Ray Optics, Wave Optics and Imaging System Design 1

Fluorescence imaging has its roots in optics and molecular physics. The light-matter interaction is the key for understanding molecular phenomenon, be it Raman or fluorescence microscopy. For the simplistic reason of strong light-matter interaction, fluorescence imaging has gained prominence over other imaging techniques and has found wide applications in fields ranging from biophysics to optical engineering. The key advantage of this interaction is the ability to follow biological processes in their native state, thereby opening-up a large number of unanswered questions that have been asked for centuries. For example, protein migration and binding mechanisms were known to occur inside the cell, but were never observed directly. Moreover, many internal functioning of the cells is still a mystery. Only now, we are able to unravel these mysteries with the development of powerful super-resolution microscopes that allow visualization of molecule/protein functioning with near molecular-scale resolution. Molecules of interest are often tagged with fluorescent markers and followed while cell performs key molecular level tasks. The advantage of fluorescence microscopy is its ability to perform functional studies that are rare in other forms of imaging.

Visible light is a minute part of what we generally refer to as a electromagnetic (EM) spectrum. So, visible light is governed by the same principles that are applicable to other members (radio waves, X-rays, γ-rays etc.) of EM spectrum All the members of EM spectrum (that include visible light) propagate in the form of two mutually coupled vector fields i.e., electric field and magnetic field. This stems from the fact that, electric field effects gives rise to magnetic field and vice-versa. The very fact that, light has been manipulated for centuries before the existence of its vectorial description of light infers that light can be treated in a simplified manner as well for explaining day-to-day events. An approximate way of treating light (which is indeed a vectorial quantity) is referred as scalar theory (i.e., ray optics and wave optics). When light propagating through and around the objects whose dimensions are much greater than the dimension of wave (i.e., wavelength), the wave nature is not evident. In such a situation, one can describe light as a bunch of rays emanating from a source and propagating towards the object. During this, the light gets slowed down and diverted obeying a set of geometrical rules and postulates. It will be shown in the later part of the chapter that, ray optics is the limit of a more general theory called, wave optics which assumes that the dimensions of objects through which the light is passing are comparable to the wavelength of light. Specifically specialization such as, beam and Fourier optics has caved an important space within wave optics. Overall, the following picture (Fig. 1.1) emerges for describing the classical theory of light.

Light is also known to produce unusual effects that are classically elusive and unexplainable. So, the theory of light can be broadly divided into two streams: classical optics and quantum optics. Classical optics encompasses, ray optics, wave optics and electromagnetic optics, while quantum optics deals with the non-classical effects (such as, squeezing, entanglement, bunching etc.). It is important to realize that, within wave optics, specializations such as, Fourier optics and beam optics have gained prominence due to its wide applications in applied sciences. The optimal choice of the theory for an observed optical effect is the simplest theory that can satisfactorily explain all the aspects of the phenomena. The choice of a particular theory is often based on experience, so a-priori it is difficult to judge an appropriate theory. In this book, we will follow the order of simplicity and progress towards more complex theory as and when the need arise. In this chapter, we begin with the most simplistic theory of light i.e., ray optics or geometrical optics.

P.P. Mondal, A. Diaspro, *Fundamentals of Fluorescence Microscopy*, DOI 10.1007/978-94-007-7545-9_1,

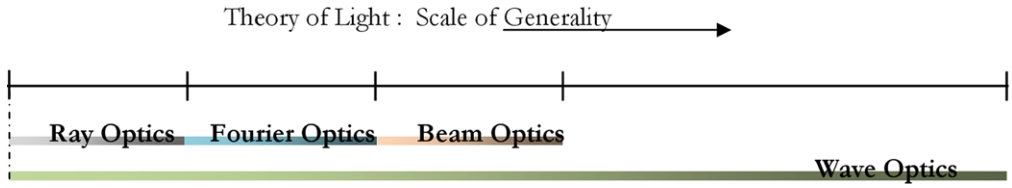


Fig. 1.1 Ray optics is the simplest theory of light and represents a sub-domain of wave optics. Fourier optics and beam optics also falls within the reach of wave optics

1.1 Ray Optics

Ray optics has been used for centuries starting from the times of Newton, Galileo and even before. This continues to impress even today and is an important tool for the development of future optical microscopes. Ray optics is by far the simplest theory of light and based on the optical phenomenon observed on day-to-day observations. Light is described by rays or simple straight lines. Ray optics in conjunction with mathematical rules of geometry can be used to describe simple optical phenomenon such as, reflection and refraction. This successfully explains the image formation using simple optical elements such as, lens, mirrors and its combinations. The image position can be obtained by simply tracing the ray path that obeys the governing postulates of ray optics. To start with, a bundle of rays can be constructed at the object and their paths can be followed. The accumulation of rays on the desired plane determines the output image. For example, multiple rays emanating from a point can be traced through an intermediate biconvex lens and followed on the other side. The density of rays at any point gives the position of the image. We put forward a set of postulates that governs the propagation of light through each optical element. These postulates are then applied to optical elements such as, lens, mirrors and graded-optical index materials to determine the image position and magnification.

Postulates of Ray Optics:

I. Light travels in the form of rays emanating from the source to the observation plane.
II. The time taken by light to travel a distance x through a medium of refractive index $n \geq 1$ is $t = \frac{x}{(c_0/n)}$, where, c_0 is the speed of light in air. The effective optical pathlength in the medium is 'nx'. In an inhomogeneous medium, the refractive index is a function of position '$\vec{r}$', so the optical pathlength is '$n(\vec{r})$' multiplied by the differential distance 'ds' i.e., $\int_{x_1}^{x_2} n(\vec{r})ds$ where, s is the path followed by light from point x_1 to x_2.
III. Fermat's Principle: Rays travelling between two points (x_1 and x_2) follow a path such that the time of travel is extremum (minimum or maximum) relative to any other neighboring paths i.e.,

$$\Delta\left(\int_{x_1}^{x_2} n(r)ds\right) = 0. \tag{1.1}$$

It is however, usually a minimum. A more practical implication would be, "Light rays travel along the path of least time". Whenever, minimum time is shared by more than one path, rays simultaneously follows all the paths. These postulate along with simple geometrical rules can explain most of the optical effects observed at macroscopic scale (scales much larger than wavelength of light). These include the effects involving, reflection, refraction and transmission at and through different optical mediums. For obtaining implications on an ideal optical system, we consider homogeneity of the medium through which the beam propagates. These are good approximations because most observations/experiments involve free-space propagation are fairly homogeneous. The results can be expanded and generalized for heterogeneous media bearing in mind that, the results for systems involving inhomogeneous medium are complex and difficult to interpret even for simple optical systems. This rules out exact solution and often one has to go for approximate solutions based on numerical analysis.

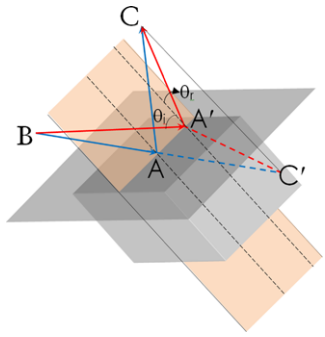
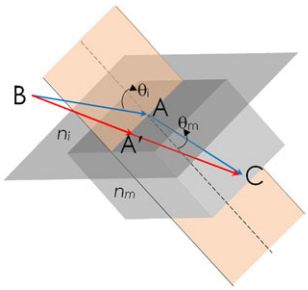


Fig. 1.2 Reflection from the surface of a planar interface



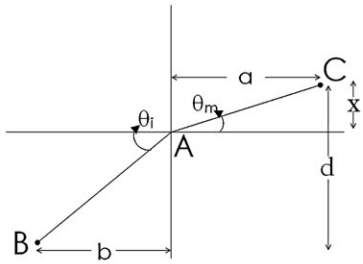


Fig. 1.3 Refraction at the interface of two medium with refractive index n_i and n_m respectively

1.2 Reflection, Refraction and Propagation of Light

In a homogeneous medium, one can safely assume that, light travels in straight line because the path of minimum time is also the path of minimum distance between two points according to Fermat's principle. As we move forward, we will realize that, laws of reflection and refraction are mere consequence of these postulates.

Law of Refraction *The angle of reflection is equal to the angle of incidence and both the rays lies in the plane of incidence.*

Consider the situation shown in Fig. 1.2, in which the light is incident on a plane of glass and gets reflected. Keeping postulates (Fermat's theorem) in mind, we ask the following question, "*Which path light follows that minimizes the travel time between point B and C*".

Rays starting from point B strikes the interface at a point A or A', and reaches the point C. The angles subtended with the normal (to the surface) before and after reflection are θ_i and θ_r respectively. A-priori, we do not know the relation between θ_i and θ_r. Since the ray remains all the time in a single medium, the path of least time is also the path of least distance according to Fermat's theorem. Our goal is to minimize the path $BA' + A'C$. Since, C' is the mirror image of C, so $BA' + A'C = BA' + A'C'$. Now the path, $BA' + A'C'$ is minimum when it is a straight line i.e., BAC'. This happens when A' coincides with A, forcing $\theta_i = \theta_r$.

Law of Refraction *The incident ray and the refracted ray are related by Snell's law i.e.,* $n_i \sin\theta_i = n_m \sin\theta_m$, *where,* θ_i *and* θ_m *are the angles in two medium of refractive index* n_i *and* n_m *respectively. In this case, we ask the question*, "What is the minimum time taken by the ray to travel from starting point B (in medium, n_i) to end point C (in medium, n_m)".

In this case, rays originating from point B hit the boundary and part of the ray gets reflected and other part refracted (see, Fig. 1.3). Here, we analyze only the refracted ray, keeping in mind that the reflected ray will behave according to the law of reflection. The refracted ray takes either the path $BA' + A'C$ or $BA + AC$. It is to be noted that, the ray $BA' + A'C$ travels shorter path in air (medium of small refractive index) and longer path in denser medium (large refractive index). This increases the total optical path. Instead, rays choose to travel longer in small refractive index medium and shorter in large refractive index medium. The total time taken to go from B to C via point A is,

$$t = \frac{BA}{(c_0/n_i)} + \frac{AC}{(c_0/n_m)} = \frac{1}{c_0}\left(n_i\sqrt{b^2 + (d-x)^2} + n_m\sqrt{x^2 + a^2}\right)$$

Now, we need to minimize this time with respect to floating variable x so,

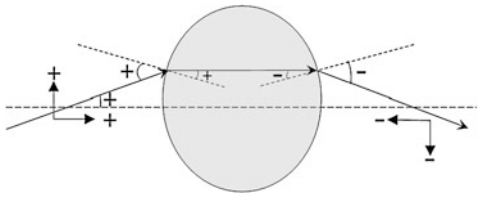


Fig. 1.4 Sign convention adopted in this book

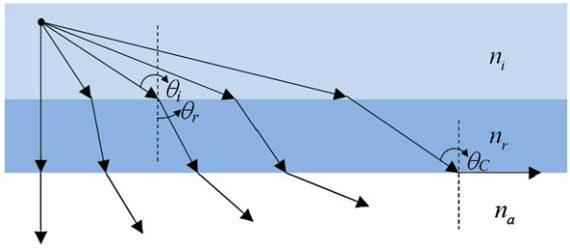


Fig. 1.5 Schematic diagram demonstrating light propagation from dense-to-rare and rare-to-dense medium leading to total internal reflection. Reflection at the interfaces is neglected

$$\partial t/\partial x = 0$$
$$\Rightarrow \quad \frac{2n_m x}{\sqrt{x^2+a^2}} - \frac{2n_i(d-x)}{\sqrt{b^2+(d-x)^2}} = 0$$

Noting that, $x/\sqrt{x^2+a^2} = \sin\theta_m$, and $(d-x)/\sqrt{b^2+(d-x)^2} = \sin\theta_i$ we get,

$$\begin{aligned} &\Rightarrow 2n_m \sin\theta_m - 2n_i \sin\theta_i = 0 \\ &\Rightarrow n_i \sin\theta_i = n_r \sin\theta_r \quad \text{(Snell's Law)}. \end{aligned} \tag{1.2}$$

Since, reflection and refraction are the consequence of Fermat's principle, one can analyze ray propagation through any optical element just by considering, reflection, refraction and straight line propagation of rays in a medium. In the succeeding section we will analyze few interesting optical elements frequently used in fluorescence microscopy. To do this we need to fix the sign convention as illustrated in Fig. 1.4. One assumption that we need to carry-on in rest of the book is, paraxial approximation. Rays that make small angles with the optical axis are paraxial rays such that, $\sin\theta \approx \theta$ and $\tan\theta \approx \theta$. Most of the expressions and calculations derived hereafter assumes this approximation and thereby facilitates simplification. In general, this is the essence of present day designing of optical components.

1.2.1 Planar Interface

We proceed to understand the laws of reflection and refraction for various optical elements such as lens and mirrors. Then we will consider complex optical elements such as, graded index (GRIN) lens. First, we need to understand reflection and refraction at planar and spherical interfaces, which forms the basis for understanding complex optical elements. In the case of a planar interface faced by the incident ray, the light bends according to the Snell's law for refraction. Figure 1.5 show the situation for two separate cases, (1) when rays propagate from low refractive index (RI) to high RI medium and, (2) rays travelling from high to low RI. For both the cases the ray propagation is shown in Fig. 1.5.

