FUNDAMENTALS OF OPTICS

An Introductory Course

FUNDAMENTALS OF OPTICS

An Introductory Course

Yobani Mejía-Barbosa

Translated from the Spanish by
Herminso Villarraga-Gómez

UNIVERSIDAD
NACIONAL
DE COLOMBIA

SPIE PRESS
Bellingham, Washington USA

Library of Congress Control Number: 2022048509

English translation of Fundamentos de óptica: Curso introductorio
Copyright © 2021 UNAL
Translated by Herminso Villarraga-Gómez

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Published by
SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org

Background cover image from Shutterstock: Fouad A. Saad

Printed in the United States of America.
First printing 2023
For updates to this book, visit http://spie.org and type “PM359” in the search field.

Contents

Translator's Preface *xi*
Author's Preface *xiii*
Introduction *xv*

1 Geometrical Optics **1**
1.1 Rays or Waves 2
1.1.1 Camera obscura 2
1.1.2 Newton's corpuscular theory of light 4
1.1.3 Huygens' wave theory 5
1.1.4 Graphical ray tracing 8
1.2 Fermat's Principle 11
1.2.1 Modern formulation of Fermat's principle 14
1.2.2 Rays and wavefronts 15
1.2.3 Image from a point source 16
1.3 Refracting Surfaces 17
1.3.1 Modeling the cornea of a human eye 19
1.3.2 Refraction at spherical surfaces 21
1.3.3 Focal lengths and focal points 24
1.3.4 Focal planes 26
1.3.5 Paraxial imaging of extended objects 28
1.3.6 Optical power and vergence 29
1.4 Reflecting Surfaces 30
1.4.1 Ray tracing for spherical mirrors 31
1.4.2 The parabolic mirror 34
1.5 Lenses: Thin Lens Approximation 36
1.5.1 Ray tracing for thin lenses 38
1.5.2 Newton's lens equation 40
1.5.3 Real and virtual images domain 41
1.5.4 Focal planes in thin lenses 43
1.5.5 Ray tracing for oblique rays 44
1.6 Lenses: Principal Planes 46
1.6.1 A lens system 51
1.7 Stops and Pupils 54
1.7.1 Aperture stop 55

1.7.2 Pupils 58
1.7.3 Marginal and chief rays 60
1.7.4 Field stop, field of view, and angle size 62
1.8 Some Optical Instruments 65
1.8.1 The human eye (schematic representation) 65
1.8.2 Magnifiers 71
1.8.3 The telescope 75
1.8.4 The microscope 81
1.9 Monochromatic Optical Aberrations 84
1.9.1 Field curvature 85
1.9.2 Spherical aberration 87
1.9.3 Distortion 88
1.9.4 Astigmatism and coma 90
References 100

2 Polarization 103
2.1 Plane Waves and Polarized Light 104
2.1.1 Maxwell's equations with plane waves 105
2.1.2 Irradiance 106
2.1.3 Natural light and polarized light 108
2.1.4 Elliptical, circular, and linear polarization 109
2.1.5 Polarization: general case 111
2.2 Dichroism Polarization 114
2.2.1 Linear polarizer 114
2.2.2 Malus' law 116
2.3 Polarization by Reflection 118
2.3.1 Laws of reflection and refraction 118
2.3.2 Fresnel equations 121
2.3.3 Reflectance and transmittance 127
2.4 Polarization by Total Internal Reflection 130
2.4.1 Total internal reflection 130
2.4.2 Reflectance and transmittance 135
2.5 Polarization with Birefringent Materials 136
2.5.1 Phase retarder plates 138
2.5.2 Birefringent crystals 141
2.5.3 Refraction in crystals 144
2.5.4 Polarizing prisms 146
2.6 Vectors and Jones Matrices 149
References 154

