1 and T_2). Each electronic energy level is broadened into a near continuum of levels by the effects of vibration and rotation of the dye molecule and also by the effects of the solvent molecules.

When the dye material is irradiated with light whose wavelength corresponds to the energy difference between S_0 and S_1, some of the ground state molecules are raised to sub-levels of S_1. The molecules from sub-levels of state S_1 decay in a very short time to the lowest lying sub-levels of the state S_1. This decay occurs by non-radiative transitions which produces heat. The upper sub-levels of the state S_0 are initially empty. Thus population inversion is achieved between S_1 and the upper sublevels of S_0. Since there are many such rotational/vibrational levels within S_0 and S_1 there are many transitions resulting in an emission line, which is very broad. This leads to the possibility of tuning dye lasers. As the termination of the laser transition in S_0, is at energy much larger than kT above the bottom of S_0, the dye laser is a four level system and threshold is reached at a very small population inversion.

8.3.1.1 Effects of the presence of T_1 and T_2 States

Although triplet states are not directly involved in the laser action, they have a profound effect, as there is a small probability of a transition $S_1 \rightarrow T_1$, even though this is forbidden by quantum mechanical selection rules. However, when dye molecules are excited to the state S_1, some of them are capable of reaching to state T_1 because of collisions. Since the transition $T_1 \rightarrow S_0$ is also forbidden, molecules pile up in T_1. This process steals the molecules that could otherwise be operative in laser operation. The transition $T_1 \rightarrow T_2$ is allowed, and unfortunately the range of frequencies required for this transition is almost exactly the same as the laser transition frequencies. Thus once a significant number of molecules have made the $S_1 \rightarrow T_1$ transition, absorption of $T_1 \rightarrow T_2$ reduces the gain and may stop the laser action. For this reason most dye lasers operate in short pulses, shorter than the time taken for T_1 to acquire a significant population, which is typically 1 μs. For long pulse or cw operation the population in T_1 will build up to equilibrium values, in which absorption is high and becomes the ultimate limitation on the efficiency of the laser.

Continuous dye lasers have been produced by rapidly circulating the dye molecules through a continuous pumping beam. The result is a dye laser, which has a continuous output. However, each molecule of the dye is irradiated by a brief pulse of pumping light as it passes through the pump beam. Then as molecules accumulate in state T_1 in a particular volume of fluid, that volume is circulated out of the laser cavity and is replaced by fresh fluid. The dye can be circulated by a pump. The circulation time is long enough so that the state T_1 can decay before a particular molecule return to the laser region.

8.3.2 Pumping Sources

Dye lasers are optically pumped, the pumping source having a wavelength slightly less than that of the laser output. Commercial pumping methods include flash lamps, nitrogen lasers, excimer lasers, solid state lasers and ion (Ar^+ or Kr^+) lasers. Out of these sources, Ar^+ laser (as well as Kr^+ laser) is a continuous source of light, all other pumping sources are generally operate in pulsed mode and produce a pulsed output from the dye laser. The choice of the pump source depends on the absorption spectrum of the dye being used and the type of output desired. In practice, lasers are used as the pumping sources for dye lasers. We shall discuss two configurations which have been employed. One geometry uses a pulsed nitrogen laser, shown in Figure 8.6. The population of T_1 builds up in about 1 μs, whereas the nitrogen laser operates only in very short pulses, with duration of the order of 100 ns. Thus, the problem with population of state T_1 is eliminated.

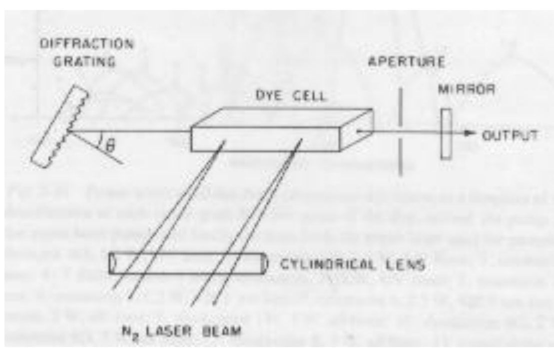


Figure 8.6 Schematic diagram of a puled dye laser excited by a Nitrogen laser beam, with tuning by a diffraction grating.

In Figure 8.6, the wavelength selecting element is a diffraction grating which serves as one of the mirrors. The diffraction grating obeys the equation

$$n\lambda = 2d \sin\theta \qquad (8.2)$$

where n is a small integer, λ the wavelength, d the spacing between rulings, and θ the angle between the light and the normal to the grating. Light of a particular wavelength which will be reflected back into the cavity can contribute to laser action. Thus variation of θ can pick a particular value of wavelength. Therefore, rotation of the grating can serve to vary the wavelength of laser operation.

A second design which represents a continuous dye laser pumped by an argon laser is shown in Figure 8.7. The argon laser beam is focused to a small spot and the dye flows through the

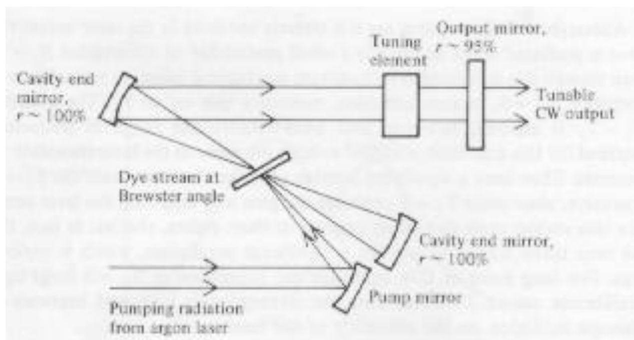


Figure 8.7 Schematic diagram of a tunable CW dye laser. The dye stream flow is perpendicular to the page.

dye laser cavity. A typical velocity for the dye flow is around 10m/s. If the pump beam is focused to a diameter around 10μm, the effective pumping pulse as the dye molecules pass through the pumping beam is around 1μsec. The laser efficiency and output powers are improved if the fluid flow is made as fast as possible and the diameter of the pump beam is made as small as possible. The dyes are forced to flow in a high jet, which is tilted at Brewster's angle to the incoming pump light. The curved mirror, shown in Figure 8.7, is used to focus the pump light. A wavelength-selecting filter is used as the tuning element. By rotating the filter, light of one particular wavelength will be able to reach the mirror and be reflected back into the laser cavity for further stimulated emission.

8.3.3 Tuning Ranges

The entire ranges of wavelengths (from near I.R. to U.V.) can't be achieved using a single dye material. Each dye material has a certain range of wavelength emission. A typical tuning range for one particular dye is several tens of nanometer. When the end of this tuning range is reached, one may physically change the dye material that is being used. This can be done by physically lifting out the reservoir of the dye and replacing it with a reservoir containing a different dye. In this way tuning across the entire visible spectrum may be obtained. This is illustrated in Figure 8.8, which shows the available power levels which may

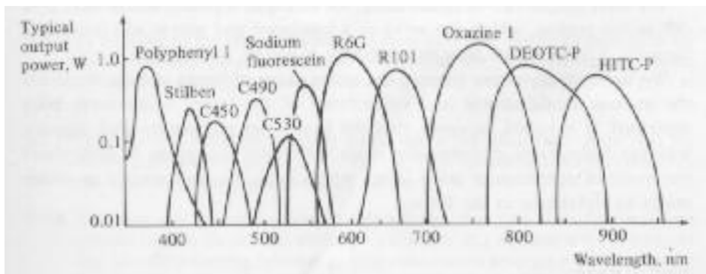


Figure 8.8 Relative outputs of some common dyes pumped by ion lasers.

be obtained in a continuous flowing dye laser pumped by an argon laser. Thus one can tune between approximately 570 to 650 nm with rhodamine 6G. This dye has an efficiency of 20%. When the end of the tuning range of this dye is reached, the rhodamine 6G is replaced with another dye and one may proceed further to the red. When the end of the tuning range of this dye is reached, another dye may be used, and so on. In this way, dye lasers can be tuned throughout the entire visible spectrum.

8.4 Solid-State Lasers

A solid-state laser is one in which the atoms that emit light are fixed within a crystal or a glassy material. The materials used for solid-state lasers are electrically non-conductive, and must be excited optically from an external source.

These lasers are rugged, easy to maintain and capable of generating high peak powers. Typical examples are Ruby Laser, Nd:YAG Laser, Nd-Glass Laser.

We will discuss Nd-YAG Laser.

8.4.1 The Nd:YAG Laser

Neodymium:YAG lasers are the most popular type of solid state lasers. These are four-level laser systems. They can be pumped by flash lamp, arc lamp or semiconductor lasers. These lasers can be operated either cw or pulsed. In Nd-YAG laser, Neodymium serves as active medium, YAG (Ytterium Aluminium Garnet) is the host. The Nd:YAG laser operates at wavelength of 1.064 μm (near I.R.). Its 2nd and 3rd harmonic can shift the output wavelength from the near I.R. into the visible or ultraviolet. Neodymium lasers can generate continuous beam of a few milliwatts to over a ten of watts, or pulsed beam with average powers in the kilowatt range or in short pulses with peak powers in the terra-watt range.

Neodymium lasers have found many applications in research, industry, medicine, military equipments etc. Commercial types range from small-diode pumped models emitting several milliwatts continuous light, to materials working system generating 1 kilowatt. Research versions cover an even broader range, from tiny fiber lasers to building sized lasers used for fusion research in advanced laboratories.

8.4.1.1 Yttrium Aluminum Garnet

Yttrium Aluminum Garnet is probably the most commonly available crystalline laser host. It is a hard isotropic crystal. Its chemical formula is $Y_3Al_5O_{12}$. Its crystalline structure is similar to that of garnet. In YAG, some of the Y^{+3} ions are replaced by Nd^{+3} ions. Typical doping level of neodymium is about 1% (by weight). Higher doping leads to strained crystal since the radius of the Nd^{+3} ion is ~ 14% larger than that of the Y^{+3} ion. This doping level makes the YAG crystal (which is otherwise transparent) to appear pale purple in color because of the Nd^{+3} absorption bands in the red. With this doping level, the neodymium concentration is of the order of 10^{20} atoms per cubic centimeter; considered optimum for lasing action. The active Nd^{+3} ion surrounds itself with several oxygen atoms that largely shield it from its surroundings.