Obeying Fermat's principle of least optical path, the ray maximizes its travel in low refractive index and minimizes its travel in high RI medium. The Snell's law governs the ray path at an interface. At the first interface, Snell's law predicts the bending of ray away from the interface i.e., $\theta_i > \theta_r$ because $n_i < n_r$ and the ray is travelling towards high RI medium. For the subsequent boundary, the rays bend towards the interface because $n_r > n_a$ and the fact that, ray is traveling from high to low RI medium. If we continuously increase the incident angle, the refracted ray may just bend enough to get reflected. This condition is termed as total internal reflection and incident angle at which it happens is the critical angle of the conjugate system. Following Snell's law, the critical angle is, $\theta_C = \sin^{-1}(n_a/n_r)$.

It is worth to note that, there are several optical systems that inherit this effect (total internal reflection). Most common of them are, prism and optical fiber. Specifically for optical fiber, one can show, that the acceptance angle for incident ray is, $\theta_a = \sin^{-1}\sqrt{(n_{co}^2 - n_{cl}^2)}$, where, n_{co} and n_{cl} are respectively the refractive indexes of core and cladding of optical fiber. The reader is encouraged to prove this (Hint: Use Snell's law at each interface).

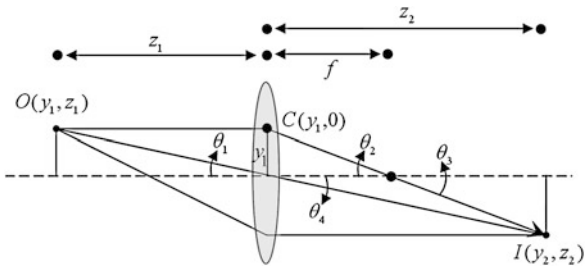


Fig. 1.6 Ray propagation through biconvex lens

1.2.2 Biconvex Lens

Most of the optical imaging system employ biconvex thin lens. The propagation of rays in a thin lens is shown in Fig. 1.6. An object of height y_1, placed at a distance z_1 casts a magnified image of height y_2 at a distance of z_2 from the center C of the lens of focal length f. In thin lens approximation, it is assumed that, the ray enters and exits the lens at the same height. From simple geometry, $\tan\theta_1 = \tan\theta_4, \Rightarrow \frac{y_1}{z_1} = \frac{y_2}{z_2}$. Similarly, $\tan\theta_2 = \tan\theta_3, \Rightarrow \frac{y_1}{f} = \frac{y_2}{z_2 - f}$. Equating, y_1/y_2 from both the expressions we get,

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f}. \tag{1.3}$$

The is the celebrated *imaging equation* for a single thin lens. The magnification of the single thin lens system is,

$$M = -\frac{y_2}{y_1} = -\frac{z_2}{z_1}. \tag{1.4}$$

1.2.3 Spherical Mirror

Based on the paraxial ray approximation, the spherical mirror can be approximated as elliptical mirror that has the unique property of focusing all the rays to a single point F. Figure 1.7(A, B, C) show the ray diagram for a spherical mirror. The rays are originating from the object $O'O$ (bottom and top) and creates an image TI at z_2, making an angle of 2θ.

From the triangle $\Delta O'MC$, $-\theta_2 = \theta_1 + \theta$ and ΔCMT gives, $-\theta_3 = \theta_2 + \theta$. Substitution yields,

$$\begin{aligned} \theta_1 + (-\theta_3) = 2\theta_2 \quad &\Rightarrow \quad \frac{y}{z_1} + \frac{y}{z_2} = \frac{y}{(-R/2)} \\ &\Rightarrow \quad \frac{1}{f} = \frac{1}{z_1} + \frac{1}{z_2} \end{aligned} \tag{1.5}$$

This is the imaging equation for a spherical mirror under paraxial approximation. As far as the magnification is concerned, we note that, from $\angle OSO' = \angle TSI$, indicating,

$$\frac{y_1}{z_1} = \frac{-y_2}{z_2} \quad \Rightarrow \quad y_2 = -\left(\frac{z_2}{z_1}\right) y_1 \tag{1.6}$$

So, the magnification factor is $(-z_2/z_1)$, where the negative sign indicates that the image is inverted.

1.2.4 Ray Equation and the Paraxial Approximation

This section deals with the important issue of trajectory tracing of light rays in an inhomogeneous medium of refractive index $n(\vec{r})$. Following Fermat's theorem,

$$T = \frac{1}{c} \int n(\vec{r}) ds$$

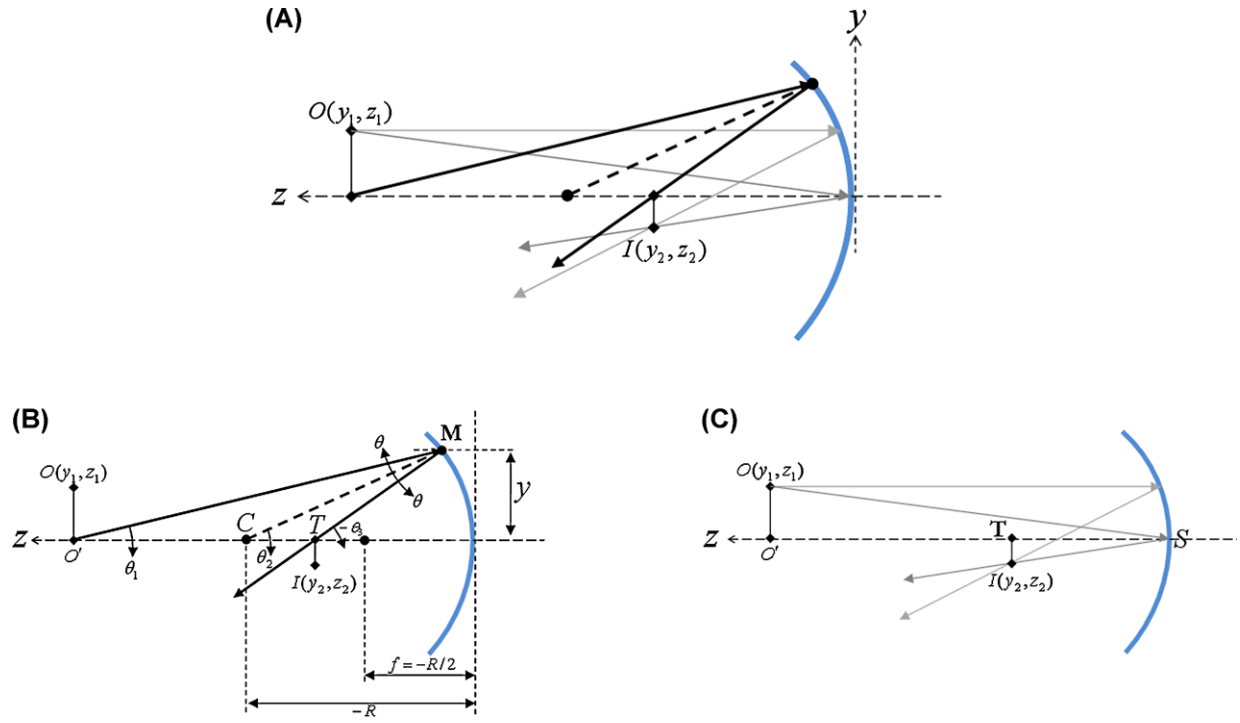


Fig. 1.7 (**A**) Image formation by the reflection of paraxial rays for a spherical mirror, (**B**) Rays originating from the bottom of the object O', (**C**) Rays originating from the top of the object O

and,

$$T + \Delta T = \frac{1}{c}\int \left[n(\vec{r})ds + \Delta(nds)\right]$$

Now, consider varying the path of the light ray from $\vec{r}(s)$ to $\vec{r}(s) + \Delta\vec{r}(s)$. The first order variation in time is given by,

$$\Delta T = \frac{1}{c}\int \Delta(n\, ds) = \frac{1}{c}\int [\Delta n\, ds + n\, \Delta ds].$$

Using the Calculus of Variation and further simplification gives,[1]

$$\Delta T = \frac{1}{c}\int \left[\nabla n - \frac{d}{ds}\left(n\frac{d\Delta r}{ds}\right)\right]\Delta r ds. \tag{1.7}$$

For minimum time, $\Delta T = 0$ producing,

$$\Delta T = 0 \quad \Rightarrow \quad \frac{d}{ds}\left(n\frac{dr}{ds}\right) = \nabla n. \tag{1.8}$$

This is the celebrated *ray equation*. In paraxial ray approximation, $ds \approx dz$ and noting that, $\vec{r} = (x, y, z)$, we get ray equation for x and y,

$$\frac{d}{dz}\left(n\frac{dx}{dz}\right) = \frac{dn}{dx} \quad \text{and} \quad \frac{d}{dz}\left(n\frac{dy}{dz}\right) = \frac{dn}{dy}. \tag{1.9}$$

[1]The calculation is outside the scope of this book. For detail refer to, R. Weinstock, Calculus of variation, Dover, 1974. Alternate approach can be found in the Book: B.E.A. Saleh amd M.C. Teich, Fundamentals of Photonics, John Wiley and Sons, 2nd Edition 2007.

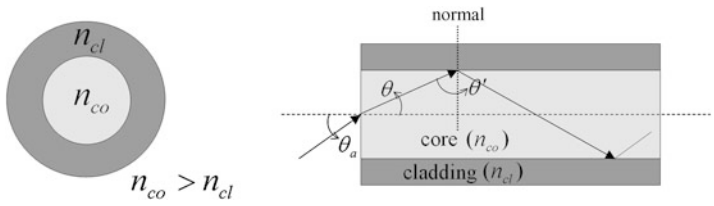


Fig. 1.8 Light propagation in an optical fiber guided by total internal reflection mechanism

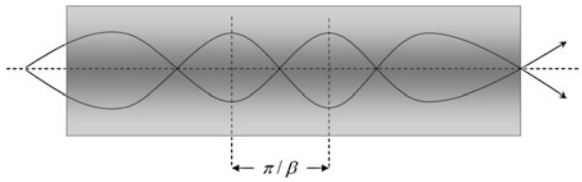


Fig. 1.9 Light trajectory in a GRIN material

1.2.5 Graded Index Lens

As the name suggests, the refractive index (RI) of the material varies with the position $\vec{r}$ as we move away from the optical axis. The optical elements made out of materials with varying RI have become very important in recent time. Numerous applications are found ranging from fiber optics to biomedical engineering. For example, consider a graded-index slab in which refractive index varies along x-direction and is uniform along y-axis. The corresponding paraxial ray equation is,

$$\frac{d}{dz}\left(n(x)\frac{dx}{dz}\right) = \frac{d}{dx}n(x) \tag{1.10}$$

Simplification gives the following differential equation,

$$\frac{d^2x}{dz^2} - \frac{1}{n(x)}\frac{dn(x)}{dx} = 0 \tag{1.11}$$

Imposing appropriate boundary conditions and with the knowledge of $n(x)$, one can solve the ray trajectory $x(z)$ inside the medium.

1.2.6 Light Guides and Graded Index Optical Fibers

The ideal way to guide a light through long distances is to make use of *total internal reflection* effect. This is precisely the technique used in the present day telecommunication employing optical fibers—*a nearly perfect light guide*. The fiber consists of inner solid thread like material termed as *core* and a rolled-up sheet called *cladding*. Both of them have slightly mis-matched RI sharing a common interface for enabling total internal reflection as shown in Fig. 1.8. The refractive index of core material is slightly more than that of cladding i.e., $n_{co} > n_{cl}$. From Fig. 1.8, $\theta' = 90^\circ - \theta = \sin^{-1}(n_{cl}/n_{co})$ for total internal reflection to occur. So, the acceptance angle that defines the transmission through the fiber without undergoing refraction at the core-cladding boundary, is given by,

$$n_a \sin\theta_a = n_{co} \sin\theta$$

Approximating, the refractive index of air by unity ($n_a \approx 1.0$), gives,

$$\theta_a = \sin^{-1}(n_{co}\sin\theta) = \sin^{-1}\left(\sqrt{n_{co}^2 - n_{cl}^2}\right). \tag{1.12}$$

The parameter $NA_{fiber} = \sin\theta_a = \sqrt{n_{co}^2 - n_{cl}^2}$ is known as the numerical aperture of the fiber.

Next we extend the theory of light transmission to include graded refractive index material as the material for optical fibers instead of step like design of core and cladding. Let us consider light propagation through a graded-index fiber (with

loss β) as shown in Fig. 1.9, in which the refractive index varies as function of x and y,

$$n = n_0\sqrt{\left[1 - \beta^2\left(x^2 + y^2\right)\right]}$$

Substituting this in the individual paraxial ray equation and assuming that, $\beta^2(x^2 + y^2) \ll 1$, we have $n = n_0(1 - \frac{1}{2}\beta^2 x^2)$ and $n = n_0(1 - \frac{1}{2}\beta^2 y^2)$ for x and y respectively after expansion and neglecting higher orders. Because the fractional change in the refractive index is very small, $(1/n)\partial n/\partial y \approx -\beta^2 y$ and $(1/n)\partial n/\partial x \approx -\beta^2 x$, giving,

$$\frac{d^2x}{dz^2} \approx -\beta^2 x \quad \text{and} \quad \frac{d^2y}{dz^2} \approx -\beta^2 y$$

Assuming, initial position (x_0, y_0) and angles ($\theta_x = dx/dz$ and $\theta_y = dy/dz$) at $z = 0$, we get,

$$x(z) = (\theta_x/\beta)\sin(\alpha z), \tag{1.13}$$

$$y(z) = y_0\cos(\beta z)(\theta_y/\beta)\sin(\beta z). \tag{1.14}$$

This demonstrates that, both x and y are therefore harmonic function of z with a period of $2\pi/\beta$ (see Fig. 1.9). The rays follow a helical trajectory inside the fiber. So, the GRIN material based optical fibers serves as a light guide for long distance transportation of light.