3 Interference 155
3.1 Interference and Coherence 156
3.1.1 Degree of coherence 158
3.1.2 Interference and coherence 159

3.1.3 Coherence length 164
3.2 Interference of Two Plane Waves 166
3.2.1 Interference with inclined plane waves 168
3.2.2 Displacement of interference fringes 170
3.2.3 Interferogram visibility 171
3.3 Interference of Two Spherical Waves 172
3.3.1 Circular fringes with the Michelson interferometer 176
3.3.2 Parallel fringe approximation with the Michelson interferometer 181
3.4 Practical Aspects in the Michelson Interferometer 184
3.4.1 Laboratory interferometer 188
3.5 Interference in a Plate of Parallel Faces 190
3.5.1 Stokes relations 191
3.5.2 Multiple-wave interference 192
3.5.3 Two-wave interference 206
3.6 Interference from *N* Point Sources 209
3.6.1 Plane wave approximation 211
3.7 Interference with Extended Light Sources 214
3.7.1 Artificial extended sources 217
3.8 Young Interferometer I 219
3.8.1 Division of wavefront and division of amplitude 222
3.9 Other Interferometers 223
3.9.1 Fabry–Pérot interferometer 223
3.9.2 Antireflective thin film 225
3.9.3 Newton and Fizeau interferometers 226
References 229

4 Diffraction 231

4.1 Huygens–Fresnel Principle 231
4.1.1 Fresnel zones 234
4.1.2 Fresnel treatment results 237
4.2 Diffraction Integral 240
4.2.1 Kirchhoff integral theorem 242
4.2.2 Fresnel–Kirchhoff diffraction 244
4.2.3 Sommerfeld diffraction 247
4.3 Fresnel and Fraunhofer Diffraction 248
4.3.1 Fraunhofer diffraction 251
4.3.2 Fresnel diffraction 252
4.3.3 Some examples 252
4.4 Young Interferometer II 257
4.4.1 Effect of the size of the diffraction aperture 258
4.4.2 Effect of light source size 258
4.5 Image Formation with Diffraction 262
4.5.1 Image of a point (source) object 262

4.5.2 Resolution in the image (two points) 266
4.5.3 Image of an extended object 268
4.6 Diffraction Gratings 269
References 273

Appendices **275**

A Ray Tracing **275**
References 280

B Refractive Index **281**
References 286

C Optical Glasses **287**
References 289

D Chromatic Aberrations **291**
References 297

E Prisms **299**
References 306

F Polarization Ellipse **307**

Index *309*

Chapter 1
Geometrical Optics

When certain optical phenomena can be explained by geometrical concepts, we are in the realm of what is called geometrical optics. Starting from the idea that light propagates as geometrical rays, and that this propagation is governed by Fermat's principle, we can study the propagation of light in media with a constant or variable refractive index, image formation in instruments comprised of optical elements (e.g., lenses, prisms, and mirrors), optical aberrations that deteriorate such instruments' image quality, the design of optical devices, etc. So, although geometrical optics cannot explain some optical phenomena tied to the wave nature of light, it has a very wide range of applications (Fig. 1.1). Therefore, starting the study of optics from the geometrical point of view is fully justified.

Starting from Fermat's principle, the geometrical properties of optical systems formed by refracting spherical surfaces, spherical mirrors, and lenses

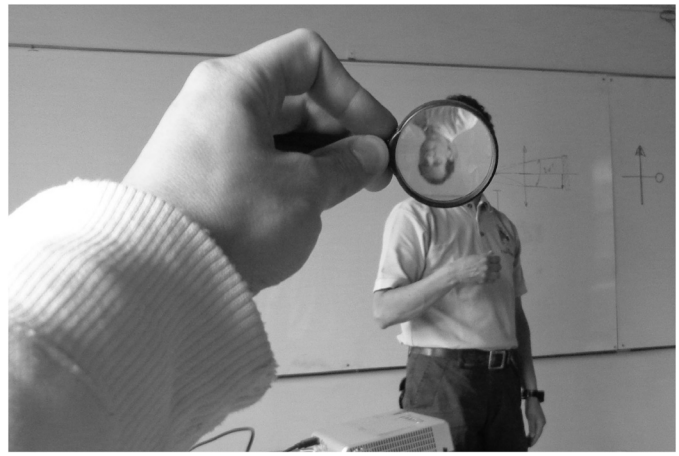


Figure 1.1 Imaging using a positive lens. The image generated by the lens is real and becomes the object for the mobile device's camera that took the photo. We can understand this imaging process from the perspective of geometrical optics. Courtesy of Felix Ernesto Charry Pastrana.

maintained, and the angle with which the light ray is reflected is equal to the angle with which the light ray hits.