Typically YAG rods are 3 to 9 mm in diameter and up to 10 cm long. The crystal has good thermal, optical and mechanical properties. It offers low value of threshold and high values of gain. Its hardness is 8.5 on Moh scale. It has high thermal conductivity; over ten times that of glass. These desirable properties allow the Nd-YAG rod to produce a good quality laser beam, something difficult at room temperature for most other solid state laser materials.

8.4.1.2 Pumping Mechanisms

The thermal and optical properties of Nd:YAG let it to be pumped either continuously or pulsed. Medium pressure (500 — 1500 torr) Xe-lamps and high pressure (4 — 6 atm.) Kr-lamps are used for the pulsed and continuous wave operations respectively. Diode lasers can also be used for pumping of solid state lasers.

Lamp Pumping

Flash lamp can produce high peak powers in pulses, while, arc lamp is operated in a continuous wave mode. Although arc and flash lamp sources offer high intensity but much of their emission is not absorbed by the neodymium ions.

For lamp pumping, hollow reflective cavities such as shown in Figure 8.10, transfer pump light from one or more pump lamps to the laser rod. The first solid state laser was pumped by helical flash lamps placed around the laser rod, as shown in Figure 8.9(a), but this approach

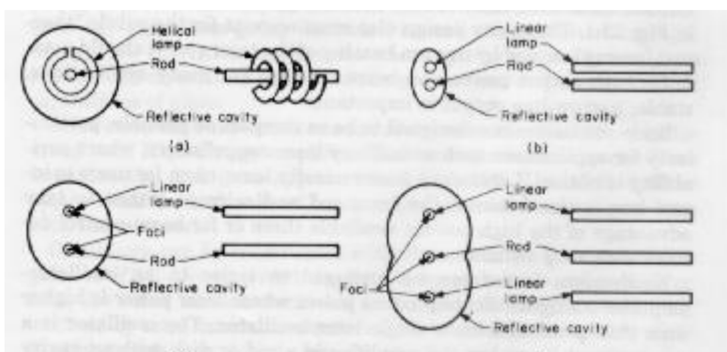


Figure 8.9 Representative cavity configurations for flash lamp pumping (a) helical lamp, (b) closely coupled lamp, (c) elliptical cavity, (d) dual elliptical cavity.

has been supplanted by linear lamps. One modern approach is to place the linear lamp with the laser rod in a closed coupling configuration as shown in Figure 8.9(b). Another is to put the lamp and the laser rod at the two foci of an ellipse, so the geometry efficiently focuses the pump light onto the rod, shown in Figure 8.9(c). Two lamps and a rod can be put into a dual elliptical cavity, shown in Figure 8.9(d), which in cross section looks like two overlapping ellipses with the rod at the common focus.

High power oscillators or amplifiers may be surrounded by arrays of many flash lamps. As the pumping flash last for only a short time (~1 ms) the laser output is in the form of a pulse, which starts about 0.5 ms after the pumping flash starts. This represents the time for the population inversion to build up. Once started, stimulated emission build up rapidly and thus depopulates the upper laser level much faster than the pumping can replace the excited atoms so that laser action momentarily stops until population is achieved again. This process then repeats itself so that the output consists of a large number of spikes of about 1 μs duration with about 1 μs separation.

Diode Laser Pumping

Pumping with diode lasers has become an important technology from mid 1980s; because it offers much higher overall efficiency than lamp pumping. The higher efficiency means that there is less waste heat to remove, allowing smaller packages. GaAlAs semiconductor lasers which emit near 800 nm have been increasingly attractive pump sources.

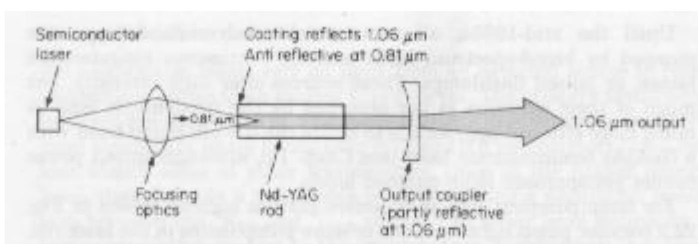


Figure 8.10 End pumping of an Nd-YAG rod with an 810 nm diode laser Diode laser

pumping can be from the end of the rod, as shown in Figure 8.10 or from the side. In end pumping, standard at low power levels, optics focus the rapidly diverging beam from the diode laser to fill as much space as possible in the crystal, which strongly absorbs at the diode laser wavelength near 810 nm. Multiple diode lasers can be arranged along the side of the rod for side pumping for higher powers.

Lamp pumped-neodymium lasers normally operate in multiple longitudinal modes. Diode-pumped-lasers, with much shorter cavities, have broader longitudinal mode spacing and can more easily be limited to oscillate on a single longitudinal mode.

Typical beam divergences from diode-pumped neodymium lasers diameters are 1 to 10 mrad at 1.06 μm. Beam diameters are 0.2 to 2 mm.

8.4.1.3 Energy Level Diagram

Neodymium laser is a four level laser system. In such a laser, the lower level lies far above the ground state and is generally empty. Let n_1 and n_2 denote the number of atoms in the upper and lower laser levels per unit volume, respectively. As $n_1 = 0$, so any population in level 2 gives rise to an inversion with $n > 0$. Such lasers are much more efficient than three level lasers. If level 1 decays to level 0 rapidly enough, i.e., $n_1 \rightarrow 0$ under all conditions and thus four level laser do not absorb (in general) the laser light itself.

When the ion is placed in the host, the crystal field splits some of the energy levels. A simplified energy level scheme for Nd:YAG laser is shown in Figure 8.11. The actual energy

level kinetics are quite complex. The levels shown in Figure 8.11, arise from transitions of the three inner shell 4f electrons of the Nd^{+3} ion in the field of YAG crystal. Since these

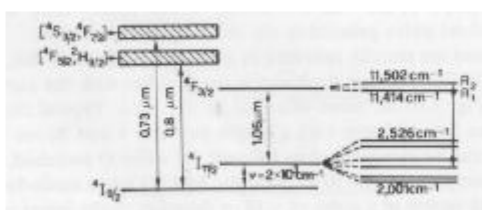


Figure 8.11 Simplified energy levels of Nd:YAG.

electrons are screened by eight outer electrons ($5s^2$ and $5p^6$), the energy levels involved are only weakly influenced by the crystal field and the corresponding transition frequencies are relatively sharp. The two main pump bands occur at 0.73μm and 0.8 μm, respectively; also other higher lying absorption bands play an important role. These bands are coupled by a fast ($\sim 10^{-7}$sec) nonradiative decay to the metastable level $^4F_{3/2}$. So the optically excited neodymium ions quickly decay to this metastable level by releasing their excess energy to the crystalline lattice. From $^4F_{3/2}$, they decay to lower I levels as shown in Figure 8.11. The rate of this decay ($^4F_{3/2} \rightarrow$ I levels) is much slower, i.e., the life time of this state is quite large($\tau \cong$

0.23 ms). The reason for this long life is that the $^4F_{3/2}$-level with a large energy gap with its nearest lower energy level that forbids the non-radiative decay and for non-radiative decay, quantum mechanical selection rules do not allow electric dipole transition to the lower lying levels. This mean that level $^4F_{3/2}$ accumulates a large fraction of the pump energy and is therefore a good candidate as the upper level for laser action. Of the various possible transitions from $^4F_{3/2}$ to lower lying levels, it turns out that the $^4F_{3/2} \rightarrow {}^4I_{11/2}$ is the strongest transition. Level $^4I_{11/2}$ is then coupled by a very fast (ns) non radiative decay to the $^4I_{9/2}$ ground level. The $^4I_{9/2}$ level can thus be considered as a favorable candidate as lower laser level for laser action.

It should also be noted that laser action can be obtained, using suitably dispersive systems in the laser cavity such as those of Figure 8.12 at several other wavelengths corresponding to various $^4F_{3/2} \rightarrow {}^4I_{11/2}$ transitions (λ = 1.05-1.1 μm), various $^4F_{3/2} \rightarrow {}^4I_{13/2}$ transitions (λ = 1.319 μm) and $^4F_{3/2} \rightarrow {}^4I_{9/2}$ transition (λ around 0.95 μm). It is also worth recalling, that the λ = 1.06 μm laser transition is homogeneously broadened at room temperature owing to interaction

with lattice phonons. The corresponding width is $\Delta\nu = 6.5\ cm^{-1}$ (or 195 GHz) at T= 300 K. The long lifetime of the upper laser level (τ = 0.23 ms) also makes Nd:YAG very suitable for Q-switched operation.

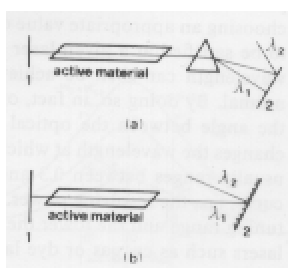


Figure 8.12 Laser tuning using the wavelength dispersive behavior of (a) a prism, (b) a diffraction grating.

8.4.1.4 Nd-Laser Resonators

The gain in neodymium lasers is high enough that they can use either stable or unstable resonators. An unstable resonator has the advantage to utilize the maximum volume of the laser rod for stimulated emission. Close to the laser source, some unstable resonators produce beams with a bright ring around a central point of minimum intensity, which looks like a ring or doughnut in cross-section. However, far from the laser, the hole vanishes to produce a bright central spot. A stable resonator can produce the standard Gaussian TEM_{00} beam with a bright central spot, but it does not extract laser energy from a large fraction of the rod volume. Unstable resonators have been going in popularity because for most applications, output power and energy are more important than near field beam quality.

Three simple resonator designs for lamp-pumped lasers are shown in Figure 8.13. Many lasers have more complex designs, which incorporate interacavity polarizers, beam splitters, or other schemes to couple light out of the cavity.

As solid-state laser cavities are designed to pump and extract maximum energy from the laser medium. Thus the resonator mirror should confine light to oscillate through most of the rod, as shown in Figure 8.13. The cavity should be designed such that it should maximize both output power and beam quality.

Some resonators have a very compact design as possible, particularly for applications such as military laser range finders, where portability is critical. Laboratory lasers usually leave room for users to insert accessories between the laser rod and mirrors.