1.3 Matrix Formulation of Ray Optics

Although formulations based on the basic postulates completely describes the effect of optical element on light propagation, but a trivial and user-friendly formulation is felt necessary for designing complex optical systems such as, fluorescence microscopy, holography systems, optical telescope and others. This gives birth to Matrix formulation of ray optics.

In matrix formulation, a ray is described by its position and angle (with respect to optical axis). In a complex system, both the angle and position changes as the ray travels through different optical elements. The input and output after each optical element (including free-space propagation) is defined by column vector. The optical system (such as, lens and mirror) is defined by a 2×2 matrix. In matrix formulation, paraxial approximation is assumed. Since an individual optical element is defined by a matrix, so any complex optical system is equivalent to a single 2×2 matrix which is a simple product of all the individual component matrix. Consider a ray travelling through the optical system starting at a plane at $z = z_1$ and observed at the output plane at $z = z_2$. In general, the ray is defined by its position (say, y) and the angle (θ) i.e., the vector $\left[\begin{smallmatrix} y \\ \theta \end{smallmatrix}\right]$. This simplification is possible due to paraxial ray approximation so that one can write, $\sin\theta \approx \theta$. Let the input and output plane be denoted as, $\left[\begin{smallmatrix} y_1 \\ \theta_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} y_2 \\ \theta_2 \end{smallmatrix}\right]$. The optical system is characterized by a matrix $\left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ that acts on the input ray to produce the output ray coordinates i.e.,

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix}. \tag{1.15}$$

The matrix is termed as *ray-transfer matrix*. In the subsequent section, we will determine ray transfer matrix for some optical components that are frequently used in fluorescence microscopy and imaging.

1.3.1 Thin Biconcave and Biconvex Lens

Consider the case of thin biconvex lens in which the rays originate from the bottom of the object (point O) and terminates at I on the optical axis as shown in Fig. 1.10. The input and output are related by,

$$\frac{1}{z_1} + \frac{1}{z_2} = \frac{1}{f}$$

Multiplying throughout by y gives,

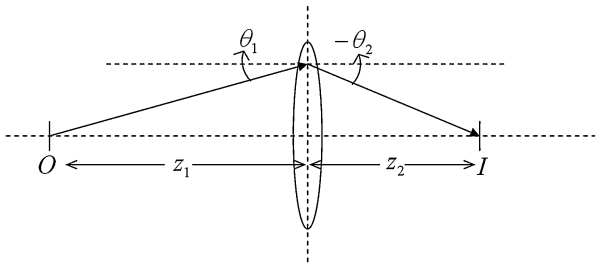


Fig. 1.10 Ray propagation through a thin lens

$$\frac{y}{z_1} + \frac{y}{z_2} = \frac{y}{f}$$

$$\theta_1 - \theta_2 = \frac{y_1}{f} \tag{1.16}$$

Since the lens is sufficiently thin so the height of the impinging and emerging rays are almost equal. Hence, it is safer to assume that, $y_1 = y_2$. Corresponding matrix equation is given by,

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (-1/f) & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix}. \tag{1.17}$$

By convention, the focal length f is negative for concave lens and positive for convex lens. The ray transform matrix determines the effect of biconvex lens on the input ray coordinates. Note that, these coordinates define the position of image formation in a single lens based imaging system.

1.3.2 Spherical Mirror

As defined in the previous section, the input and output coordinates indicating the ray position are given by, $\theta_2 = \theta_1 - \frac{y_1}{f}$ and $y_2 = -\frac{z_2}{z_1} y_1$ (see Sect. 1.2.3). The same can be rewritten in the ray-matrix formulation as,

$$\begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -z_2/z_1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix}. \tag{1.18}$$

Note that, for a spherical mirror, the ray matrix also defines the magnification factor of the system in addition to the coordinates of the image formation.

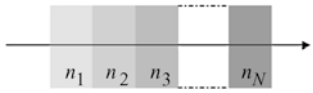
As an example, the readers are encouraged to compute the ray coordinates $r_i = (x_i, v_i)$ (using MATLAB) at an arbitrary distance $z = d_i$ behind a lens of focal length f when the input ray coordinates $r_0 = (x_0, v_0)$ for a ray starting from an object located at a distance d_0 in front of the lens. Note that, $r_0 = (0, 1)$, $d_0 = 150$ mm, $f = 100$ mm, $z_i = 300$ mm. Determine the system matrix. Find out the magnification of the system for $(1, 1)$ as the input ray-coordinate.

1.3.3 Cascade Optical Systems

Practically optical systems are complex and consists of a cascade of many optical elements such as lens, mirrors, optical window etc. Each of these optical elements is characterized by a unique ray-matrix M, so a cascade of N optical elements is simply equivalent to a resultant system ray-matrix obtained by multiplying individual matrix in a reverse order i.e., starting

Fig. 1.11 A cascade of optical system, in which each optical element is represented by a Matrix M_i

Fig. 1.12 Ray through a set of parallel transparent plates



from the image plane to the object plane i.e.,

$$M = M_N \times M_{N-1} \times \cdots \times M_2 \times M_1 \tag{1.19}$$

The optical components are indicated by black box in Fig. 1.11.

An interesting cascade system consists of a set of N parallel transparent plates of increasing refractive index $n_1 > n_2 > \cdots > n_N$ at normal as shown in Fig. 1.12. The readers are encouraged to determine that the ray-transfer matrix of the cascade system is $M = \begin{vmatrix} 1 & d\sum_{i=1}^{N} 1/n_i \\ 0 & 1 \end{vmatrix}$. Note that, the light is incident from left at normal incidence. It would be worthwhile to workout the ray-transfer matrix for incidence at an angle θ_i.

1.4 Wave Optics

Centuries ago, it was conceived that light propagates in the form of waves. In the previous sections, we witnessed the fact that most of the imaging parameters such as, image formation coordinates and magnification can be determined using ray optics. So, we never felt the necessity to formulate another theory unless some of the striking optical effects such as interference and diffraction are observed. These effects are outside the bound of ray optics and so the necessity to formulate a new theory was felt to explain these effects. The new theory is based on wave nature of light. Essentially, this theory is inspired by effects observed in water (water waves) and air (sound waves). We all know, what happens when waves are created at two nearby points on the bank of a still water body (pond or lake). Wherever two wave crests meet the amplitude of the wave elevates than the amplitude of individual wave, and the amplitude decreases when trough of one wave meet the crest of the other wave.

Incidentally, we will see (as we progress in the chapter) that, interference and diffraction effects are shown by all the members of electromagnetic spectrum. So, the wave theory is applicable to the entire spectrum and encompasses the ray optics. Ray optics is capable of defining most of the optical phenomena in which the wavelength of the light is sufficiently small as compared to the objects through or around which it is propagating. The starting point for wave optics is the description of the field which is a scalar function called wavefunction and all the physically measurable quantity (such as, intensity and power) can be calculated from the wavefunction. It is to be noted that, that scalar wavefunction is not a physically measurable quantity, it is the modulus of this function that is measurable. Similar to ray optics, we start with the postulates that govern wave optics and its effects. These postulates forms the basis for understanding several effects that falls outside the confines of ray optics including, diffraction and interference dependent optical effects. However, wave optics is not the complete story and we will need a much finer theory for a complete understanding of light. Soon we will realize the limitation of wave optics and we will be forced to formulate a sophisticated theory for describing vectorial properties of light. This is called electromagnetic theory of light which is the subject matter of the next chapter. Wave theory which is a scalar theory will fall short in this regard, although a few concepts of wave theory can be carried forward.

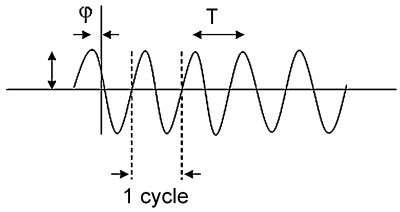


Fig. 1.13 Time representation of a monochromatic plane wave for fixed position $\vec{r}$ as the time advances

1.4.1 Postulates

I. The Wave Assumption Light propagates in the form of waves (very much like water and sound waves). The wave is fully determined by wavefunction that satisfies wave equation:

$$\frac{\partial^2}{\partial t^2}\psi(\vec{r},t) - c^2\nabla^2\psi(\vec{r},t) = 0 \tag{1.20}$$

where, $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$. Incase of multiple scalar waves satisfying the wave equation, all of them are equally a possible wave, and even their sum is a optical wave because the principle of superposition applies.

II. The Medium Effect In vacuum, light travels with a constant speed of $c_0 = 3 \times 10^8$ m/s. A medium is characterized only by its refractive index, and the speed of light (c) propagating through the medium is scaled by the refractive index i.e., $c = \frac{c_0}{n}$.

At the interface of two media, the wavefunction changes in a way that depend only on their refractive index and maintain continuity of the wavefunction at the interface. Moreover, the refractive index and hence the speed becomes position dependent for inhomogeneous medium, i.e., $n(r), c(r)$.

1.4.2 Monochromatic Wave

The simplest wave that one can think of is monochromatic wave that has single wavelength and is a harmonic function of time. Mathematically, a monochromatic wave can be expressed as,

$$\begin{aligned}\psi(\vec{r},t) &= \psi_0(\vec{r})\cos\big(\omega t + \varphi(\vec{r})\big) \\ &= \psi_0(\vec{r})\,\mathrm{Re}\big(e^{-i\omega t}e^{-i\varphi(\vec{r})}\big)\end{aligned} \tag{1.21}$$

where, ψ_0 and φ are respectively the amplitude and phase of the optical wave. $\omega = 2\pi\nu$ is the angular frequency and ν being the frequency (cycles/s or Hz) of the wave. The time period (time between two consecutive crests or troughs) is, $T = 1/\nu = 2\pi/\omega$ (in seconds). A sample wave demonstrating these parameters is shown in Fig. 1.13.

The scalar wavefunction can be rewritten as,

$$\psi(\vec{r},t) = \mathrm{Re}\,\Psi(\vec{r},t) = \big[\Psi(\vec{r},t) + \Psi^*(\vec{r},t)\big]/2 \tag{1.22}$$

where, $\Psi(\vec{r},t) = (\psi_0(\vec{r})e^{-i\varphi(\vec{r})})e^{-i\omega t} = \Psi(\vec{r})e^{-i\omega t}$ is the complex wavefunction that describes the wave completely, and $*$ represents complex conjugate. The reason to switch on to the exponential form is the mathematical simplicity and because this form clearly specifies the space and the time dependent part. Like $\psi(\vec{r},t)$, complex wavefunction $\Psi(\vec{r},t)$ also satisfies the wave equation and same boundary conditions. Therefore, the wave equation corresponding to complex wavefunction is given by,

$$\frac{\partial^2}{\partial t^2}\Psi(\vec{r},t) - c^2\nabla^2\Psi(\vec{r},t) = 0. \tag{1.23}$$

1.5 Helmholtz Equation and Its Consequences

It will be realized in the next section that, spatial component of light has a prominent role in fluorescence microscopy. Usually, it is ideal to have uniform or Gaussian profile of the light beam. Here, we focus on the spatial part of the complex wavefunction. Substituting, $\Psi(\vec{r},t) = \underbrace{[\Psi_0(\vec{r})e^{-i\varphi(\vec{r})}]}_{\Psi(\vec{r})} e^{i\omega t}$ in the wave equation gives,

$$\nabla^2\Psi(\vec{r}) + (\omega/c)^2\Psi(\vec{r}) = 0 \tag{1.24}$$

where, the wave number is defined as, $k = \omega/c$. This is called *Helmholtz equation*. This also enables the determination of physically measurable quantities such as, intensity and power in general terms. The optical intensity is defined as, $I(r) = |\Psi(\vec{r})^2|$, where, $\Psi(\vec{r})$ is the complex amplitude. So, there is no time dependence in the intensity of the monochromatic wave. One also need to better understand the phase and the wavefront associated with it. In brief, the wavefronts are surfaces of equal phase and so, $\varphi(r) = C_1$. Accordingly, the wavefront can be chosen such that, $\Psi(\vec{r}) = 2\pi n$, where n is an integer. In the next section, we will expand on the functional form of the function Ψ, and find out simple solutions of the Helmholtz equation.

1.5.1 The Plane Wave

The simplest mathematical way to solve the above differential equation is by well-known trial method. The first step is to disentangle the x, y and z components by assuming, solution of the form $\Psi = \Psi_1(x)\Psi_2(y)\Psi_3(z)$. Substitution in the Helmholtz equation produces three similar equation and the x-component is given by,

$$\frac{\partial^2}{\partial x^2}\Psi_1 + k_x^2\Psi_1 = 0. \tag{1.25}$$

Similar equation exists for y and z components with the condition, $k^2 = k_1^2 + k_2^2 + k_3^2$. In trial method, we guess a solution of the form, $\Psi_1(x) = e^{mx}$. Substitution in the Helmholtz equation produces $m = \pm ik$, so the general solution is a linear combination of both the roots i.e., $\Psi_1(x) = A_1e^{-ikx} + A_2e^{ikx}$. By analogy, other solutions are of the form, $\Psi_2(y) = B_1e^{-iky} + B_2e^{iky}$ and $\Psi_3(z) = C_1e^{-ikz} + C_2e^{ikz}$. The complete solution is, $\Psi(\vec{r}) = [A_1e^{-ikx} + A_2e^{ikx}][B_1e^{-iky} + B_2e^{iky}][C_1e^{-ikz} + C_2e^{ikz}]$. Considering only the waves propagating in the positive direction, the solution becomes,

$$\Psi(\vec{r}) = De^{-i(k_xx+k_yy+k_zz)} = De^{-i\vec{k}.\vec{r}} \tag{1.26}$$

where, $D = A_1B_1C_1$ is a complex constant (also known as the complex envelop). k or more appropriately $|\vec{k}|$ is the wave number of the resultant plane wave Ψ.