On the other hand, in the case of refraction, Newton maintains that the interface that separates two media of different refractive indices (with the refractive index of the incident medium being less than the refractive index of the transmission medium) exerts an attraction on the light particles, thus increasing the normal component of velocity while the tangential component remains unchanged, as shown in Fig. 1.5. This implies that the modulus of the velocity vector increases when light passes from a medium with a lower refractive index to a medium with a higher refractive index, contrary to what actually happens.

1.1.3 Huygens' wave theory

Newton's corpuscular theory is not only wrong about the change in velocity when light is refracted, but it also cannot explain interference phenomena. A different proposition was put forth by Huygens, known as Huygens' principle.

Before stating this principle, let us qualitatively define what a wavefront is by using mechanical waves on the surface of water. When a drop falls on the surface of a water pool that is at rest, we can observe a series of ripples (crests and troughs) on the surface that propagate away from the place where the drop fell. The shape of the crests (and troughs) are circles [Fig 1.6(a)]. Ideally, the points at the peak, or highest point, of one of the crests will be at the same height. We will say that all the points of a peak (and valley) that are at the same height describe a profile that is called a wavefront.

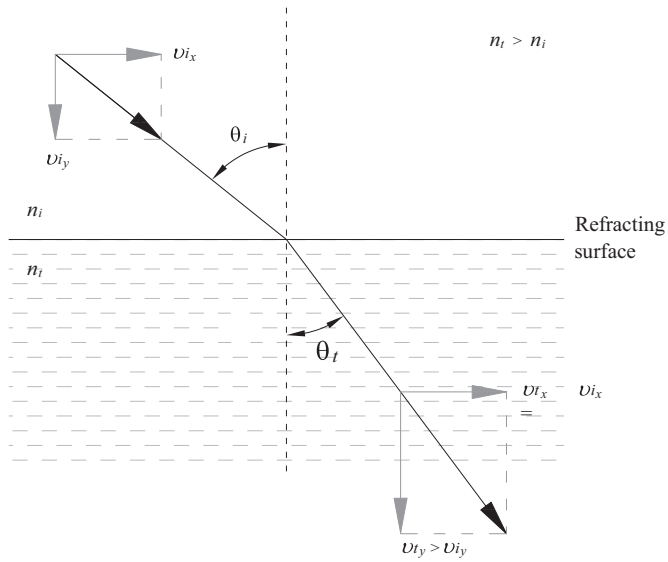


Figure 1.5 Newton's law of refraction.

1.2 Fermat's Principle

In his treatise on geometry, Euclid (325–265 BCE) postulates that "a straight line segment can be drawn joining any two points." In Euclidean space, the shortest distance between two points is the length of the straight line segment that joins them. Following this postulate, Hero of Alexandria (10–70 BCE) establishes that *for light to go from one point to another, it follows the shortest geometrical path*. This is another way of saying that light propagates as geometrical rays. Given that in a homogeneous medium light propagates with constant speed, it can also be said that *for light to go from one point to another, it follows the path for which the least time is used*. For a homogeneous medium, the two statements are equivalent. However, if the two points are in different media (both homogeneous), the result is that light no longer follows the shortest geometrical path.

In Fig. 1.12(a), three possible light paths are illustrated. A planar interface separates the media with refractive indices n and n'. The shortest geometrical path is the straight line PT_2Q, but it does not correspond to the path that light follows if $n' \neq n$. The path followed by light resembles lines PT_1Q when $n' < n$ and the lines PT_3Q when $n' > n$. So how can the real path followed by light be obtained when $n' \neq n$? The answer to this question is obtained by using the statement above referring to the path for which the least time is used. This statement was formulated by Fermat (1601–1665) and is known as *Fermat's principle*.