Solid state laser materials are not always used in the form of cylindrical rods. Disk geometry

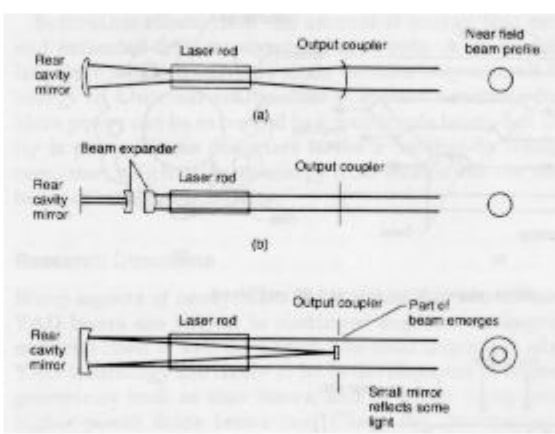


Figure 8.13 Three simple resonant cavity used with lamp pumped Nd:YAG laser: (a) a simple

stable resonator to generate TEM_{00} mode; (b) stable resonator with internal telescope, which helps in generating low divergence beam; (c) unstable resonator with output coupled around a small mirror producing a near field beam with a null at its center. In the far field all three beams have central bright regions.

can also be used, shown in Figure 8.14. Use of thin disks rather than thick chunks of laser glass minimize thermal problems with the material.

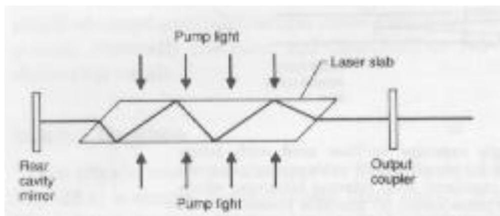


Figure 8.14 A zigzag slab laser.

8.4.1.5 Output Power

A wide range of output power is available from Nd-YAG lasers, depending on the laser configurations, pumping source, and wavelength. Steady or average powers in the kilowatt range can be obtained either from pulsed or continuous wave lasers with the highest power from multimode, oscillator-amplifier configuration and slab designs.

Peak powers available from neodymium lasers are much higher and depend on both the pulse energy and duration. Neodymium lasers can operate in a variety of pulsed modes, depending both on excitation and on control of the energy by Q-switching, modelocking, etc.

Saturation effects limit the amount of energy that can be stored in and extracted from neodymium laser rods. A resonator, which oscillates in a single transverse mode, extracts only a small fraction of the energy in a normal rod because it produces a narrow diameter beam. More power can be extracted in a multimode beam; but the beam quality is poorer. Unstable resonators can be used to extract energy from most of the rod with reasonable beam quality.

Peak power in an emitted pulse of Nd-laser tends to decrease as repetition rate increases beyond a certain point. In general, the higher the peak power, the longer the interval needed between pulses to dissipate excess heat that otherwise might distort the optical properties of the rod and degrade the beam quality of the output beam.

A laser, which is pumped by a flash lamp (with an input energy of about 10 kJ), may produce a laser pulse of about 10 Joules. As the pulse lasts for only 0.5 ms, the peak power is then $2x10^4$ W, this peak power can be greatly increases by Q-switching. The pulses of 10 ns with the peak power up to 10^9 watts can be achieved using Q-switching techniques.

8.4.1.6 Oscillator-Amplifier Configurations

Nd-glass can be arranged in oscillator-amplifier configuration to produce pulses whose laser power is higher than that possible with a single laser oscillator. The oscillator is a conventional laser, but the amplifier is a rod or disk without cavity mirrors, which is separately pumped. A typical oscillator-amplifier configuration is shown in Figure 8.15. The oscillator and amplifier need not use the same host material but the host wavelengths must be close enough that the oscillator's wavelength falls within the gain bandwidth of the laser amplifier. In high energy, low repetition rate systems, oscillators are Nd:YAG but the amplifiers typically are glass rather than crystalline because of (i) the high quality glass laser rods can be made in about any diameter to allow high power with relatively low power density and (ii) Glass is highly resistant to damage from higher power density.

Care must be taken in matching the wavelengths of oscillators and amplifiers made of different materials. Nd:YAG (λ = 1.064 μm) oscillators can be used with silicate glass (λ = 1.062 μm) amplifiers.

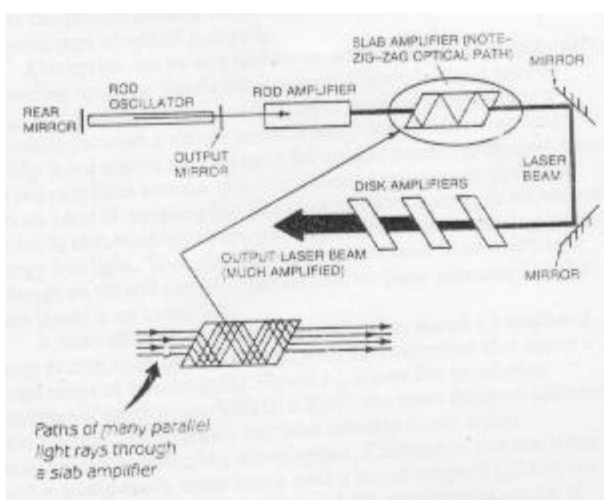


Figure 8.15 Laser oscillator and amplifiers.

Multiple amplifiers may be used where extremely high laser powers are required, such as in laser fusion experiments. A system based on Nd-Glass laser amplifiers, delivering pulses with energy of ~ 100 kJ and peak power of 100 TW, has already been built. These lasers make use of a chain of several Nd-glass amplifiers, one of which consists of Nd-glass disk of ~ 4 cm in thickness and ~ 75 cm in diameter.

8.4.1.7 Efficiency

Overall efficiency of a commercially lamp-pumped neodymium laser, measured as laser power output divided by electrical power in, typically is 0.1 to over 1 percent. Diode-pumped lasers can be considerably more efficient because neodymium absorbs the 0.8-μm pump beam much more efficiently than the broad-band output of a lamp. The optical losses contribute to the low efficiency of lamp-pumped lasers. They account for the following typical fractions of the input energy to a continuous lamp:

- Only half the input energy emerges as pump light at 0.3 to 1.5 μm; the lamp dissipates the other 50 percent as heat.
- Only 8 percent of the original energy is absorbed by the laser rod. The laser cavity absorbs 30 percent, the coolant and flow tubes absorb 7 percent, and the lamp re-absorbs 5 percent.
- Only 2.6 percent of the lamp input energy goes into stimulated emission from the laser rod, 5 percent is lost as heat and 0.4 percent is lost as fluorescence.
- Optical losses reduce the typical overall efficiency to the 1 percent range. The fractions differ somewhat with design details.

8.4.1.8 Cooling

A large amount of heat is dissipated by the flash lamp and consequently the laser rod quickly becomes very hot. To avoid the damage resulting from this, and to allow a reasonable pulse repetition rate, cooling arrangements are necessary for solid-state lasers. For low-power lamp or diode pumped YAG lasers, connective air-cooling is sufficient. However this limits operation to low repetition rates. For high power lasers, it is necessary to use forced air cooling or water cooling.

8.4.1.9 Safety

Pulsed neodymium lasers probably have been responsible for more eye injuries than any other type, both because of their wide spread use and because of the particular hazards they pose.

Although the human eye can not see the 1.06 μm fundamental wavelength of the neodymium laser, that light can pass through the eyeball to damage the retina. The same considerations apply to the 1.3 μm neodymium line. The risk of eye damage due to an unseen beam is increased by the fact that most neodymium lasers used in the laboratory produce short, intense pulses. The best protection against such accidents is safety goggles which effectively block all optical paths to the eye, transmitting most visible light by strongly blocking the near infrared neodymium lines.

Harmonics of neodymium lasers also come in short, intense pulses and present serious eye hazards. The second harmonic at 532 nm is visible green light, and can penetrate the eye. It requires safety goggles, which strongly attenuate the green part of the spectrum.

The 355 nm third harmonic and the 256 nm fourth harmonic pose eye hazards. The fourth harmonic also can cause a sunburn-like effect on the skin. Ultraviolet-blocking safety goggles can provide eye protection against the third and fourth harmonics.

Users should be aware that some light at longer wavelength may remain after harmonic generation unless the optical system is explicitly designed to get rid of it. Some commercial lasers come with optional beam dumps for this purpose. Multiwavelength laser output can pose a severe eye hazard because it is difficult to design safety goggles to block two or more wavelengths.

Continuous wave neodymium lasers pose serious eye hazards, but the power levels are sufficiently low that an instantaneous exposure is less likely to cause permanent damage. Some diode pumped lasers produce continuous beams on the order of a milliwatt at the second harmonic wavelength, but most neodymium lasers generate considerably high powers, eye protection is necessary when working with any neodymium laser.

Lamp pumping poses two additional hazards. All lamps require high drive voltages, posing serious electrical hazards if any high voltages are exposed.

We have by no mean covered the entire range of lasers or fully discussed the various modification and refinements of lasers which have been described. It is hoped, however, that the basic laser physics covered together with the descriptions of some laser types will enable the reader to understand the mode of operation of other lasers which might be encountered or which might be developed in the future.

Problems

8.1 Which major components are required for a laser system? Discuss the function of each component.

8.2 Discuss in detail the pumping mechanism for a He-Ne laser.

8.3 Explain how a He-Ne laser can be operated in green region?

8.4 With the help of a graph, discuss the effect of changing transmitivity of out-put mirror on the output power of a He-Ne laser.

8.5 Calculate the length of the cavity for a He-Ne laser, which is to be operated in a single longitudinal mode.

8.6 With the help of energy level diagram of a liquid dye laser, justify the statement: “Dye lasers are tunable”.

8.7 Describe the effects of T_1 and T_2 states on lasing action in a dye laser. How these effects are eliminated or minimized?

8.8 Give a comprehensive comparison of different optical pumping sources which can be used for a solid-state laser.

8.9 What is oscillator amplifier configuration? Give a schematic diagram of this configuration for Nd-YAG laser. Explain it in detail.