It must be noted that, the phase of the wave is, $-\vec{k}.\vec{r} + \arg D$. The wavefront (the surfaces of constant phase) of the resultant wave must obey, $-\vec{k}.\vec{r} + \arg D = 2\pi q$, with q an integer. This ($-\vec{k}.\vec{r} = 2\pi q - \arg(D)$) represents parallel planes $\vec{r} = (x, y, z)$ perpendicular to the wave vector, $\vec{k} = (k_x, k_y, k_z)$. Hence, the resultant wave is known as plane wave. It may be noted that, consecutive planes are separated by a distance, $2\pi/k$, which is the wavelength λ of the resultant plane wave. The intensity of the plane wave is, $I(r) = |D|^2$ everywhere in space and time, so it carries infinite energy. This is clearly an idealization and does not hold in reality.

Consider that the wave is propagating along $+z$-axis, then $\Psi(\vec{r}) = |D|e^{-ikz}$. The real wavefunction is,

$$\begin{aligned}
\psi(\vec{r},t) &= \left(\Psi(\vec{r})\, e^{i\omega t} + \Psi^*(\vec{r})e^{-i\omega t}\right)/2 \\
&= |D|\cos(\omega t - kz + \arg D) \\
&= |D|\cos\left[\omega\left(t - \frac{z}{c}\right) + \arg(D)\right] \\
&= |D|\cos\left[2\pi\left(\frac{t}{(1/\nu)} - \frac{z}{(c/\nu)}\right) + \arg(D)\right]
\end{aligned} \tag{1.27}$$

So, it is clear that, the wavefunction is periodic in space with period $c/\nu = \lambda$ and periodic in time with period $1/\nu$. The phase of the wavefunction $\omega(t - z/c) + \arg(D)$ varies with space and time as a function of $(t - z/c)$; c is called the phase velocity of the wave. The representation shown in Fig. 1.14 emerge from the description of plane waves in this section.

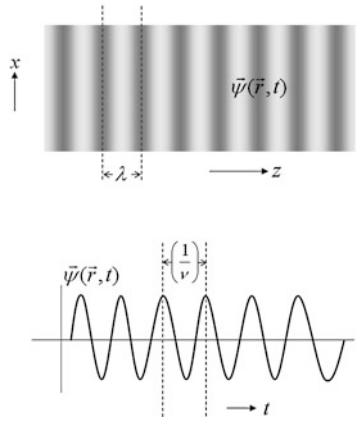


Fig. 1.14 Spatial and temporal periodic propagation of a plane wave

The other solution of the above equation in spherical coordinates is spherical wave, $\Psi(r) = \frac{1}{r}De^{-ikr}$ for which the wavefronts are concentric spheres separated by, $2\pi/k$. We leave it to the reader to derive and verify this. Also, the readers are encouraged to determine that spherical wave close to z-axis is $\Psi(r) \approx \frac{A_0}{z}e^{-ikz}e^{-ik(x^2+y^2)/2z}$. Further, it is interesting to note that, spherical wave behaves like plane wave at large distances.

1.5.2 Paraxial Plane Waves and Helmholtz Equation

Those waves for which the wavefront normals are paraxial (parallel to the optical axis), are termed as paraxial waves. Simplest way to obtain a paraxial wave is to start with plane wave propagating along the z-axis i.e., $\Psi(r) = D(r)e^{-ikz}$.

Paraxial Helmholtz equation is obtained by substituting the plane wave in the Helmholtz equation,

$$\nabla_T^2\left(De^{-ikz}\right) + \nabla_z^2\left(De^{-ikz}\right) + k^2De^{-ikz} = 0$$

where, $\nabla_T^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the transverse Laplacian operator.

Rearrangement and expansion gives,

$$\begin{aligned} \nabla_T^2\left(De^{-ikz}\right) + \frac{\partial^2 D}{\partial z^2}e^{-ikz} - 2ik\frac{\partial D}{\partial z}e^{-ikz} = 0 \\ \nabla_T^2\left(De^{-ikz}\right) + \left[\frac{\partial^2 D}{\partial z^2} - 2ik\frac{\partial D}{\partial z} + k^2D\right]e^{-ikz} - k^2De^{-ikz} = 0. \end{aligned} \tag{1.28}$$

It is assumed that, the variation of the envelop and all its derivatives is less than as compared to the wavelength so that plane wave nature stays. Accordingly, $\Delta D \ll D$ for $\Delta z = \lambda$. Since, $\Delta D = \frac{\partial D}{\partial z}\Delta z = (\frac{\partial D}{\partial z})\lambda$, so, $\frac{\Delta D}{\lambda} = \frac{\partial D}{\partial z}, \Rightarrow \frac{\partial D}{\partial z} \ll \frac{D}{\lambda} = \frac{Dk}{2\pi}$, implying,

$$\frac{\partial D}{\partial z} \ll Dk \tag{1.29}$$

$$\Rightarrow \quad \frac{\partial^2 D}{\partial z^2} \ll k\frac{\partial D}{\partial z} \ll k^2D \tag{1.30}$$

Using paraxial plane wave approximation and neglecting the second order, and by using (1.30) we get,

$$\nabla_T^2\left(De^{-ikz}\right) + \left[-2ik\frac{\partial D}{\partial z}e^{-ikz} + k^2De^{-ikz}\right] - k^2De^{-ikz} = 0$$

Simplification and cancellation gives,

$$\nabla_T^2 D - 2ik\frac{\partial D}{\partial z} = 0 \tag{1.31}$$

This is the paraxial Helmholtz equation that deals with slowly varying envelop approximation of Helmholtz equation. This gives rise to the most important beam shapes such as, Gaussian and Bessel Beam, which are extensively used in fluorescence microscopy.

1.6 Solution of Paraxial Plane Wave (Beam Optics)

1.6.1 Gaussian Beam

Let us go back to paraxial plane waves which is based on the approximation that, wavefront normals are paraxial i.e., makes small angles with the z-axis. Here, we investigate some of the important solutions of paraxial Helmholtz equation from application point-of-view. Considering a paraxial plane wave travelling along the z-axis e^{-ikz} with complex envelop $D(x, y)$, has a complex amplitude given by,

$$\Psi(x, y) = D(x, y)e^{-ikz} \tag{1.32}$$

The envelop $D(x, y)$ is assumed to be approximately constant within a distance of wavelength λ, so that the wave remains paraxial.

Substituting into the Helmholtz equation we see that, the envelop function $D(x, y)$ satisfies paraxial Helmholtz equation i.e.,

$$\nabla_T^2 D(x, y) - 2ik\frac{\partial D(x, y)}{\partial z} = 0 \tag{1.33}$$

One can verify by direct substitution that, one of the solutions of the paraxial Helmholtz equation is the *Gaussian beam*,

$$D(x, y) = \frac{C_0}{q(z)}e^{-ik\frac{x^2+y^2}{2q(z)}} \tag{1.34}$$

where, $\frac{1}{q(z)} = \frac{1}{R(z)} - i\frac{\lambda}{\pi W^2(z)}$. Here, $R(z)$ and $W(z)$ are the measure of wavefront radius of curvature and beam width respectively.

Simplification show that, the complex amplitude can be written in an expanded form as,

$$\Psi(x, y) = C_0\frac{W_0}{W(z)}e^{-[\sqrt{x^2+y^2}/W(z)]}e^{-i[kz-k\sqrt{x^2+y^2}/2R(z)+\chi(z)]} \tag{1.35}$$

where, $W(z) = W_0\sqrt{[1 + (z/z_0)^2]}$, $W_0 = \sqrt{\frac{z_0\lambda}{\pi}}$, $\chi(z) = \tan^{-1}(z/z_0)$, $R(z) = z(1 + (z_0/z)^2)$ and C_0 is a complex constant.

Properties of Gaussian Beam

Here, we explain some of the important characteristics of Gaussian beam. In this section, we will explore some of the key properties of Gaussian beam that are often used in fluorescence microscopy and imaging.

Intensity: The optical intensity of the Gaussian beam is simply the modulus of the complex amplitude i.e.,

$$I(r, z) = \left|\Psi(x, y)\right|^2 = |C_0|^2\left(W_0/W(z)\right)^2 e^{-2r^2/W^2(z)}, \tag{1.36}$$

where, $r = \sqrt{x^2 + y^2}$ is the radial coordinate.

Phase: The phase of the Gaussian Beam is given by,

$$\varphi(x, y, z) = kz - k\frac{\sqrt{x^2 + y^2}}{2R(z)} + \chi(z) \tag{1.37}$$

Considering $r = \sqrt{x^2 + y^2} = 0$ (on the beam axis), the phase becomes,

$$\varphi(r=0,z) = kz + \chi(z) \tag{1.38}$$

where, the first term (kz) is the phase of the plane wave and the second term ($\chi(z)$) represents phase retardation of the wave. This ranges from $(-\pi/2, +\pi/2)$ for $z = (-\infty, +\infty)$. This is the excess delay faced by the Gaussian wave with respect to the plane wave. The total accumulation of the phase contributed by this retardation process is half cycle i.e., π (from the far-field to the far-field on the other side of the focus). This extra phase is known as *Gouy phase*.

Beam Width and Depth of Focus: One can easily show that, most of the power is carried within a circle of radius $W(z)$ so, we regard this as the beam width. The beam-width of a Gaussian wave along the optical axis z is given by,

$$W(z) = W_0\sqrt{\left[1 + \left(\frac{z}{z_0}\right)^2\right]} \tag{1.39}$$

for which the minimum W_0 is attained at $z = 0$. In fluorescence microscopy, the waist diameter $2W_0$ is commonly termed as *spot size*. The beam-width increases monotonically with z.

The depth-of-focus is defined as the length over which the beam-width is within a factor of $\sqrt{2}$ times W_0 which is also equal to twice that of Rayleigh range (z_0) i.e.,

$$F_D = 2z_0 = \frac{2\pi W_0^2}{\lambda} = \frac{2A_0}{\lambda} \tag{1.40}$$

where, $A_0 = \pi W_0^2$ is the area of the beam-waist at $z = 0$. So, a beam focused to a small spot size has a short depth-of-focus. This has far-reaching implications for large-penetration depth-imaging in fluorescence microscopy.

Beam Divergence: It is interesting to know, the waist size at large z i.e., $z \gg z_0$. Imposing this condition results in the following relation at large distances,

$$W_L = \lim_{z\to\infty} W_0\left[1 + \frac{z}{z_0}\right] \approx \left(\frac{z}{z_0}\right)W_0 \tag{1.41}$$

Recognizing that, $\theta_D = \frac{W_0}{z_0} = \frac{\lambda}{\pi W_0}$, the beam diverges as a cone of half-angle θ_D. So, the divergence is directly proportional to the wavelength and inversely proportional to the beam-width at focus $z = 0$.

1.6.2 Bessel Beam

The other solution is the non-diffracting beam called Bessel beam. Bessel beams are waves whose wavefronts are planar but intensity distribution is non-uniform in the transverse plane i.e.,

$$\Psi(x,y) = D(x,y)e^{-ik_L z} \tag{1.42}$$

where, $k^2 = k_L^2 + k_T^2$. K_L and k_T are respectively the longitudinal and transverse wavevectors.

For this to satisfy Helmholtz equation, the quantity $D(x,y)$ must satisfy,

$$\nabla_T^2 D(x,y) + k_T^2 D(x,y) = 0$$

The readers are encourage to verify that, the solution in polar coordinate system is,

$$D(x,y) = D_m J_m(k_T\rho)e^{-ik_L z}$$

where, J_m is a Bessel function of first kind and order m. Specifically for $m = 0$, the complex amplitude is given by,

$$\Psi(x,y) = D_0 J_0(k_T\rho)e^{-ik_L z} \tag{1.43}$$

The intensity function is given by, $I(\rho,\phi,z) = |D_0|^2 J_0(k_T,\rho)$, which is independent of the z and is circularly symmetric. This makes sure that there is no spread of optical power. This wave is called *Bessel Beam*.

1.7 Characteristics of Plane Waves

In this section, we manifest the interference and diffraction nature of light within the wave optics approximation. When two or more monochromatic wave are simultaneously present in the same space-time coordinates, the principle of superposition applies. Note that, superposition does not apply to the optical intensity since the phase part (argument) is lost. Non existence of the phase information is the reason why ray optics is incapable of explaining interference effects. Once the notion of wave optics is considered, the concept of both intensity and phase is taken into account while ray optics mostly uses intensity information to explain optical effects. Based on the postulates of wave optics and its consequence, we proceed to explain interference and diffraction in the subsequent sections.