8.10 Discuss the factors, which cause inefficiency of Nd-YAG lasers.

Books for further reading

A. E. Seigman, *Lasers*, (University Science Books, California, 1986).

O. Svelto, *Principles of Lasers*, 3rd ed. (Plenum Press, New York, 1989).

P. W. Milonni and J. H. Eberly, *Lasers,* (John Wiley & Sons, New York, 1991).

J. Wilson and J.F.B. Hawkes, *Optoelectronics: An Introduction*, (Prentice-Hall, India, 1996).

J. Hecht, *Laser Guide Book*, 2nd ed. (McGraw-Hill, New York, 1992).

B. A. Lengyel, *Introduction to Laser Physics*, (John-Wiley & Sons, New York, 1966).

K. Shimoda, *Introduction to Laser Physics* , (Springer-Verlag, New York, 1984).

Unit-9

Laser Applications

Objective

Lasers are widely used in many types of industrial processing, engineering, metrology, scientific research, communications, holography, medicine, and for military purposes. It is clearly impossible to give an exhaustive survey of all of these applications and the students are referred to the selection of books. Rather than attempt the impossible, we gave a list of common applications and discuss a few one in this unit.

9.1 Introduction

Laser sits near the top of any list of the greatest scientific inventions of the 20th century. Like other technologies such as satellite, computer, integrated circuits, fiber optics etc., laser is also a symbol of high technology. Due to its high intensity, coherence, directionality and monochromaticity, it has applications in almost all branches of science. Laser can be used in reading and writing information. It can read computer data on optical disks. We can have computer output using laser printers. It can mark identification codes. Lasers can help in cutting, drilling and welding of plastics, metals and other materials with accuracy unimagined before. Drilling of the materials from diamonds to baby bottle nipples is possible using lasers. Laser can engrave wood. It can measure small distances and can make ultra precise measurements. We can study atomic and molecular physics using laser.

In the field of medical science, lasers can be used for eye operations, blood clots treatment, brain surgery and surgery of the various parts of the body with minimal chances of infection. Lasers can illuminate our body cells for biomedical measurements. It can shatter kidney stones.

Lasers have many applications in military & defense field such as in anti-sensor weapons, in anti-satellite weapons, in antimissile weapons, in battle simulation, in simulating the effect of nuclear weapons. Lasers can measure the range to distant objects. It can pinpoint the targets with accurate distance measurements and designations for bombs and missiles.

Laser can make holograms: a three dimensional photograph of an object. Holography in turn provides a new technique of non-destructive testing of objects of complicated shapes. We can have laser light shows for entertainment. It can control chemical reactions. It can change concentration of the desired isotopes. It can produce nuclear fusion. It is widely used in modern research etc.

Laser beams when focused at a point can generate the temperature, which exist in the core of the sun. As a result, fusion of hydrogen atom has been done. It is expected that lasers will play a major role in future fusion reactors. The well-known Star War Program of US, meant to destroy distant target from space, was also based on lasers. In the field of entertainment, lasers have contributed immensely. Laser beams shine in smoke and dance with music in laser shows known as Laserium. In the field of communications, lasers coupled with fiber optics are changing the whole concept of cable laying in town planning.

In below a list of laser applications is given:

Reading and Writing Information

Playing audio compact discs

Playing videodisks

Reading Universal Product Code in stores

Reading computer data on optical disks

Laser printers for computer output

Measurement and Inspection

Projecting straight lines for construction alignment and irrigation

Measuring the range to distant objects

Measuring small distances very precisely

Illuminating cells for biomedical measurements

Laser induced fluorescence measurements

Studies of atomic and molecular physics

Measuring concentration of chemicals or pollutants

Medicine

Treatment of diabetic retinopathy to forestall blindness

Laser surgery

Laser bleaching of port wine stain birthmarks

Shattering of kidney stones

Treatment of infertility

Materials Working

Cutting, drilling, and welding plastics, metals, and other materials

Cutting cloth

Engraving wood

Marking of identification codes

Military Applications

Laser range finders

Fire control systems

Laser guidance

Anti-sensor weapons

Anti-satellite weapons

Anti-missile weapons

Battle simulation

Simulating effects of nuclear power

Other Applications

Making holograms

Laser light shows

Displays

Laser pointers

Controlling chemical reactions

Changing concentrations of isotopes producing nuclear fusion

Basic research

New applications of lasers continue to be found. In a single chapter we cannot describe all the above mentioned applications. However, we will discuss applications of lasers in holography, Doppler anemometers and medical applications.

9.2 Holography

Light is an electromagnetic radiation, characterized by its amplitude, wavelength, phase, polarization, speed and direction of propagation. When light is scattered or reflected by an object or passes through a transparent medium, a few or all of its characteristics get change. Light detectors like eye, photographic materials and photomultiplier tubes are square-law detectors, which respond only to the average intensity of light. None of these can record directly the spatial distribution of amplitude and phase of the electromagnetic field of light wave.

Holography provides an opportunity of recording the information about the object in the form of phase and amplitude distribution simultaneously on a photographic plate and reconstructs the three-dimensional image of the object. The word holography is derived from Greek roots meaning, *complete writing*, as it provides complete information about an object in its all-visual properties, including the realism of the three dimensions. Holography is now established as a display medium as well as a tool for researchers.

Denies Gabor (1948) showed that *if one allows highly coherent light to be diffracted by an object and the light from a coherent reference wave to interfere, the resulting interference pattern would contain all the information that needed to reconstruct the object completely*. In the presence of weakly coherently sources available at that time, Gabor demonstrated the validity of the concept of holography.

9.2.1 Difference of Holograms and Photographs

In holography, the interference pattern formed by the superposition of reference beam and object beam is recorded on a photographic film. After developing process, when this film is illuminated with a reference beam, it generates the image of the object in all of its perspective. Holograms and photographs are different in many ways.

1. When a photograph of a scene is taken, firstly a negative is made then the dark and bright portions of this negative are reversed to form a positive print, which portrays the scene as original one. In the case of holograms, both negative and positive versions generate identical three-dimensional image.
2. If a hologram is divided into pieces, all pieces usually give to a complete information in all respects about the object, which is not possible by a photograph.
3. A hologram appears to be a uniform gray sheet and hardly any pattern is seen, which is contrary to the photographic negative.
4. A hologram possesses the property to reconstruct the light wave coming from the object.
5. Holography allows several hundred of pictures to be recorded in an emulsion, while an ordinary photography is capable of storing only one image.

9.2.2 The Principle of Holography

Holography is the art of lens-less, three-dimension photography. Dennis Gabor described the essential principle of holography. Before the development of laser it was difficult to make holograms. With the advent of laser, it was quickly recognized to be practical, and holograms since then become a whole field of research with many applications. In this section we will take up the basic idea underlying the holographic process.

Consider the light scattered by some object illuminated by a monochromatic wave. In some plane (say z = 0) we may represent the electric field of the scattered light as shown in Figure 9.1.

$$E_s(x,y,z,t) = E_{so}(x,y).\, e^{-i\omega t} \qquad (9.1)$$

The subscript, s, labels this field as due to the scattered radiation from the object, and E_{so} is the spatial component of the time varying electric field E_s. The complex amplitude $E_{so}(x, y)$ defined over the plane z = 0 (Figure 9.1) may be written as

$$E_{so}(x,y) = A_o(x,y).e^{i\phi_o(x,y)} \tag{9.2}$$

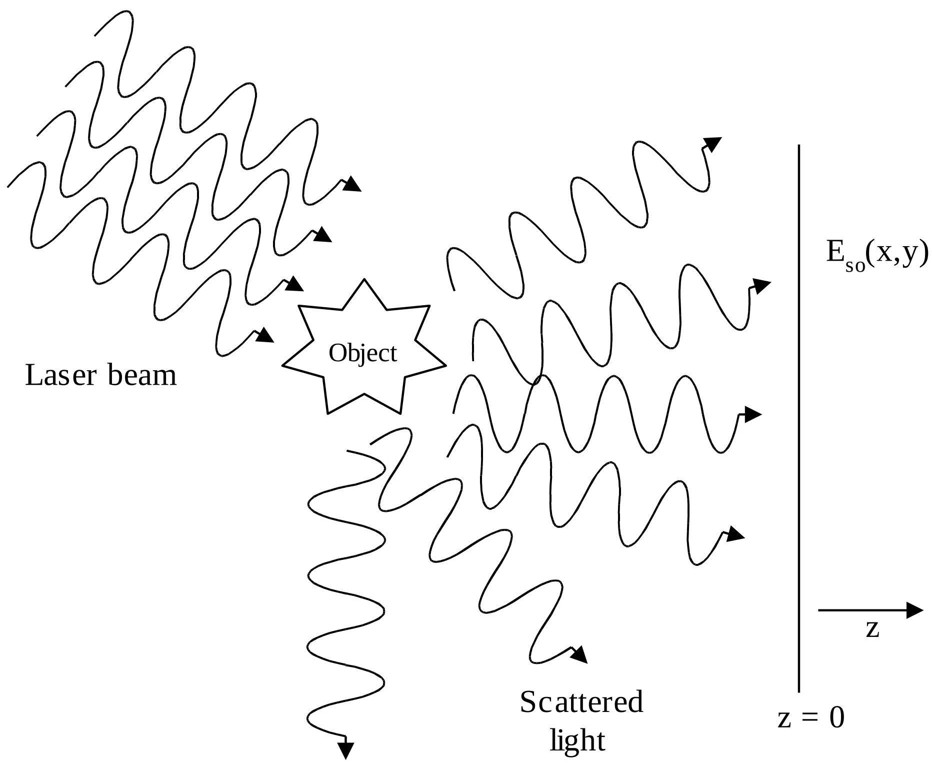


Figure 9.1 The light scattered by an object is represented in a plane z = const. by the electric field $E_{so}(x,y)$.

where $A_o(x, y)$ and $\phi_o(x, y)$ are the (real) amplitude and phase, respectively, in this plane. The real electric field defined by equations (9.1) and (9.2) is given by

$$E_s(x,y,t) = A_o(x,y).\cos[\omega t - \phi_o(x,y)] \tag{9.3}$$

The intensity of the field (9.3), averaged over a number of optical periods is proportional to $(A_o)^2$:

$$I_o(x,y) = k.A_o^2(x,y) \tag{9.4}$$

Except for the multiplicative constants, k, this is the intensity distribution measured by developing photographic film placed in the observation plane of Figure 9.1. Such a photograph is only a flat two-dimensional record. This record of the object does not look exactly like the real-life object because it registers only the intensity, and not the phase, of the object field. This is where a hologram of the object differs from a photograph, i.e., a hologram contains information about the intensity and the phase of the object field.

Suppose that, in addition to the object field (equation 9.1), we shine some reference beam

$$\begin{aligned} E_r(x,y,t) &= E_{ro}(x,y).e^{-i\omega t} \\ &= A_r(x,y).e^{i\phi_r(x,y)}.e^{-i\omega t} \end{aligned} \tag{9.5}$$

on the observation plane of Figure 9.1, where $A_r(x,y)$ and $\phi_r(x,y)$ are the amplitude and phase of this reference field. The total amplitude at the observation plane is

$$E_t(x,y) = A_o(x,y).e^{i\phi_o(x,y)} + A_r(x,y).e^{i\phi_r(x,y)} \tag{9.6}$$

and the total intensity is $I_t = k.\,(E_t)^2$, where

$$\begin{aligned} \left|E_t(x,y)\right|^2 &= A_o^2(x,y) + A_r^2(x,y) \\ &+ 2A_o(x,y).A_r(x,y).\cos\left[\phi_o(x,y) - \phi_r(x,y)\right] \end{aligned} \tag{9.7}$$

Because of the interference of the object and reference waves, this intensity, unlike equation (9.4), does contain information about the phase $\phi_o(x,y)$ of the object field. The recording of this intensity distribution on photographic film is the first step of the holographic process. This intensity distribution does not look like the object, to obtain a fully three-dimensional image of the object we must reconstruct the wave front of equation (9.2). This reconstruction is the second step of the holographic process. The challenge is to use the intensity distribution (of equation 9.7) on a developed film to reconstruct the object wave front of equation (9.2).

Suppose an intensity distribution (like Eq. 9.7) is recorded on a photographic film. Exposure of the film to this intensity results in the conversion of silver halide grains in the film emulsion to silver atoms; the density of the silver atoms is greatest where the intensity of the incident light was greatest. Development as a photographic positive reverses the situation: the positive is lightest where the incident intensity was greatest. The transmission function t(x,y) of the positive is therefore largest where the intensity of the incident light was high. In fact, we may assume that t(x,y) is approximately proportional to the intensity I(x,y) of the light to which holographic film was exposed. Then if the positive is illuminated with a wave front E(x,y), the transmitted wave front is

$$\begin{aligned} E_{trans}(x,y) &= t(x,y).E(x,y) \\ &= C.I(x,y).E(x,y) \end{aligned} \tag{9.8}$$

where C is the constant of proportionately between t and I. Now suppose that

$$E(x,y) = E_r(x,y) \tag{9.9}$$

and

$$I(x,y) = I_t(x,y) \tag{9.10}$$

where E_r and I_t are given by equation (9.5) and (9.7), respectively. That is, the original intensity distribution to which the film is exposed is the total intensity from the object and reference fields, and after development the incident field E(x,y) is that of the reference field alone. Then from equation (9.8) we have

$$\begin{aligned} E_{trans} &= C.\left[A_o^2 + A_r^2 + 2A_o.A_r.\cos(\phi_o - \phi_r)\right].A_r.e^{i\phi_r} \\ &= C\left[A_o^2 + A_r^2 + A_o.A_r.e^{-i(\phi_o-\phi_r)} + A_o.A_r.e^{i(\phi_o-\phi_r)}\right].A_r.e^{i\phi_r} \\ &= C\left[\left(A_o^2 + A_r^2\right).A_r.e^{i\phi_r} + A_o.A_r^2.e^{i(2\phi_r-\phi_o)} + A_o.A_r^2.e^{i\phi_o}\right] \end{aligned} \tag{9.11}$$

The important part of the transmitted field is the last term of this expression:

$$C.A_r^2(x,y)\left[A_o(x,y).e^{i\phi_o(x,y)}\right] = C.A_r^2(x,y).E_{so}(x,y) \tag{9.12}$$

where E_{so}(x,y) is the object wave front (equation 9.2).

Thus the transmitted field is proportional to the object wave front E_{so}(x,y). That is, we have managed to reconstruct the complete object wave front, including both amplitude and phase. The first term in the equation (9.11) represents the direct beam and the middle term represents the other diffracted beam, which forms the real (conjugate) image.

9.2.3 Construction and Reconstruction of Holograms

Figure 9.2 and 9.3 illustrate the basic setup for construction and reconstruction of the hologram. A holographic plate is exposed simultaneously to waves of light scattered by the *object* and to waves of light from a *reference* source. The reference beam, shown in Figure 9.2 as a plane parallel beam, is derived from the same laser source as the light illuminating the object. Because of their high degree of mutual coherence, the two sets of waves produce an interference pattern on the plate, which is recorded in the photographic emulsion and forms a hologram.

The holographic plate is now processed and illuminated with only the reference beam present as shown in Figure 9.3. Most of the light from the reference beam passes straight through the

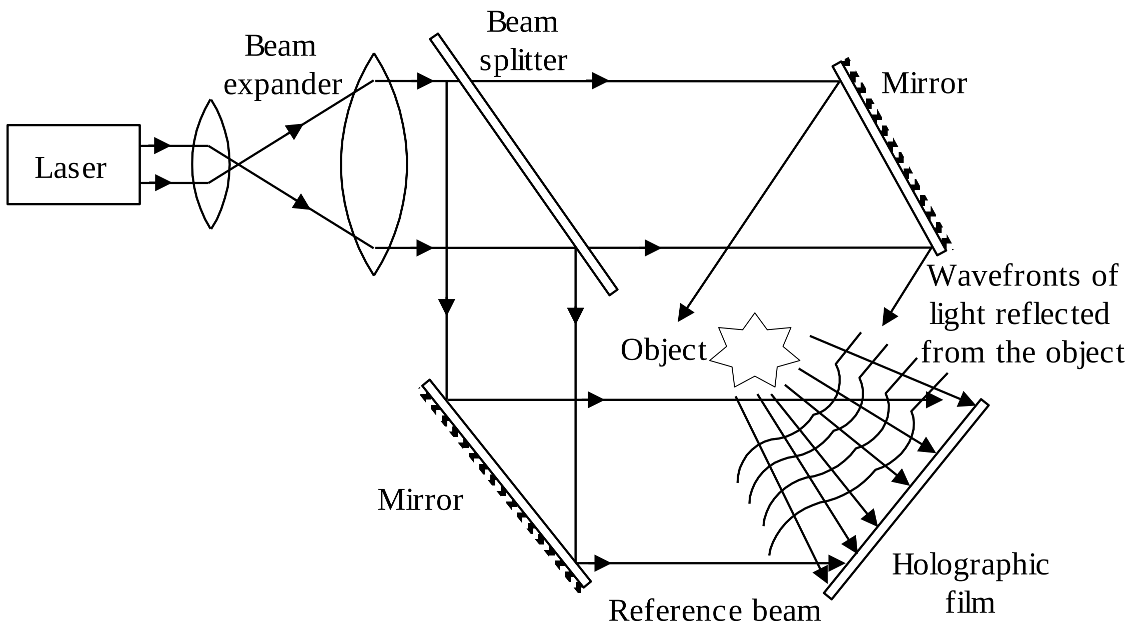


Figure 9.2 Schematic disgram showing the basic geometry for making a hologran by recording the interference pattern produced by the interference of the objec and reference wavefronts.

hologram; some of it is diffracted by the interference pattern in the emulsion. By the normal diffraction grating, light of wavelength λ will experience constructive interference at angles such that $\lambda = d.\sin\theta$, where d is the local fringe spacing of the interference fringes whose exact shape and distribution depends on the shape of the object and the wave fronts reflected from it. Thus the constructive interference of these diffracted waves reconstructs the original wavefronts from the object and to an observer the wavefronts appear to be coming from the object itself. These wavefronts constitute the virtual image. However, just as a diffraction grating gives diffracted orders on either side of the straight through position, the hologram generates a second image, which is a real image and can be taken on a screen.

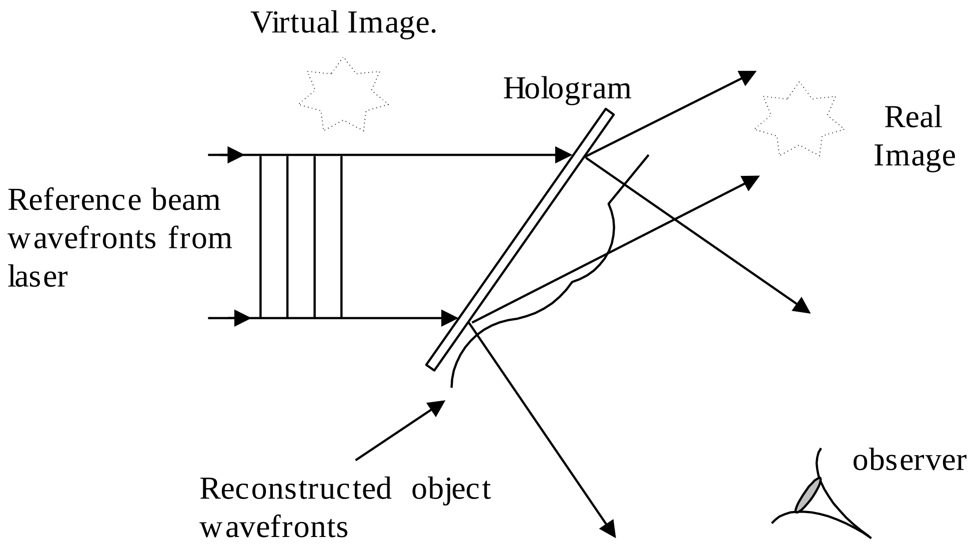


Figure 9.3 The figure shows the reconstruction of the object wavefronts. To observer the reconstructed wavefronts appears to be coming from the object it and he sees the virtual image.

Holography has found applications in a number of areas, and new uses are being developed all the time. One of the applications is holographic interferometry, which is a technique for measuring small displacements. The technique can be used to observe and analyze, either in real time or after some delay, the microscopic flexure of loaded support beam, to observe the strain in a fractured bolt, to record the shock wave from a bullet, or to detect the hidden flaws in an aircraft wings, tires, etc.

9.3 Laser Doppler Anemometer

With the invention of laser in 1960 it was natural that people should consider measuring the velocity of a material by means of the *Doppler effect*. This is a well-known effect in acoustics and was first discovered in the nineteenth century by a German physicist, Doppler. As an observer moves towards a sound source, he encounters the waves more frequently than he would if he was stationary. The source therefore, appears to have a higher pitch. Conversely, if an observer moves away from the source it appears to have a lower pitch. A similar effect occurs if the source is moving and the observer is stationary. The Doppler effect is often heard from fast moving trains. The pitch of the noise suddenly drops as the train passes.