1.7.1 Interference

In general, the plane wave has a complex amplitude given by,

$$\Psi(\vec{r},t) = \Psi(\vec{r})\, e^{-i\omega t} \tag{1.44}$$

Consider two monochromatic waves of same angular frequency ω with their spatial part (complex amplitude) be $\Psi_1(r)$ and $\Psi_2(r)$. Now, consider the superposition of both the waves i.e.,

$$\Psi(\vec{r}) = \Psi_1(\vec{r}) + \Psi_2(\vec{r})$$

The intensity of the individual waves are, $I_1 = |\Psi_1(\vec{r})|^2$ and $I_2 = |\Psi_2(\vec{r})|^2$ respectively. We proceed to find out the intensity distribution resulting from the superposition of both the waves i.e.,

$$\begin{aligned} I &= \left|\Psi_1(\vec{r}) + \Psi_2(\vec{r})\right|^2 \\ &= \left|\Psi_1(\vec{r})\right|^2 + \left|\Psi_2(\vec{r})\right|^2 + \Psi_1^*(\vec{r})\Psi_2(\vec{r}) + \Psi_2^*(\vec{r})\Psi_1(\vec{r}) \end{aligned} \tag{1.45}$$

Noting that, $\Psi_1(\vec{r}) = \sqrt{I_1}e^{-i\varphi_1}$, $\Psi_2(\vec{r}) = \sqrt{I_2}e^{-i\varphi_2}$, and substituting this along with their complex conjugates in the above equation, we get,

$$\begin{aligned} I &= I_1 + I_2 + \sqrt{I_1 I_2}e^{i\varphi_1}e^{-i\varphi_2} + \sqrt{I_1 I_2}e^{-i\varphi_1}e^{i\varphi_2} \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\varphi_1 - \varphi_2) \end{aligned}$$

Now, considering that the intensity of both the beams are equal i.e., $I_1 = I_2 = I'$ (say), we get,

$$I = 2I'\left[1 + \cos(\varphi_1 - \varphi_2)\right] = 4I'\cos^2\left(\frac{\varphi_1 - \varphi_2}{2}\right) \tag{1.46}$$

This is the interference equation for two monochromatic waves of same frequency. It is clear that, maximum intensity is $4I_0'$, and the minimum is zero. This is incontrast to the fact that, the intensity addition can utmost have a maximum of $2I_0'$. The output pattern is sinusoidal in nature with maximum and minimum of $4I_0'$ and 0 respectively. For odd integer multiple of $\pi/2$ in the phase term i.e., $(\varphi_1 - \varphi_2)/2 = qn/2$, the total intensity is the sum of constituent intensities $2I_0'$ while, for even integer multiple, the maximum intensity reaches $4I_0'$. This describes the wave nature of light which is due to the remarkable property of light called phase. It is the wave optics that incorporates the notion of phase unlike ray-optics. Dependence of resultant intensity on the phase of light is the backbone of many optical devices and imaging techniques including phase-contrast microscopy.

It must be noted that, coherence in time (temporal coherence) is a necessary condition for interference to occur. This necessitates that, both the waves must be derived from a single source. At this stage, it becomes important to realize that, light inherit randomness because of the fluctuation of light source which is due to the emission of light is from a large number of atoms or molecules emitting independently. So, there is no fixed phase between two emissions occurring at two different locations (atoms) in a random light source. But emissions can be linked or correlated in time. Accordingly, the temporal coherence is the memory time of the light source (atoms) in which the wavefunction fluctuates in unison at two instants of

time separated by a time interval. In this context, it becomes important to know the extent (in time) to which it maintains some correlation. This is measured by the auto-correlation function i.e.,

$$G(\tau) = \left\langle \Psi^*(t)\Psi(t+\tau) \right\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \Psi^*(t)\, \Psi(t+\tau)\, dt \tag{1.47}$$

where, the spatial coordinate is dropped (since $\vec{r}$ is fixed), so that $\Psi(r,t)$ is replaced by $\Psi(t)$. $G(\tau)$ is also called *temporal coherence function*.

We now understand that, it is the extent of fluctuation that determine temporal coherence than the absolute intensity function. So, one can modify the temporal coherence function to make it more sensitive to intensity fluctuations rather than average intensity. This can be accomplished by introducing a new function $g(\tau)$,

$$g(\tau) = \frac{G(\tau)}{G(0)} = \frac{\langle \Psi^*(t)\Psi(t+\tau) \rangle}{\langle \Psi^*(t)\Psi(t) \rangle} \tag{1.48}$$

In this formulation the absolute value of $g(\tau)$ is bounded between 0 and 1 i.e., $0 \leq |g(\tau)| \leq 1$. $|g(\tau)|$ is the measure of degree of correlation between two instances of Ψ i.e., $\Psi(t)$ and $\Psi(t+\tau)$. For deterministic and monochromatic light, $\Psi(t) = Ae^{-i\omega t}$, implying, $g(\tau) = e^{-i\omega\tau}$ and $|g(\tau)| = 1\ \forall\tau$. This states that the light is completely correlated for all times i.e., it is the perfect replica of itself at two different instances of time.

The measure of the 'memory-time' over which the fluctuations remembers itself is known as the *coherence time*. For most of the practical purpose it is defined as, $1/e$ of temporal coherence function. Formally, coherence time is defined as,

$$\tau_c = \int_{-\infty}^{\infty} \left|g(\tau)\right|^2 d\tau \tag{1.49}$$

So, the fluctuations are strongly correlated for $\tau \leq \tau_c$ and weakly correlated for $\tau > \tau_c$.

Another important quantity associated with temporal coherence is the spectral-width (or line-width) $\Delta\nu$ of the spectral density $S(\nu)$. Spectral density is a Fourier domain representation of $g(\tau)$ and their widths are inversely related ($\Delta\nu_c = 1/\tau_c$). Spectral-width is defined as,

$$\Delta\nu_c = \int_0^{\infty} S(\nu)\, d\nu \tag{1.50}$$

So, a light of narrow spectrum has a long coherence time and vice-versa.

Now, consider the superposition of two plane waves for which the intensity can be rewritten as,

$$\begin{aligned} I &= \left\langle |\Psi_1 + \Psi_2|^2 \right\rangle = \left\langle |\Psi_1|^2 \right\rangle + \left\langle |\Psi_2|^2 \right\rangle + \left\langle \Psi_1^*\Psi_2 \right\rangle + \left\langle \Psi_1\Psi_2^* \right\rangle \\ &= I_1 + I_2 + G_{12} + G_{12}^* = I_1 + I_2 + 2\,\mathrm{Re}\{G_{12}\} \end{aligned}$$

Since, $g_{12} = \frac{\langle \Psi_1^*\Psi_2 \rangle}{\sqrt{I_1 I_2}}$, so this modifies to,

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\,\mathrm{Re}\{g_{12}\}$$

One can express this in terms of the magnitude and phase of g_{12} i.e., $|g_{12}|\, e^{-i\varphi_g}$,

$$= I_1 + I_2 + 2\sqrt{I_1 I_2}|g_{12}|\cos(\varphi_g) \tag{1.51}$$

It must be noted that, for two completely correlated waves with, $|g_{12}| = 1$, we recover the original interference expression (1.45). On the other hand, $|g_{12}| = 0$ for two uncorrelated waves and the intensity is just the sum of two intensities.

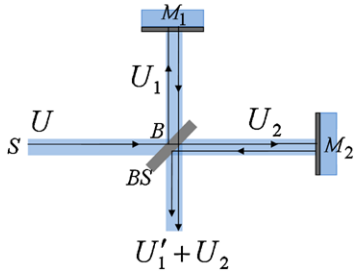


Fig. 1.15 A simplified version of Michelson interferometer

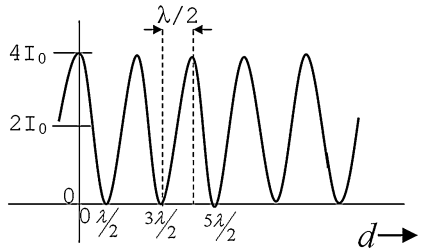


Fig. 1.16 Interference pattern (phase dependent intensity variation) for Michaelson's interferometer

Michelson's Interferometer

This is probably the simplest interferometer that reveals all the key features of interferometry. The schematic of the Michelson's interferometer is shown in Fig. 1.15. An input beam is splitted into two beams U_1 and U_2 using a beam-splitter (BS). They travel different paths via route BM_1B and BM_2B followed by recombination. The beams are recombined by the same beam-splitter BS where they undergo superposition to create a superimposed beam.

Considering the complex amplitude of beam 1 and 2 as $\Psi_1 = U_1 = \sqrt{I_0}e^{-ikz}$ and $\Psi_2 = U_2 = \sqrt{I_0}e^{-ik(z+d)}$. The phase-difference is given by, $\varphi = \varphi_2 - \varphi_1 = kd = 2\pi d/\lambda$. So, the resultant intensity of the sum of these two waves is given by,

$$I = 2I_0\big[1 + \cos(2\pi d/\lambda)\big] \tag{1.52}$$

The resultant interference pattern is shown in Fig. 1.16. It is evident that, the intensity varies periodically with the path-length d. For odd integer multiple of $\lambda/2$, the intensity is 0 indicating destructive interference, whereas for even integer multiple of $\lambda/2$, the intensity is maximum which refers to constructive interference.

1.8 Diffraction of Light and Fourier Optics

The phenomenon of diffraction is associated with the bending of light at the macroscopic edges. For example, consider the transmission of light through a circular aperture and traveling through some distance before hitting a screen. It is observed that, light gets diffracted at the edges of the macroscopic aperture and eventually a large hole is observed on the screen along with associated circular pattern. This pattern is called diffraction pattern. The same thing can be translated to any arbitrary shape aperture. To better illustrate, consider the diffraction from an aperture as shown in Fig. 1.17. Light from a point source is collimated and is made to fall on the aperture-stop. Now, according to Huygens-Fresnel principle, each point on the wavefront (emerging from the aperture (see, Fig. 1.17)) is a source of secondary waves. Superposition of these waves on the observation plane (placed far from the aperture plane) create diffraction pattern.

The complex amplitude $g(x, y)$ at the observation plane (placed at a distance d from the aperture plane) may be described based on two different approximations: Fraunhofer or Fresnel. These approximation determine the nature of free-space propagation. Corresponding diffraction pattern, $|g(x, y)|^2$ is termed as Fraunhofer or Fresnel based on the approximation used. Although this approach is fairly accurate for most of the aperture geometries but cannot be exact. Ideally, one must use paraxial Helmholtz equation and subject it to boundary conditions imposed by the aperture to get the exact solution. Mathematically, it is complex and difficult. Moreover, venturing into the depth of diffraction theory will divert our attention from the theme of the book. So, we will limit ourselves to the simple theory of diffraction and encourage interested readers to other texts [1]. Even to develop this simple theory of light, we need to introduce optical Fourier transform.

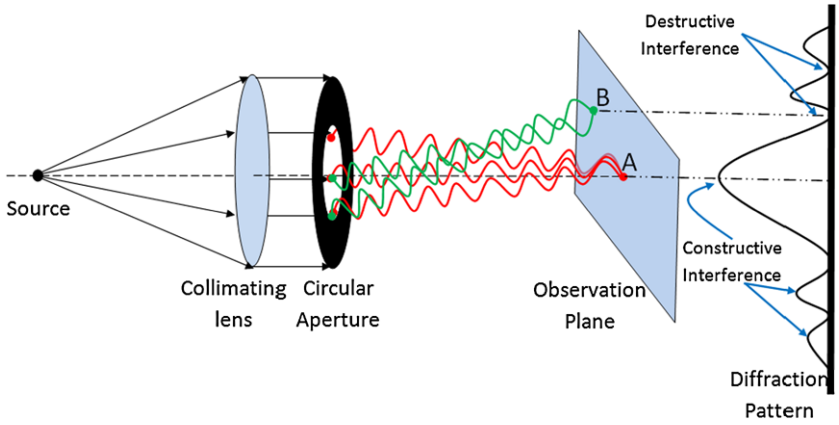


Fig. 1.17 Diffraction pattern of a circular aperture

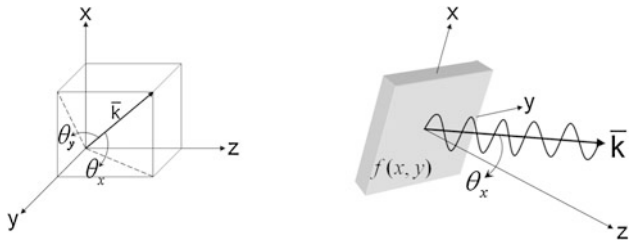


Fig. 1.18 A plane wave travelling at angles, θ_x & θ_y after emerging from an optical element can be represented as a harmonic function of spatial frequencies, ν_x & ν_y respectively

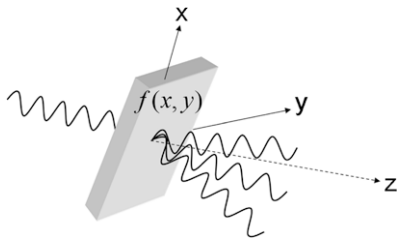


Fig. 1.19 The decomposition of incident plane wave into many plane waves as it emerges from an optical element