In his Special Theory of Relativity, Einstein showed that a similar Doppler effect could occur with light waves. This Doppler effect can be used to measure the velocity of a body, which scatters the light. Let us consider light sent out from a source to a body moving towards it. The body will see light of a higher frequency than that initially emitted (blue shift). When the body scatters the light, it re-emits light of the same frequency as it receives. An observer at the source will then see the light at a still higher frequency since the body is moving towards him. The fractional shift in frequency is approximately 2v/c, where v is the velocity of the body and c is the speed of light. Thus a body moving at 1 m/s and the speed of light of $3x10^8$ m/s, the frequency shift is 1 part in 150 million. The frequency of light waves is approximately 500 terahertz ($5x10^{14}$ Hz). Thus a frequency shift of 1 part in 150 million corresponds to a shift of 3.3 MHz which is quite a sizeable frequency.

We take the Doppler shifted signal, mix it with the original laser beam and pass them both through a photodetector. The output of which will consist of a d.c. part due to the total light and an a.c. part alternating at the difference frequency of the two signals. If the reference beam has the same frequency as the original laser, then this frequency will be exactly the Doppler shift. Thus the heterodyne technique enables one to measure small frequency shifts very easily. All that one has to do is to measure the frequency of the alternating component of the photodetector output.

Yeh and Cummins were the first who measure the velocity of a fluid using the Doppler shift of laser light. In their experiment, output from a He-Ne laser is splitted into two beams by a half-silvered mirror. One beam is passed through the fluid, while the other beam falls on the detector as a reference beam. The beat signal from the light scattered of the fluid and the reference beam was detected on a spectrum analyzer. Yeh and Cummins used this laser velocimeter to measures the velocity profiles of laminar flow through the flow tube. Very low velocities could be measured with this instrument, i.e., as low as 0.007cm/s, corresponding to a Doppler shift of 17.5 Hz.

The single-beam laser anemometer was first used by Foreman et al. In this technique one beam from the laser is focused upon the fluid whose velocity is to be measured. Part of the light is scattered to produce the signal beam and part is transmitted to produce the reference beam. This system was used to measure both the laminar flow of a liquid and the flow of air. Sufficient signal strength was obtained without adding any scattering particles to the liquid, but smoke particles were added to the air. The instrument was shown to be practical tool for use in wind tunnels.

In a dual-beam laser anemometer, the laser beam is divided into two using a beam splitter before the beams enter into the fluid to be measured (Figure 9.4). The two beams are deflected by mirrors and focussed onto the same region of fluid. One of the beams, the reference beam, passes directly through the fluid and falls upon a photomultiplier. The other

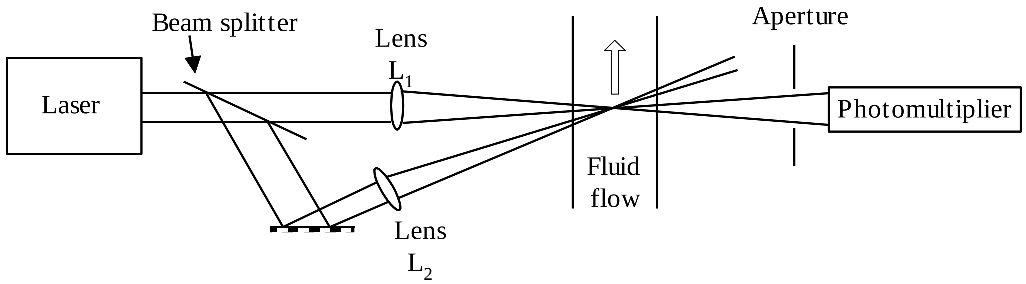


Figure 9.4 Dual-beam laser anemometer.

beam, the signal beam, is partially scattered by the fluid. Now some of this scattered light will travel in the same direction as the reference beam. Beat frequency and the angle between the two beams gives the information of the fluid velocity. This system is much easier to align than the single-beam laser anemometer.

9.3.1 Experimental Setup for A Laser Doppler Anemometer

A common experimental setup for laser Doppler anemometer is shown in Figure 9.5, which is a refined form of the dual beam anemometer mentioned earlier. In this arrangement light

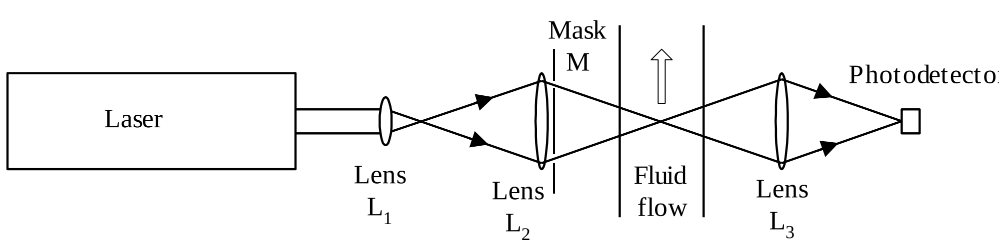


Figure 9.5 Experimental setup for laser Doppler anemometer.

from a laser is spread by a microscope objective lens L_1, L_2, to cover a mask M. This mask contains two slits which produce two beams. The beams are produced by a single optical component and are focused by the same component, therefore, these should be focused on the same volume, as long as the lens is free from aberrations. Thus the alignment is automatic and inherently unaffected by vibration or thermal changes.

Since the two beams are equivalent (either may be the reference beam with the other being the signal beam) they are both collected by a lens L_3 and focused onto a photodetector. The signal is always much stronger and the system is always perfectly aligned. Therefore, a simple semiconductor photodiode can be used to detect the light. Such system has been successfully used to measure turbulent flow in a pipe.

9.3.2 Doppler Shift Caused by a Moving Scattering Center

Laser Doppler velocimeter will only work when scattering centers are present. These may be found naturally in fluids or artificially added. Scattering centers may vary in size from a fraction of a μm to several tens of μm for efficient operation. Too small particles will not produce enough scattering power to exceed the noise level of the detection system. Too large particles, on the other hand, will tend to block the light completely and thus render the instrument inoperative for the forward scattering. Therefore with large particles the system can be used in the back-scattering configuration.

The laser Doppler velocimeter does not measure the velocities of the fluids. They only measure velocities of the contamination in fluids, i.e. scattering centers, which may or may not follow at the fluid velocity. The discrepancy between particle velocities in fluids and fluid velocities may be significant for large particles (10 μm approx.) especially in turbulent flows. The particles as small as 0.1 μm can faithfully follow turbulent flow in liquids such as water.

Consider a plane wave with uniform intensity polarized in one plane with a wave vector **i**k (**i** is the unit vector along the incident beam direction) striking a particle P moving with a vector velocity **v** (Figure 9.6). Some of the incident radiation will be scattered in the direction **s** (here **s** is a unit vector in the scattered direction) towards the photodetector. To detect the change in the frequency of the scattered light we shall bring in the reference beam along the

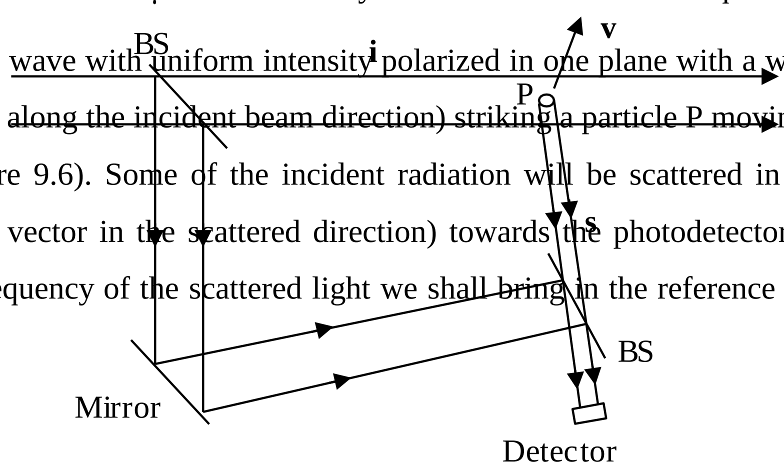


Figure 9.6 Simple arrangement for the measurement of Doppler shif

path using two beam splitters BS. Both semi-silvered mirrors BS must be adjusted so that the reference beam is co-linear with the scattered beam.

Using complex notation the incident plane wave $E_i(t)$ may be represented as

$$E_i(t) = E_{io} \exp(-i\omega_o t) \tag{9.13}$$

The scattered wave $E_s(t)$ falling on the photodetector is then

$$E_s(t) = E_{so} \exp\{-i[\omega_o + k(\mathbf{s}-\mathbf{i}).\mathbf{v}]t\} \tag{9.14}$$

where ω_o is the frequency of the incident radiation, E_{io}, E_{so} are the amplitudes of the incident and scattered radiation, **i** & **s** are the unit vectors in the incident and scattered directions, **v** is velocity of the fluid, and t is time. Thus, $k(\mathbf{s}-\mathbf{i}).\mathbf{v}t$ gives the frequency difference of the incident (which is the reference beam) and the scattered radiation, i.e., Doppler shift ω_D. Here phase differences are assumed to be small enough not to influence the relative coherence between the two beams.

We can calculate the current at a point x on the photodetector, which is proportional to the irradiance of the two beams, and is given by:

$$i(t) = K(E_i+E_s).(E_i+E_s)^* \tag{9.15a}$$

$$i(t) = K(E_{io}e^{-i\omega_o t} + E_{so}e^{-i(\omega_o+\omega_D)t}).(E_{io}e^{i\omega_o t} + E_{so}e^{i(\omega_o+\omega_D)t}) \tag{9.15b}$$

$$i(t) = K(E_{io}^2 + E_{so}^2 + E_{io}E_{so}[e^{-i\omega_D t} + e^{i\omega_D t}]) \tag{9.15c}$$

$$i(t) = K(E_{io}^2 + E_{so}^2 + 2E_{io}E_{so}\cos(\omega_D t)) \tag{9.15d}$$

where $\omega_D = k(\mathbf{s}-\mathbf{i}).\mathbf{v}t$ and K is a constant. The Doppler shift can be represented as

$$\omega_D = (2\pi/\lambda).(\mathbf{s}-\mathbf{i}).\mathbf{v} = 4\pi/\lambda.v_x \sin\alpha \tag{9.16}$$

Here 2α is the angle between the incident and scattered radiation, v_x is the velocity component in the plane of the two beams. Thus by knowing the wavelength of light and incident and scattering directions, the velocity of the fluid can be determined.