1.8.1 Spatial Approximation of Paraxial Plane Waves

To start with, consider a plane wave with complex amplitude, $\Psi(\vec{r}) = \Psi_0 e^{-i\vec{k}.\vec{r}}$, where, $\vec{k} = (k_x, k_y, k_y)$, wave number $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$. The vector components are, $k_x = |\vec{k}| \sin\theta_x$, $k_y = |\vec{k}| \sin\theta_y$ and $k_z = |\vec{k}| \sin\theta_z$. In other words, the vector $\vec{k}$ makes an angle, $\theta_x = \sin^{-1}(k_x/|\vec{k}|)$ with the YZ-plane. Similarly, for y- and z-direction, we have $\theta_y = \sin^{-1}(k_y/|\vec{k}|)$ and $\theta_z = \sin^{-1}(k_z/|\vec{k}|)$ (see, Fig. 1.18). For $z = 0$ plane, the complex amplitude is given by, $\Psi(x, y) = \Psi_0 e^{-i2\pi(\nu_x x + \nu_y y)}$, with spatial frequencies, $\nu_x = k_x/2\pi$ and $\nu_y = k_y/2\pi$ along x- and y-axis respectively. *At this point it is important to note that, the spatial frequencies has the unit cycles/mm and optical frequency* $\nu = \omega/2\pi = ck/2\pi$ *is measured in cycles/sec or Hz.* The angles can be expressed in spatial frequencies as, $\theta_x = \sin^{-1}(\nu_x\lambda)$ and $\theta_y = \sin^{-1}(\nu_y\lambda)$. In the paraxial approximation where, the spatial wave number is negligible as compared to optical wave number, $k_x \ll k$ and $k_y \ll k$, it can be safely assumed that, $\sin\theta_x \approx \theta_x$ and $\sin\theta_y \approx \theta_y$, so that,

$$\theta_x \approx \lambda\nu_x \quad \text{and} \quad \theta_y = \lambda\nu_y. \tag{1.53}$$

The basic fact that, a wave (of complex amplitude, $\Psi(x, y, 0) = e^{-i(\nu_x x + \nu_y y)}$) incident on a thin optical element gets converted into many plane waves propagating at angles θ_x and θ_y with spatial frequencies, $\nu_x = k_x/2\pi$ and $\nu_y = k_y/2\pi$, and amplitudes $O(\nu_x, \nu_y)$ (see, Fig. 1.19). More generally, $f(x, y)$ can be expressed as a superposition integral over all spatial frequencies i.e.,

$$f(x, y) = \Psi(x, y, 0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} O(\nu_x, \nu_y) e^{-i2\pi(\nu_x x + \nu_y y)} d\nu_x d\nu_y$$

The transmitted wave at (x, y, z) is the superposition of plane waves i.e.,

$$\Psi(x, y, z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} O(\nu_x, \nu_y) e^{-i2\pi(\nu_x x + \nu_y y)} e^{-ik_z z} d\nu_x d\nu_y \tag{1.54}$$

where, $k_z = \sqrt{(k^2 - k_x^2 - k_y^2)}$ is the component of propagating wave-vector along z-axis. It is important to note that, $O(\nu_x, \nu_y)$ is the Fourier transform of $f(x, y) = \Psi(x, y, 0)$.

1.8.2 Free-Space Propagation

Once spatial representation of plane wave is taken care-off, we are ready to understand their free-space propagation. Consider the propagation of plane wave from $z = 0$ to $z = z_0$. Corresponding complex amplitude functions are respectively $f(x, y) = \Psi(x, y, z = 0)$ and $g(x, y) = \Psi(x, y, z = z_0)$. Linearity and shift-invariance nature of the Helmholtz equation makes the input and output system as linear and shift-invariant. In such a system, the input and output are linked by impulse response function, $h(x, y)$ for which the transfer function is $H(\nu_x, \nu_y)$, ν_x and ν_y being the spatial frequencies along x- and y-direction [2]. The output harmonic function ($g(x, y) = \Psi(x, y, z = z_0) = Ae^{-i(k_x x + k_y y + k_z z_0)}$) is the product of input harmonic function ($f(x, y) = \Psi(x, y, z = 0) = Ae^{-i(k_x x + k_y y)}$) and its transfer function ($H(\nu_x, \nu_y)$). Hence, the transfer function which is the ratio of the input and output function is given by,

$$H(\nu_x, \nu_y) = \frac{g(x, y)}{f(x, y)} = e^{-i2\pi k_z z_0} \tag{1.55}$$

Realizing that, $k_x = 2\pi\nu_x$ and $k_y = 2\pi\nu_y$, we have,

$$k_z = \left(k^2 - k_x^2 - k_y^2\right)^{1/2} = 2\pi\left(\frac{1}{\lambda^2} - \nu_x^2 - \nu_y^2\right)^{1/2}$$

Substitution k_z into Eq. (1.55) gives,

$$H(\nu_x, \nu_y) = e^{-i2\pi z_0(\lambda^{-2} - \nu_x^2 - \nu_y^2)^{1/2}} \tag{1.56}$$

This is the general expression for the transfer function in free-space. Next, we derive the expression for two cases of interest i.e., transfer function for near field (*Fresnel Approximation*) and far-field (*Fraunhofer Approximation*).

1.8.3 Fresnel Approximation and Its Free-Space Propagation

The complex amplitude of plane wave propagating from $z = 0$ to $z = z_0$ are related by (1.55). In this sub-section, we move forward to understand the behavior of plane wave propagation under Fresnel approximation. This approximation states that, "*The input complex amplitude* $f(x, y)$ *contains spatial frequencies that are much smaller than the cutoff frequency i.e.*, $\nu_x^2 + \nu_y^2 \ll \lambda^{-2}$". So, the plane wave makes an angle θ with the optical axis, and hence, $\theta^2 = \theta_x^2 + \theta_y^2 \approx \lambda^2(\nu_x^2 + \nu_y^2)$. Substitution produces the following phase factor,

$$2\pi z_0\left(\frac{1}{\lambda^2} - \frac{\theta^2}{\lambda^2}\right)^{1/2} \tag{1.57}$$

Expanding using Taylor series (see, Appendix A.1), we get,

$$\frac{2\pi z_0}{\lambda}\left[1 - \frac{\theta^2}{2} + \frac{\theta^4}{8} - \cdots\right]$$

Neglecting the terms beyond second order we get,

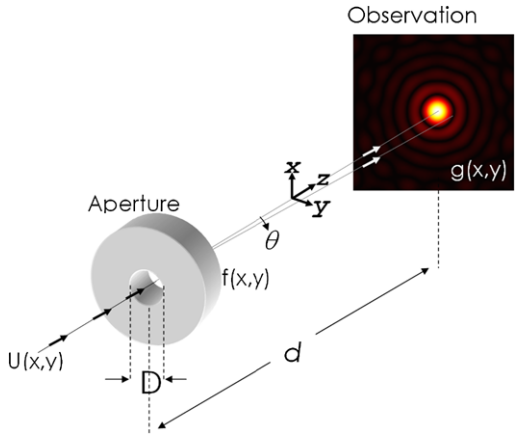


Fig. 1.20 The airy-disc pattern from a circular aperture under Fraunhofer approximation

$$H(\nu_x, \nu_y) \approx e^{-ikz_0} e^{i\pi\lambda z_0(\nu_x^2+\nu_y^2)} \tag{1.58}$$

This is the Fresnel approximation in frequency domain (ν_x, ν_y) for free-space propagation. Similarly one can express the Fresnel approximation in the spatial domain (x, y) by taking the inverse Fourier transform (see, Appendix B.2). The inverse Fourier transform of $H(\nu_x, \nu_y)$ turns out to be,

$$h(x, y) = (i/\lambda z_0) e^{-jkz_0}\ e^{-ik(x^2+y^2)/2z_0} \tag{1.59}$$

Next, we set out to show that, free space propagation is equivalent to a convolution operation. Consider, $f(x, y)$ and $g(x, y)$ as the input and output complex amplitudes. $f(x, y)$ can be thought of as the super-position of several points (mathematically, delta functions) each producing a near-spherical wave. The wave originating from (x', y') is centered at (x', y') and has an amplitude $f(x', y')$. This generates a wave with amplitude, $f(x', y')\ h(x - x', y - y')$ at the point (x, y) in the output plane, where $h(x, y)$ is the impulse response relating (x', y') (in the input plane) and (x, y) (in the output plane). So, the complex amplitude $g(x, y)$ is obtained by collecting contributions from all the points (considering infinite points) in the input plane i.e.,

$$g(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(x', y'\right) h\left(x - x', y - y'\right) dx'dy'$$

The output is the convolution of input function $f(x, y)$ and the impulse response $h(x, y)$. Using Fresnel Approximation, this becomes,

$$g(x, y) = h_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(x', y'\right)\ e^{-ik((x-x')^2+(y-y')^2)/2z_0} dx'dy' \tag{1.60}$$

where, $h_0 = (i/\lambda d)e^{-ikz_0}$. As evident, the input wave is expanded in terms of Gaussian/paraboloidal waves in a space-domain.

1.8.4 Fraunhofer Approximation and Its Far-Field Effect

To understand diffraction far from the input plane, let us consider the simplest theory of diffraction where, only the waves transmitting through the aperture survives (see Fig. 1.20). If $U(x, y)$ and $f(x, y)$ represents the complex amplitude of paraxial transverse plane wave just before and after the circular aperture $T(x, y)$ then,

$$f(x, y) = T(x, y) \times U(x, y) \tag{1.61}$$

where, the aperture function is given by, $T(x, y) = \begin{cases} 1, & r \le r_0 \\ 0, & r > r_0 \end{cases}$ and $r = \sqrt{x^2 + y^2}$ for a circular aperture.

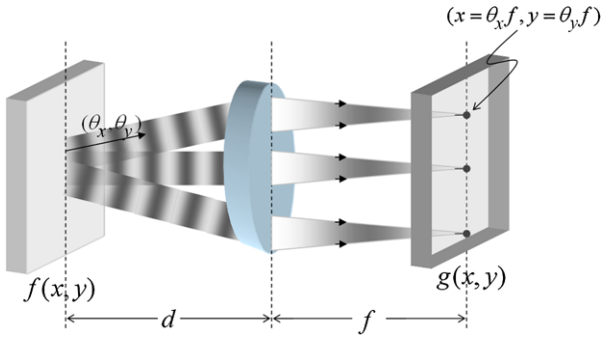


Fig. 1.21 Plane waves (which are the harmonic Fourier components of the function $f(x, y)$) with the direction (θ_x, θ_y) is focussed at the point $(x, y) = (\theta_x f, \theta_y f)$

The screen is placed at a distance d from the aperture plane. *Fraunhofer approximation states that, for long propagation distance, the only plane waves that contribute to the complex amplitude at a point* (x, y) *in the observation plane are the wave with direction making small angles* $(\theta_x \approx (x/d), \theta_y \approx (y/d))$ (considering the small angle approximation).

The angles along x- and y-axes are respectively, $\theta_x \approx (x/d)$ and $\theta_y \approx (y/d)$ with the optical axis. The corresponding wave vectors are, $k_x \approx (x/d)k$ and $k_y \approx (y/d)k$ and amplitude $F(\nu_x, \nu_y)$ with $\nu_x = x/\lambda d$ and $\nu_y = y/\lambda d$ (see, Sect. 1.8.1). Given that, $f(x, y)$ is the complex amplitude at $z = 0$, the complex amplitude at $z = d$ is,

$$g(x, y) = \Psi(x, y, z = d) \approx \left(\frac{i}{\lambda d}\right) F\left(\frac{x}{\lambda d} \frac{y}{\lambda d}\right) e^{-ikd} \tag{1.62}$$

where, $F(\nu_x, \nu_y)$ is the Fourier transform of the aperture function $T(x, y)$.

This is the expression for free-space propagation under Fraunhofer Approximation. Unlike Fraunhofer diffraction, Fresnel diffraction accounts for the curvature of the wavefront, in order to correctly calculate the relative phase of interfering waves. As an example, the readers are encouraged to derive the expression for Fraunhofer diffraction pattern due to rectangular aperture of size D_x and D_y observed at a distance d. Verify that, the first zero occurs at $x = \pm\lambda d/D_x$ and $y = \pm\lambda d/D_y$. Compute and show that, $2D$ diffraction pattern due to a rectangular window of arbitrary size. The window size may be varied to observe the changes in the diffraction pattern. It would be informative to see the pattern for $D_x \gg D_y$ (slit).

1.9 Optical Imaging Systems and Microscopy

In this section, we will use both ray and wave optics to achieve a bigger goal i.e., to design and analyze optical systems such as, $4 - f$ imaging system, dark field microscopy and ultimately DIC microscopy. These microscopy techniques are often moulded and used to compliment fluorescence microscopy.

1.9.1 Single Lens Based Imaging System

We will start with the simplest optical imaging system one can think of i.e., single lens based imaging. It must be bourne in mind that, the optical system with single lens can only be treated under Fresnel approximation that assumes that the spatial frequencies of the input complex function $f(x, y)$ are much smaller than the cutoff frequency (λ^{-1}). By now, it is clear that, a thin lens transforms a plane wave into a spherical wave (or more appropriately paraboloidal wave). A plane wave arriving at an angle θ_x and θ_y gets focussed at the point (x, y) (see, Fig. 1.21). Similarly, other plane waves arriving from different directions also gets mapped to the focal plane. This concludes that, the lens maps a series of plane wave to points on the focal plane. Let $f(x, y)$ be the amplitude of the plane wave in the input plane $(z = 0)$. The amplitude at the output plane is a function of spatial frequency i.e., ν_x and ν_y or equivalently the propagation direction i.e., $\theta_x = \lambda\nu_x$ and $\theta_y = \lambda\nu_y$. Let us symbolize this by, $F(\nu_x, \nu_y)$. Similarly, we take into consideration all the plane waves propagating along different directions i.e., we have to integrate over all the propagation direction or equivalently spatial frequencies to get the output function $g(x, y)$ at the focal plane,

$$g(x, y) \propto F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right). \tag{1.63}$$

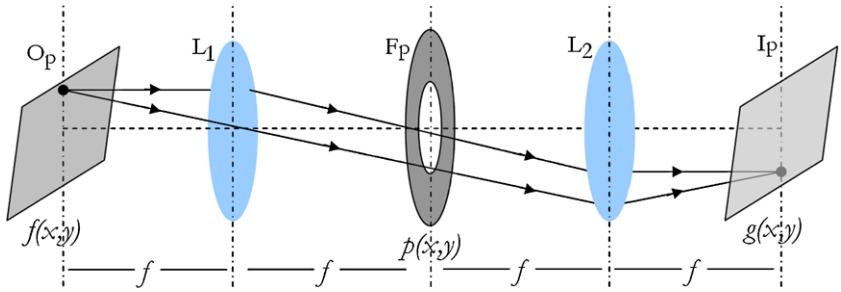


Fig. 1.22 $4f$ imaging system with unit magnification

Next, we need to determine the proportionality factor. Assuming paraxial approximation for the waves and using Fresnel approximation (see, Sect. 1.8.3) we get,

$$g(x, y) = (j/\lambda f)\, e^{-jk(d+f)}\, e^{j\pi(x^2+y^2)(d-f)/\lambda f d}\, F(x/\lambda f, y/\lambda f). \tag{1.64}$$

Noting that, the input plane is at the back-focal plane, i.e., $d = f$, we get,

$$g(x, y) = (j/\lambda f) e^{-jk(2f)} F(\nu_x, \nu_y). \tag{1.65}$$

This geometry is known as $2f$ optical system in which, the input plane at $-f$ is the Fourier transform of the output plane at f, and vice-versa.

The optical intensity at the focal plane is given by,

$$I(x, y) = \left|g(x, y)\right|^2 = \frac{1}{\lambda f}\left|F(\nu_x, \nu_y)\right|^2 \tag{1.66}$$

So, both the planes located at a back- and front-focal length of the lens are related by Fourier transform in both phase and magnitude. Object placed at the input plane can be imaged at the output plane and vice-versa. In short, we have shown that, single lens performs Fourier Transform.

1.9.2 $4f$ Imaging System

$4f$ system serves as a focussed imaging system with unit negative magnification. Recall that, negative magnification corresponds to the fact that, image is inverted. From earlier sub-sections, it is clear that a single lens performs Fourier transform. $4f$ system is a combination of two lens placed at a distance of $2f$ from each other along the optical axis (see, Fig. 1.22). The object is placed at the back-focal length of the first lens and the image is obtained at the focus of the second lens. Since, both the lens have same focal length, the distance from object plane to image plane is $4f$, hence the name, $4f$ imaging system.

It is important to recognize that, $4f$ imaging system is essentially a cascade of two sub-systems: first object O_P to Fourier plane F_P of first lens L_1 and, second is from Fourier plane of first lens to the image plane I_P through the second lens L_2. When exposed to a plane wave, the first lens L_1 performs Fourier transform, and subsequently, the second lens L_2 does an inverse Fourier transform to get back the perfect replica of the object. Let us consider that, the complex amplitude of the plane wave incident on the first lens be, $f(x, y) = A_0 e^{-ikz}$ travelling in the z-direction. Let $g(x, y)$ be the complex amplitude in the image plane. The first lens decomposes $f(x, y)$ into its Fourier components (spatial frequencies). The second lens then recombines these spatial frequencies thereby perfectly reconstructing the object.

$4f$ imaging system can be used in a way to block certain spatial frequencies and allow others. Note that, the Fourier components of $f(x, y)$ in the Fourier plane are evaluated at the spatial frequency (ν_x, ν_y) and is located in the fourier plane at the point $(\lambda f \nu_x, \lambda f \nu_y)$. A mask of transmission function $p(x, y)$ placed at the Fourier plane must be proportional to the transfer function $H(\nu_x, \nu_y)$ i.e.,

$$H(\nu_x, \nu_y) = p(x, y) = p(\lambda f \nu_x, \lambda f \nu_y) \tag{1.67}$$

The inverse Fourier transform of transfer function gives the impulse response function in the spatial domain i.e.,

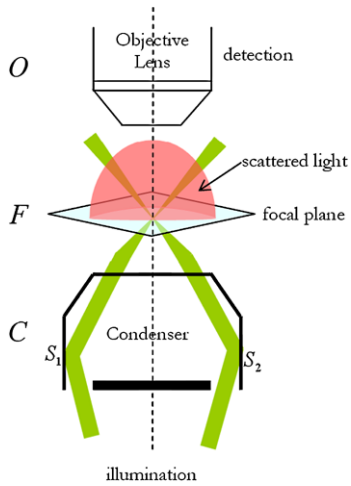


Fig. 1.23 Schematic of dark field microscopy, in which angular illumination is preferred

$$h(x,y) = \frac{1}{(\lambda f)^2} P\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) = \frac{1}{(\lambda f)^2} P(\nu_x, \nu_y) \tag{1.68}$$

This is the impulse response function for $4-f$ system (see, Fig. 1.22). For high-pass filtering, the transfer function has the form, $H(\nu_x, \nu_y) = \begin{cases} 1, & \nu_x^2+\nu_y^2 > \nu_c^2 \\ 0, & \nu_x^2+\nu_y^2 \leq \nu_c^2 \end{cases}$. This filter stops spatial frequencies that are smaller than cutoff frequency ν_c and allows higher spatial frequencies. As an example, consider $\nu_c = D/2\lambda f = 40$ lines/mm, that corresponds to, $\lambda = 0.5$ µm, $D = 40$ mm and $f = 1000$ mm. This filter retains spatial frequencies that are greater than 40 lines/mm. A similar filter can be designed for low-pass filtering operation, which is the compliment of high-pass filter. Moreover, it is also possible to design filters that transmits or blocks all the horizontal or vertical frequencies.

1.9.3 Dark Field Microscopy

A simplest way to obtain high contrast image in an optical imaging module is to completly eliminate the incident light in the detection process. This makes sure that, the background is dark enough for the visualization of faint objects. In a normal transmission light microscopy, the background is quite bright and overshadows the signal. In most of the transmission microscopy, both the non-diffracted and diffracted rays are collected by the objective lens for image formation. The idea behind the darkfield microscopy is the complete elimination of non-diffractive contribution for image formation. It is the diffracted rays that interact with the sample and not the non-diffracted rays which are basically the unperturbed rays that pass through the specimen. Due to the presence of only the diffractive component, this technique is very sensitive, and hence the minute diffracting specimens such as, biological polymers (microtubules, actin filaments etc.) are better visualized.

The schematic ray diagram of a typical darkfield microscopy is shown in Fig. 1.23. The specimen F is illuminated at an oblique angle θ_0 so that the non-diffracting rays are not collected by the objective lens O. Special cardioid condensers C are used to direct the light at an angle. These condensor receives nearly collimated beam reflected from both the surfaces S_1 and S_2 for creating steep angle. In the darkfield setup (see Fig. 1.23), the output consists of two components: Scattered and Unscattered. The detection objective is placed in such a way (in-between the angles on the detector side), so as to collect only the scattered light. This severely reduces the background generated by the unscattered light. Note that, the illumination by this condensor is aplanatic and so is free from spherical aberration. These condensers are capable of maintaining darkfield condition at an NA upto 1.0. Darkfield images have superior contrast because of the dark background. It is important to realize that such a concept can be extended to fluorescence microscopy as well.

1.9.4 Differential Interference Contrast (DIC) Microscopy

DIC microscopy is based on the principles of shear interferometry. Two wavefronts derived from a single source are slightly shifted to create an interference pattern. The phase of one of the wavefronts is interfered with a shifted version of itself. The shift is brought-in by the refractive index of the specimen. Let the phase of one of the wavefront be $\varphi(x)$ and the shifted version has the phase $\varphi(x+\Delta x)$. Then, the phase shift between two interfering wavefronts is given by,

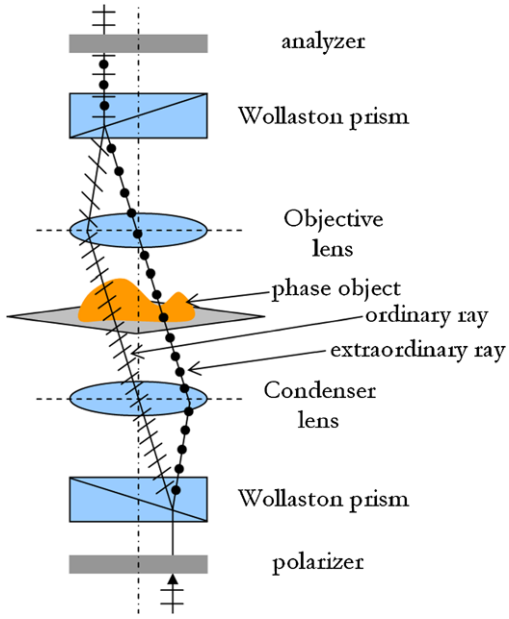


Fig. 1.24 Schematic of differential interference contrast microscopy for imaging phase objects

$$\Delta\varphi = \varphi(x + \Delta x) - \varphi(x) = \frac{d\varphi}{dx}\Delta x \tag{1.69}$$

So, the intensity distribution of the resultant interference pattern is solely governed by the interference term (see, Sect. 1.7.1) i.e.,

$$I(x) = 4I' \cos^2\left(\frac{d\varphi}{dx}\Delta x\right) \tag{1.70}$$

where, $\Delta\varphi$ is the phase term, and Δx is the shear between the wavefronts.

Practically, the simplest DIC microscopy system uses two Wollaston prism (see Fig. 1.24). The first prism splits the beam into ordinary and extraordinary components. One of the components is made to pass through the sample of interest and the other component goes unperturbed, thereby causing a phase-delay to the first component. This is then followed by a condensor lens and recombined using Wollaston prism-analyzer module (see Fig. 1.24). The polarizer is rotated by 45° with respect to the shear axis of the prisms. A simpler expression can be obtained by assuming certain simple approximations [3]. These approximations are as follow: (1) $\Psi(x, y) \approx \Psi(x + \Delta x, y) = C$ (*say*), (2) $|\varphi(x + \Delta x) - \varphi(x)| \ll \pi/4$ and, (3) $\varphi_0 = -\pi/2$. Using these and the phase bias φ_0 that is introduced by the translated prism, the interference term becomes, $\cos(\frac{d\varphi}{dx}\Delta x - \varphi_0) = \cos(\frac{d\varphi}{dx}\Delta x + \pi/2) = -\sin(\frac{d\varphi}{dx}\Delta x)$. So, the intensity in the image is given by,

$$I = 4I' \sin^2\left[\left(\frac{\partial\varphi}{\partial x}\Delta x\right)\right] \tag{1.71}$$

The advantage of this imaging modality is that, transparent objects that have varying thickness and varying refractive index can be imaged.

It must be noted that, the above mentioned optical systems are some of the simple optical imaging systems directly based on the theory of ray optics and wave optics. There are several other imaging systems that use these two theories along with the advanced EM theory and quantum optics that we plan to cover in the next few chapters. We will explain these advanced microscopy techniques in Chaps. 3, 4, 8 and 9, when we are better equipped with EM theory and the basics of fluorescence. In these chapters, we will describe confocal microscopy, multiphoton microscopy, TIRF, STED, 4π, structured illumination, SPIM and others.

1.10 Optical Resolution Limit

One of the key property of any optical microscopy technique is the system resolution. Classically, resolution is the ability to distinguish two nearby objects as distinct objects. We plan to derive an approximate expression for determining the lower bound on resolution based on the diffraction from a point source (small aperture).