9.3 Medical Applications of Lasers

Physicians and surgeons were not far behind engineers and scientists when it came to starting experiments with lasers. The new Lasers offer scientists a novel means of studying individual cells as well as whole organs that provide the foundation of laser medicine.

Today lasers are increasingly important medical tools. They can be considered as "knives" of lights, which can be used to treat individual cells as well as whole organs.

The principal therapeutic uses of laser energy in medicine have to be cut, cauterize (seal-off) or injure certain tissues. Indeed, biomedical applications for lasers include much diverse tasks as unclogging obstructed arteries, breaking up kidney stones, clearing cataracts, destroying red birth marks called port-wine stains, treating lung disorders and even altering genetic materials. Lasers can be safely used in ophthalmology, dermatology, otolaryngology, gastroenterology, photodynamic therapy, acupuncture procedure, angioplasty, gynecological surgery, neurosurgery, burn surgery, microsurgery etc. Laser can also provide information about the interior working of the cells.

9.3.1 Brief History

Conventional light sources, before the invention of laser light, were in use for biomedical applications. In 1946 a German physician, Gerd Meyer Schwickerath, used the sun light to treat detached retinas and to destroy tumors in some of his patient's eyes. The idea of using conventional light in medical opened the way for the lasers to use in medical surgery. In 1961, only a year after T.H Maiman built the first laser, Milton Zaret of the New York University School of Medicine used a laser to produce ocular lesions in animals. Two years later, Chris Zweng of the Palo Alto Medical Research Foundation in California treated retinal disease in his patients for human trials. Having successful results, the laser was quickly accepted as a standard ophthalmic magic tool.

Today laser can be used for a number of medical applications as stated above.

Before explaining some of the laser applications in medical field, first we discuss some drawbacks of scalpel (surgical knife).

9.3.2 Drawbacks of the use of Scalpel

Cutting with a scalpel has been used throughout the history of surgery and is the most common instrument still using in surgical procedures. But the use of scalpel in medical surgery has some drawbacks as:

(i) When a surgeon cuts with a scalpel, bleeding into the incision may impair the accuracy of his work by obscuring his view.

(ii) If the bleeding becomes excessive, the quantity of blood loss may require transfusion, with the risk of hepatitis.

(iii) Each bleeding vessel have to be clamped and tied with sutures and the incision sponged dry to permit continued work. This obviously increases the duration and complexity of the procedure.

(iv) Foreign materials (sutures) have to be left in the wound after surgery. Sutures left in the closed wound may become a nidus for infection or foreign body reaction.

(v) There may be problems in maintaining adequate blood pressure and blood flow to such vital organs as the brain, heart and kidneys.

To overcome the drawbacks of use of scalpel, Laser "knives" can be used to cut tissue. This technique also produces coagulation and cautery by heating cells in the tissue. But in this process, the heating is caused by the absorption of optical radiation by the cells. Since the optical absorptivity of most cells is very high, most of the energy is absorbed very nearly the surface. This means that with laser treatment, few cells are affected and distant tissues are not damaged as explained below.

9.3.3 Removal of Tissue by Laser Surgery

Laser has very high intensity (power/area) and can produce intense heat on a specific spot. Today most laser surgery makes use of this heat primarily because its destructive effects can be extremely selective and precisely controlled. To have an idea about the removal of tissue using lasers, one must know that how laser light interacts with the tissue.

Tissue like any other material absorbs light differently at different wavelengths. The factors to be considered in laser light interactions are given in Table 9.1.

Table 9.1 Factors in Laser-Tissue Interactions.

1. wavelength of laser
2. Optical absorpitivity of tissue at that wavelength
3. Intensity incident on the treatment site (W/cm^2)
4. Duration of the irradiation on the treatment site (sec.)
5. Volume of tissue treated (spot size and penetration depth)

If the wavelength of a laser is matched very closely with the absorption band of the target structure; the laser light will be absorbed by and therefore damage that structure. The penetration depth of the different lasers in skin, liver and stomach is shown in Table 9.2.

Table 9.2 Penetration Depth in Skin, Liver and Stomach Tissue

Laser	Penetration Depth (mm)
CO_2 at 10.6 μm	0.05
Nd:YAG at 1.06 μm	0.8
Argon at 0.5 μm	0.2

The exact wavelength at which a tissue absorbs light strongly depends on the composition of tissue. Water is the most important component of a tissue, which absorbs light in the infrared region. As water is the major component of a tissue, so the tissue can be considered as absorbing light like water, especially in the infrared region. It means a laser having wavelength in the infrared region can be proved useful to cut or cauterize a tissue. Carbon dioxide (CO_2) laser is the one, which fulfills this requirement because it emits light of wavelength 10.6 μm (in the infrared region). Water reflects about 10% of the incident CO_2 laser light and absorbs about 80% in the first 20 micrometers thickens. As a result, CO_2 laser became a favorite tool for laser surgery because it is readily available and absorbed strongly by water and tissue.

The CO_2 laser beam is scanned across the tissue to be removed as shown in Figure 9.7. Water, the

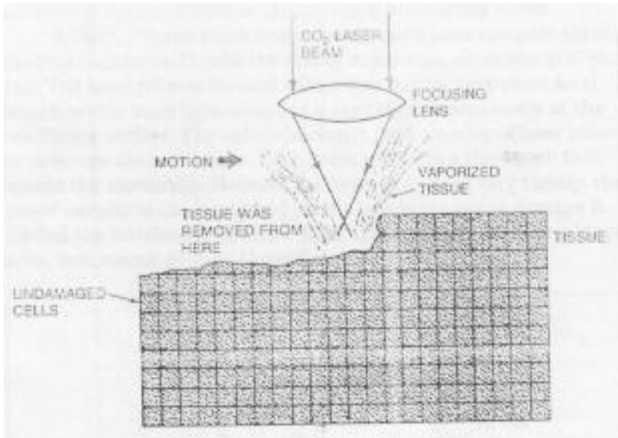


Figure 9.7 A CO_2 laser removes the top layer of cells with minimal damage to those underneath.

major component of the cells, absorbs light beam of the wavelength 10.6 μm strongly, with the result that top layers of cells are vaporized but doing little damage to the underlying tissue. This process also can stop bleeding effectively by sealing off (cauterizing) the blood vessels smaller than about a millimeter in diameter.

Following are the useful aspects of laser-tissue interaction:

(i) The laser energy will vaporize the target cells only and causes no damage (or little) to the other cells.

(ii) CO_2 laser penetrates just deep enough into tissue to seal small blood vessels and stop bleeding. This makes the CO_2 laser as especially valuable tool for surgery in regions rich in blood vessels such as the female reproductive tract.

(iii) CO_2 laser's ability to remove thin layers of blood rich tissues gives it a special slot in the surgeon's collection of tools for laser surgery.

Continuous laser as well as pulsed laser both can remove the tissue.

Advantages in the fabrication of different types of lasers and the art of laser medicines are leading to better criteria for matching lasers to specific medical treatment needs. Various types of lasers either pulsed or continuos are in use for such purposes e.g. CO_2 laser, Argon laser, Nd-YAG laser, dye lasers (tunable to different wavelengths) etc. Having the choice of variety of lasers, the physician picks the best laser for the job, selecting a desired pulse length as well as wavelength to control energy deposition.

When the tissue and its cells absorb the intense laser light, energy must be dissipated in some way; this release may take place in the form of heat, photodissociation, shock waves, chemicals reactants or fluorescence. Physicians use all these mechanisms to study cells and tissues for the diagnoses and treatment of diseases. Each of these varying effects also allows scientists to perform sub-cellular microsurgery.

Figure 9.8 shows an instrument known as “Biolaser". This instrument has wide applications in the

Figure 9.8 Biolaser.

biomedical field and has been used for (a) microsurgical procedures, for providing openings in cell walls with much greater accuracy than is possible with presently available micro-manipulators, (b) irradiating cell interiors to cause healing of internal structures, (c) causing the coagulation of individual blood corpuscles, and (d) eliminating tissue parts.

Here we discuss some specific examples of medical applications of lasers.

9.3.4 Lasers in Ophthalmology

The main uses of lasers in ophthalmology are:

(1) The treatment of retinal detachment,

() Diabetic retinopathy,

(3) Cataract surgery,

(4) Corneal sculpting,

(5) The treatment of corneal ulcers,

We discuss some of the above applications in below.

9.3.4.1 Treatment of Detached Retina

Figure 9.9 shows the structure of the human eye. Retina is the light sensitive layer at the back of the eye. Sometimes the retina can become torn and detached from the back of the eyeball. The damage can spread, and without treatment the entire retina can come loose from the back of the

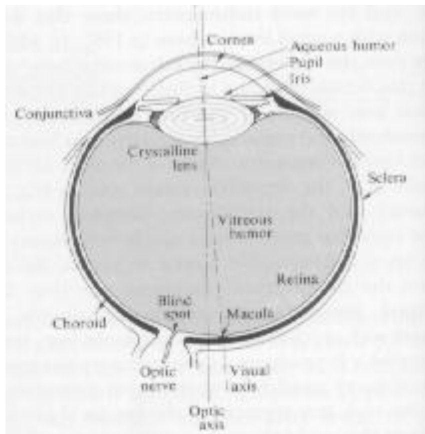


Figure 9.9 A diagram showing structure of a human eye.

eye causing blindness. A laser pulse focused on the retina can stop the detachment from the eyeball. An instrument that has the opportunity of working closer to the macula is shown in Figure 9.10.

The procedure of this treatment is explained below:

Focus a pulsed laser onto the retina so it causes a burn forming a scar tissue which welds the retina down to the back of the eye ball, so it cannot break free. Although laser may damage a small area of the retina but even then this technique represents a big advantage over earlier techniques which require risky open-eye surgery.

Figure 9.10 Laser photocoagulator.

9.3.4.2 Cataract Surgery

Cataracts occur when the natural lens of the eye becomes cloudy and obstructs vision. This usually occurs in older people. The standard treatment is to remove the natural lens material, leaving behind the back membrane of the lens, and to implant a plastic lens. In about a third of all cases, the natural membrane itself can become cloudy again obstructing vision.