From Sect. 1.8, the diffraction pattern at a point (x, y) under Fraunhofer approximation is proportional to the square magnitude of the Fourier transform of the aperture function $k(x, y)$ evaluated at spatial frequencies $\nu_x = x/\lambda d$ and $\nu_y = y/\lambda d$ i.e.,

$$I(x, y) = \left(\frac{\sqrt{I_0}}{\lambda d}\right)^2 \left|K(x/\lambda d, y/\lambda d)\right|^2 \tag{1.72}$$

For most of the imaging conditions, we deal with an aperture or optical elements (such as, objective lens, condensors etc.) of circular dimensions. So, we choose the aperture function to be $k(x, y) = \begin{cases} 1, & r \le (D/2) \\ 0, & r > (D/2) \end{cases}$ and $r = \sqrt{x^2 + y^2}$, where D is the aperture dimension. The next step is to find the Fourier transform of the aperture function i.e.,

$$K(\nu_x, \nu_y) = \int_{x=0}^{\infty} \int_{y=0}^{\infty} k(x, y) e^{-i2\pi(x\nu_x + y\nu_y)} dx dy \tag{1.73}$$

To exploit the circular symmetry of the aperture, we make a transformation to polar coordinates in both the spatial domain (x, y) and the Fourier domain (ν_x, ν_y), as follows:

$$x = r\cos\theta, \qquad y = r\sin\theta \quad \Rightarrow \quad r = \sqrt{\left(x^2 + y^2\right)}, \qquad \theta = \tan^{-1}(y/x). \tag{1.74}$$

and,

$$\nu_x = \rho\cos\phi, \qquad \nu_y = \rho\sin\phi \quad \Rightarrow \quad \rho = \sqrt{\left(\nu_x^2 + \nu_y^2\right)}, \quad \phi = \tan^{-1}(\nu_y/\nu_x). \tag{1.75}$$

In polar coordinates, the Fourier transform is given by,

$$\begin{aligned} K(\rho, \phi) &= \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} k(r, \theta) e^{-i2\pi r\rho(\cos\theta\cos\phi + \sin\theta\sin\phi)} r\, dr\, d\theta \\ &= \int_{r=0}^{D/2} \left[\int_{\theta=0}^{2\pi} e^{-j2\pi r\rho\cos(\theta-\phi)} d\theta\right] r\, dr \end{aligned} \tag{1.76}$$

Let, $\xi = 2\pi r\rho$, so, $d\xi = 2\pi\rho dr$. Substitution gives,

$$\begin{aligned} &= \frac{1}{2\pi\rho^2} \int_{\xi=0}^{\pi D\rho} \left[\frac{1}{2\pi} \int_{\theta=0}^{2\pi} e^{-i\xi\cos(\theta-\phi)} d\theta\right] \xi\, d\xi \\ &= \frac{1}{2\pi\rho^2} \int_{\xi=0}^{\pi D\rho} J_0(\xi)\xi\, d\xi \end{aligned}$$

Using the identity, $\int_0^{x_0} x J_0(x)\, dx = x_0 J_1(x_0)$, and noting that $\rho = \sqrt{\nu_x^2 + \nu_y^2} = r/\lambda d$ we get,

$$K(\rho, \phi) = \frac{1}{2\pi\rho^2} \pi D\rho J_1(\pi D\rho) \tag{1.77}$$

So, the diffraction pattern under Fraunhofer diffraction is given by,

$$I(x, y) = \left(\frac{\pi D^2 I_0}{(\lambda d)^2}\right) \left|\frac{J_1(\pi Dr/\lambda d)}{\pi Dr/\lambda d}\right|^2 \tag{1.78}$$

where, $r = \sqrt{x^2 + y^2}$.

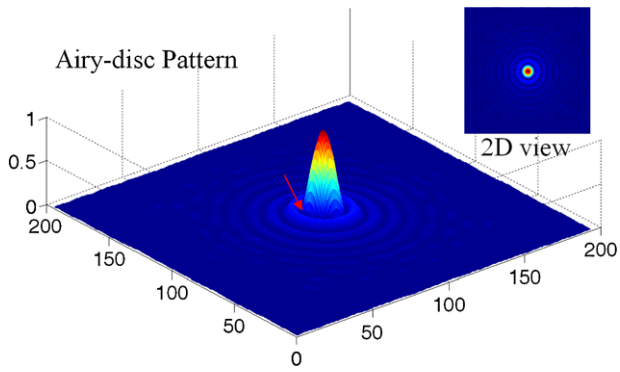


Fig. 1.25 Classical Airy-disc pattern generated by a circular aperture

The diffraction pattern generated using (1.78) consists of bright central lobe surrounded by circular rings of decreasing intensity pattern as shown in Fig. 1.25. This is popularly called as Airy pattern. Considering a plane wave focussed at the back of the aperture (size $\approx D$) followed by focussing using a convex lens (focal length, f), the intensity pattern is given by,

$$I(x, y) = \left(\frac{\pi D^2 I_0}{(\lambda f)^2}\right)\left|\frac{J_1(\pi Dr/\lambda f)}{\pi Dr/\lambda f}\right|^2 \tag{1.79}$$

At this point, it is important to realize that, the first minima of the diffraction pattern defines the resolution limit of an imaging system (lens in this case). This is due to the fact that, two nearby airy-pattern can be distinguished when the maximum (central lobe) of first pattern falls on the first minimum (side lobe) of the second pattern. Closer than this distance, the central lobes of both the airy-pattern appear as a single lobe and hence are indistinguishable. Note that, the first minima occur at, $I(x, y) = 0$, i.e., $J_1(\pi Dr/\lambda d) = 0$. First zero of Bessels function occurs at 3.8317 (from standard table of Bessels function as shown in Appendix C, Table C.1) i.e.,

$$\left.\frac{\pi Dr}{\lambda d}\right|_{r=r_1} = 3.8317 \quad \text{(first minimum)}$$

$$r_1 = \frac{1.22\lambda f}{D}$$

In this case, the radius of the focal spot is given by,

$$\zeta = r_1 = 1.22\frac{\lambda}{(2\frac{D}{2})/f} = 1.22\frac{\lambda}{2\sin\alpha} \tag{1.80}$$

where, α is the semi-aperture angle of the lens.

Now, considering a medium of regractive index n for liquid-immersion lens/objective lens (of aperture angle, α) and the fact that, the wavelength in the medium is scaled by refractive index i.e., λ/n, we get,

$$\zeta \approx \frac{1.22\lambda}{2n\sin\alpha} \tag{1.81}$$

Classically, optical resolution has a limit approximately given by the above equation and one just can't resolve features beyond this limit. This is popularly known as Abbe's diffraction limit.

1.11 Optical Aberrations: Chromatic, Spherical and Astigmatism

The dispersion of the optical material used for lens manufacturing is responsible for chromatic aberration. Just to remind, dispersion is the bending or refraction of different spectral components of light based on the refractive index, as shown in Fig. 1.26(A). This means that refractive index is wavelength-dependent $n \approx n(\lambda)$ (see, Fig. 1.26(B)). Additionally, the speed

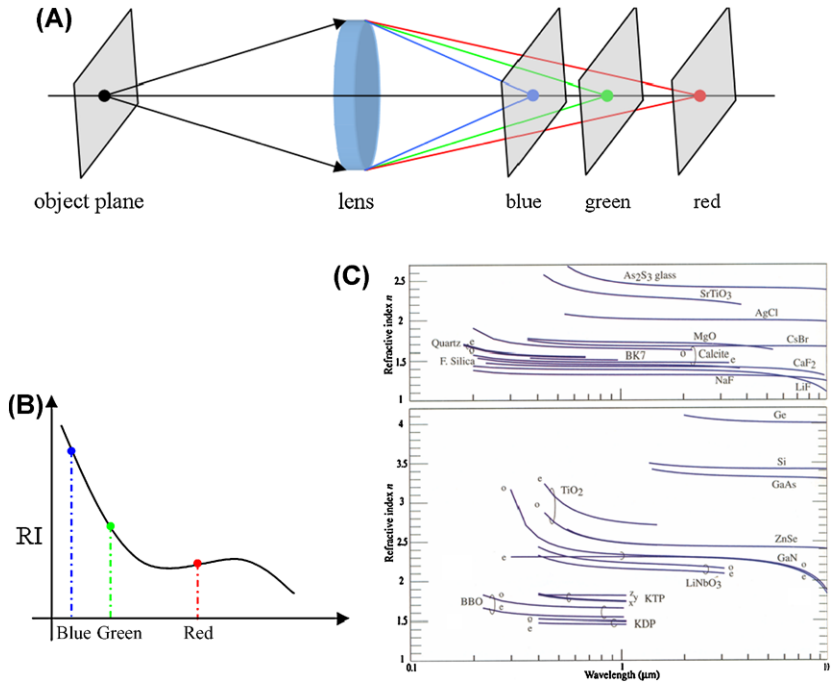


Fig. 1.26 (**A**) Component splitting of a polychromatic light, (**B**) Schematic of wavelength dependent refractive index, (**C**) Wavelength-dependent RI for selected optical materials

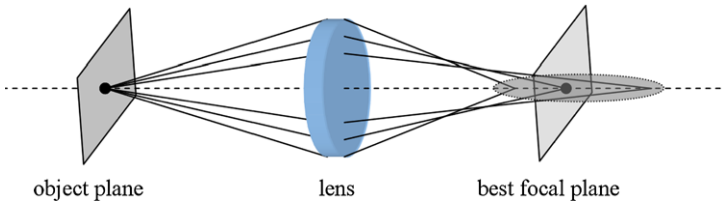


Fig. 1.27 A ray undergoing spherical aberration

of a particular spectral component (λ) is given by, $c_0/n(\lambda)$. So, different wavelengths of light gets refracted by various angles. When a lens is illuminated by a polychromatic light source, the light is focussed at different points along the optical axis. Similarly, a prism bends light of different wavelength by different angles according to Snell's law. The effect (i.e., refractive index dependence on wavelength) of these components which are made of different materials are shown in Fig. 1.26(C).

Chromatic aberration is severe especially when dealing with pulsed light which are polychromatic in nature. Inside the dispersive medium, each frequency component in the pulse travels at different velocity because each experiences a different time-delay. So, the red-component of the pulse can travel faster as compared to the blue component. As a result, the output pulse after travelling a long duration in the dispersive medium widens with red-component (long wavelength) appear before than blue-component (short wavelength). This is widely apparent in multi-photon fluorescence microscopy. For multi-photon excitation process, one needs laser with high peak power and thus very short pulse (often *femto-second* pulse) lasers are employed. Such pulses have a broad spectrum and the effects due to chromatic aberration are apparent. Figure 1.26(C), below show the dispersive nature of few well-known materials used for manufacturing optical components. In fluorescence microscopy, this effect is corrected by using achromatic objectives. These objectives are primarily built with a lower NA ($\approx$0.6) and moderate magnification ($\times$40). Additionally, when corrected for field curvature, they are called plan-achromats.

Next, we study spherical aberration effects in optical microscopy. This artifact is due to the fact that, rays at different height of the pupil gets focussed at different planes along the optical axis resulting in an axial blur. The effect is shown in Fig. 1.27. This is a consequence of manufacturing optical components with spherical curvature. The most common way to correct spherical aberration is to use a combination of plano-concave and plano-convex lens. In fluorescence microscopy, the thickness and refractive index of the coverglass along with the medium between the coverglass and sample is important for obtaining spherical-aberration-free imaging.

Astigmatism refers to the distortion of system PSF at different optical depths (depths measured from optical focus). In optical terms, this is the measure of ellipticity of a PSF on opposite sides of the best focal plane. This results in varying magnification at off-focal planes as shown in Fig. 1.28. Recently, the Astigmatism is used to realize super-resolution fluorescence microscopy. In 3D STORM, astigmatism imaging method was used to achieve 3D STORM imaging [4]. To achieve

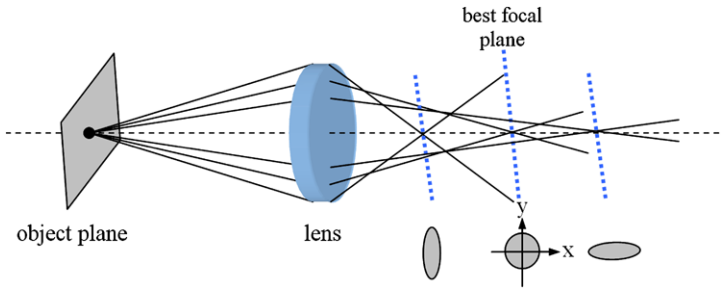


Fig. 1.28 A ray demonstrating astigmatism

3D, a cylindrical lens was introduced into the imaging path. This creates two slightly different focal planes along x and y directions. The position of the fluorescent dots change along z-axis and this is measured by its ellipticity. In the best focal plane, the dots inside the image appears round. Off-focal dots appears ellipsoidal with its long axis along x- or y-axis, depending upon the z-position. The z-coordinate can be obtained and pin-pointed from the ellipticity of the dots. This technique powered by Astigmatism has profound application in 3D fluorescence imaging.

Appendix: Recommended Reading

1. M. Born and E. Wolf, Principles of Optics, Cambridge University Press, 7th Edition, 2003.
2. E. Hecht, Optics, Addison-Wesley, 4th Edition, 2002.
3. T.C. Poon and T. Kim, Engineering Optics with MATLAB, World Scientific, 2006.
4. A.K. Ghatak and K. Thyagarajan, Introduction to Fiber Optics, Cambridge University Press, UK, 1998.
5. E. Hecht, Optics, Schaum's Outlines of Optics, Mc-Graw Hills, 2011.
6. J.A. Edminister, Schaum's Outlines of Electromagnetics, Second Edition, 2010.
7. D. Meschede, Optics, Light and Lasers, Wiley-VCH, Second Edition, 2007.
8. L. Novotny and B. Hecht, Principles of Nano-optics, Cambridge University Press, 2009.
9. J.D. Jackson, Classical Electrodynamics, John Wiley and Sons, 2007.
10. D.J. Griffiths, Introduction to Electrodynamics, Addison Wesley, Third Edition, 1999.
11. B.K. Johnson, Optics and Optical Instruments, Dover Publications, 1960.
12. J.W. Goodman, Introduction to Fourier Optics, Addison-Wesley, 2005.
13. R. W. Boyd, Nonlinear Optics, Academic Press, Third Edition, 2008.

References

1. Born, M., Wolf, E.: Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, 7th edn. Cambridge University Press, Cambridge (1999)
2. Goodman, J.: Introduction to Fourier Optics, 3rd edn. (2005). Roberts and Company Publishers (USA)
3. Holmes, T.J., Levy, W.J.: Signal-processing contrast of differential-interference-contrast microscopy. Appl. Opt. **26**, 3929 (1987)
4. Huang, B., Wang, W., Bates, M., Zhuang, X.: Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science **319**, 810–813 (2008)

http://www.springer.com/978-94-007-7544-2

Fundamentals of Fluorescence Microscopy
Exploring Life with Light
Mondal, P.P.; Diaspro, A.
2014, XV, 218 p. 140 illus., 55 illus. in color., Hardcover
ISBN: 978-94-007-7544-2