Prior to 1980, the only treatment for secondary cataracts was the surgical opening of the posterior membrane, a maneuver that requires general anesthesia. But it has been showed that short pulsed infrared lasers could be focused on or near the opaque posterior membrane, causing it to be torn apart by a shock wave. Neodymium-YAG laser can issue pulses on the order of nanoseconds or picoseconds and operates at a wavelength of 1.06 μm. After laser treatment, a patient's vision is almost always instantly improved.

The procedure for this treatment is explained below.

A short intense pulse from a Neodymium-YAG laser can pass through the lens implanted and break the cloudy membrane as shown in Figure 9.11. The Nd-YAG pulsed laser is focused with a lens having very short focal length, so the laser light comes to a very tight focus exactly at the membrane surface. Milli joules of energy are delivered for a period that ranges from 10^{-12} to 10^{-9} second on a spot of 25 to 50 micron in diameter. The ophthalmologist fires a series of laser pulses to puncture the membrane. The beam passes through the outer cornea and the artificial lens which

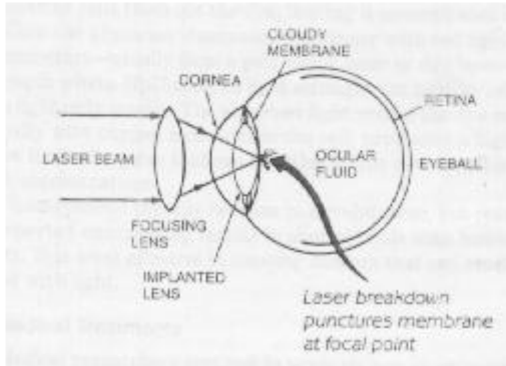


Figure 9.11 A tightly focussed laser pulses punctures a cloudy membrane.

are both transparent at 1.06 μm wavelength of Nd-YAG laser. Each pulse focused on the membrane produces a tiny spark that breaks the membrane. Behind the membrane, the laser light again spreads out over a larger area so it can not damage the retina.

The advantages of this surgery-using laser are:

(i) The operation can be done in an ophthalmologist's office and the membrane opens up instantaneously so no hospitalization is needed.

(ii) It avoids the risk of surgically opening the eye to cut the membrane.

9.3.4.3 Corneal Sculpting

Excimer laser (λ = 0.193 μm) can be used for corneal sculpting. The procedure is aimed to correcting major visual defects by reshaping the cornea. Using these procedures, visual defects such as near sightedness and farsightedness or astigmatism could be corrected.

9.3.5 Lasers in Dermatology

Dermatology is the medical field concerned with diseases of the skin. Laser has given dermatologists an improved new tool to treat some conditions. This technology was easily introduced in dermatological therapeutic because dermatologists have long been used ordinary light for skin treatments.

Lasers can be used to treat port-wine stains, tattoos, seborrhea, warty keratoses, basal cell carcinomas, warts, freckles, nevi, acne and various growths both benign and malignant. Generally the laser is used to remove some sort of growth or colored area in the skin and in general the results are comparable to conventional modes of therapy. Some of the applications are described in following:

9.3.5.1 Treatment of Port-Wine Stains

Dark-red birthmarks called port-wine stains often appear on the face or neck. Networks of abnormal blood vessels just under the surface of the skin cause port-wine stains as shown in Figure 9.12. Because port-wine stains are dispersed on the surface of the skin, conventional surgery can't remove them effectively. Laser can treat this condition.

Figure 9.12 Treatment of portwine stains with laser.

The procedure for this treatment is explained below:

Red birthmarks called port-wine stains absorb the Argon laser's beam, which can be blue or green depending on its wavelength. The laser beam heats the blood vessels causing burning and blistering of the skin and destroys hundreds of extra blood vessels that lie just beneath the skin's outer layer. Although it causes the blood vessels to close over a number of weeks, but as the burn heals the birthmarks bleach away.

The advantages of using laser in dermatology are its speed, superficial nature of the injury and interaction with the desired tissue only.

Although laser surgery is preferable to skin grafting and incision, it has its own drawback. The heat generated by the beam can sometimes spread to the parts of the skin other than the abnormal blood vessels and causes scarring or loss of pigment. However scarring can be prevented when the laser beam is delivered in short pulses rather than continuously. Short exposure less than one one-thousandth of a second to intense light would destroy the absorption site but produce little or no damage to the adjacent tissue. The reason is that it would take less time for the energy to be absorbed than it does for the heat to be transferred or dissipated to surrounding areas. Therefore the selective destructions of pigmented targets would require two conditions: preferential light absorption and sufficiently short light pulsation. This technique is called photothermolysis. It has also proved useful for removing tattoos.

9.3.6 Optical-Fiber based Surgical Laser Systems

Optical fiber based surgical laser systems have many applications in medical. Some of them are discussed below.

9.3.6.1 Laser-Tissue Interaction

The, wavelength of Nd-YAG laser is 1.06 μm. The light of this wavelength is not strongly absorbed by water with the advantage that it can penetrate more deeply into the tissue. But a problem can also be faced because its greater penetration depth can make it harder for the surgeon to control where the laser beam is going. To avoid it, the 1.06 μm wavelength can be transmitted through optical fibers. This lets optical fibers to deliver laser light within the body without the need for surgical incisions.

9.3.6.2 Shattering of the Stones

Delivery of light from a dye laser through optical fiber can shatter kidney stones and gall stones into pieces small enough to pass freely from the body. The laser is tuned to a wavelength strongly absorbed by the stones. Prospects for this treatment look very good.

9.3.6.3 Cancer Treatment

Lasers are an integral part of a new treatment for cancer called photodynamic therapy. Scientists noted that cancerous tissue concentrated some of the body's own pigments, particularly porphyrin, in the brownish-red pigment of blood. The treatment relies on a dye, called hematopophyrin derivative (HpD) which cancer cells retain much longer then healthy cells.

The procedure for this treatment is explained below:

A patient is given an injection of HpD, (a dye) as shown in Figure 9.14. This dye selectively concentrates in cancerous tissue 48 to 72 hours after it is injected while healthy cells flush out the

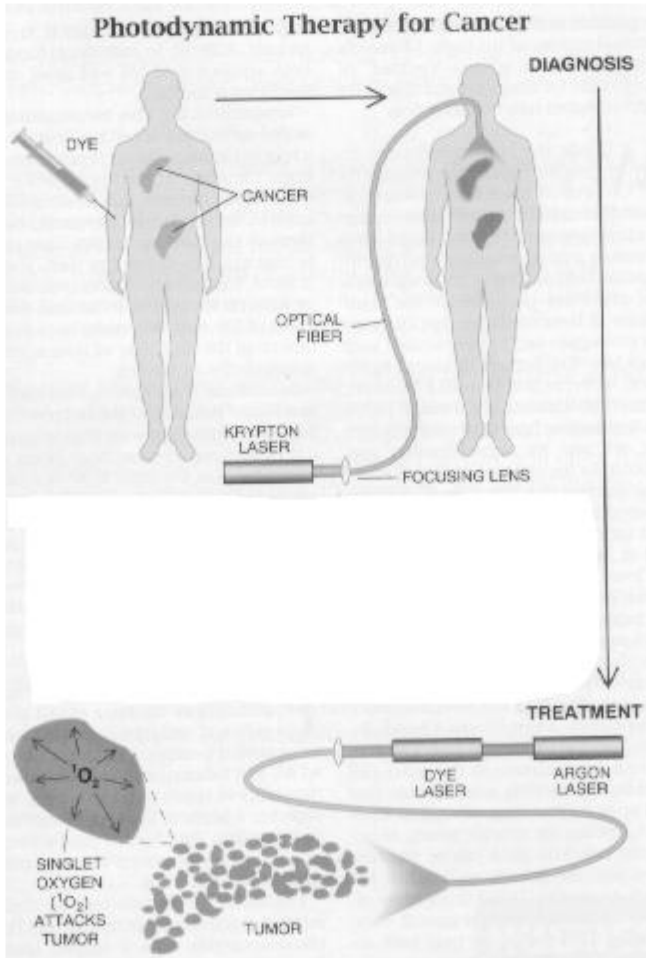


Figure 9.14 Photodynamic therapy for cancer.

dye. Blue violet light from the krypton laser transmitted through an optical fiber, causes the dye-laden growth to fluorescence so it can easily be observed and diagnosed. Then the physician illuminates the cancer with red light near 631 nm, a wavelength where HpD absorbs light strongly but healthy cells absorb light on weakly. For this purpose, usually gold vapor laser or argon laser pumped dye laser is delivered through optical fiber. Energy from the laser excites the dye molecules in the tumor. These molecules in turn pass the energy to molecular oxygen. The excited oxygen becomes highly reactive and cyto-toxic (cell killing) products are released in the tumor resulting in destruction of the tumor.

Problems

9.1 In the discussion of holographic principle we assume the hologram is a photographic positive. Discuss the case in which the hologram is a negative transparency.

9.2 Suppose the reconstructing beam is incident from the object side of a reflection hologram formed. Show that a real image of the object is formed.

9.3 How light get the information of the particle velocity from which it is scattered?

9.4 Calculate the Doppler shift in He:Ne laser light from a gas at temperature of 400 K.

9.5 Discuss in detail those properties of laser due to which it is accepted as a very useful "surgical tool". How these properties are exploited for medical surgery?

9.6 Discuss the basic principle, procedure and useful aspects of laser-tissue interaction.

9.7 Discuss the role of laser to treat the abnormal cloudy membrane behind the natural lens of the eye.

9.8 How laser light can be used for the treatment of port-wine stains. What are the advantages and drawbacks of this surgery?

9.9 Explain the procedure for diagnosing and treating photodynamic therapy.

Books for further reading

J. Wilson and J.F.B. Hawkes, *Optoelectronics: An Introduction*, (Prentice-Hall, India, 1996)

P. W. Milonni and J. H. Eberly, *Lasers*, (John Wiley & Sons, New York, 1991).

B. M. Watrasiewicz and M. J. Rudd, *Laser Doppler Measurements*, (Butterworths, London, 1976).

H. K. Koebner, *Laser in Medicine*, (John Wiley & Sons, New York, 1980).

J. Hecht, *Laser Guide Book*, 2nd ed. (McGraw-Hill, New York, 1992).

H. A. Elion, *Laser System and Applications*, (Pergamon Press, Oxford, 1967).