To apply Fermat's principle to the problem illustrated in Fig. 1.12(a), consider the geometry of Fig. 1.12(b). The interface is located at $y = 0$ and the point T at x in the horizontal axis. The coordinates of point P are $(0, h)$ and those of point Q are $(a, -b)$. The time for light to go from P to Q through T will be

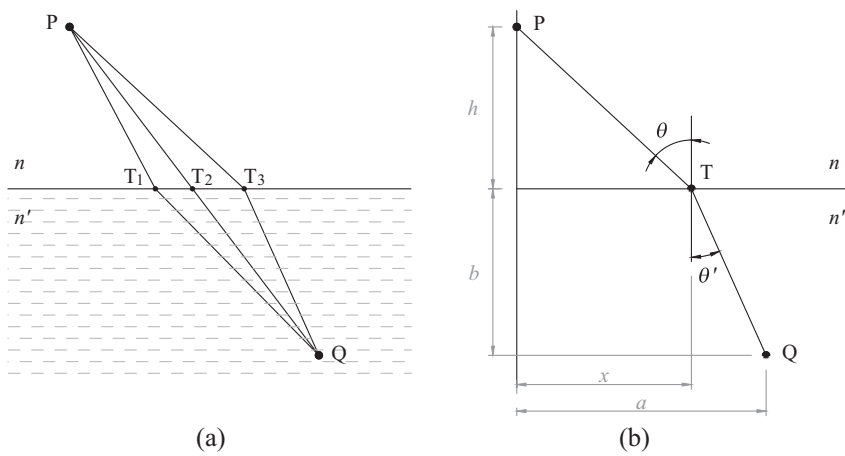


Figure 1.12 Sketch to derive Snell's law from Fermat's principle. (a) Three possible trajectories to go from P to Q. (b) Geometry to calculate the real trajectory.

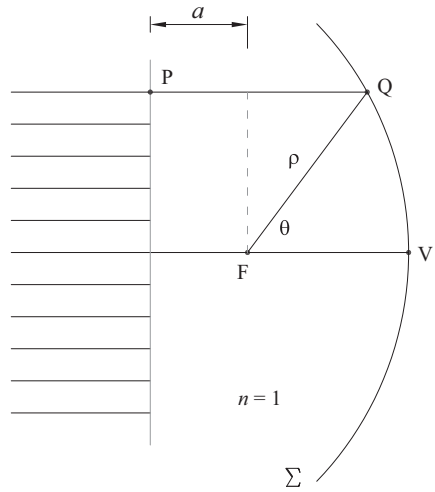


Figure 1.37 A parabolic mirror focuses the parallel rays coming from infinity at the focal point.

In Cartesian coordinates, the profile of the paraboloid can be written as

$$z = \frac{y^2}{2R_v}, \tag{1.36}$$

where z is the coordinate along the optical axis, y is the meridional coordinate, and R_v is the radius of curvature of the parabola at the vertex. Then, the focal point of the mirror is located at the distance $R_v/2$ from the vertex V (Fig. 1.37).

1.5 Lenses: Thin Lens Approximation

If we want to obtain images in a medium other than the spherical refracting surface discussed in Section 1.3.2, another refracting surface should be included. In this section, we will deal with refracting elements limited by two spherical surfaces with a common optical axis. These types of elements are called lenses.

To obtain the position of the image of a point object, we should use Eq. (1.21). The Gaussian equation for surface 1 is

$$\frac{n_1'}{s_1'} - \frac{n_1}{s_1} = \frac{(n_1' - n_1)}{R_1}; \tag{1.37}$$

for surface 2, it is

$$\frac{n_2'}{s_2'} - \frac{n_2}{s_2} = \frac{(n_2' - n_2)}{R_2}. \tag{1.38}$$

1.5.5 Ray tracing for oblique rays

Although they have been mentioned in Sections 1.3.4 and 1.5.4, oblique rays can be specifically defined as those rays that leave the tip of an object and do not travel in a direction parallel to the optical axis. In particular, we are going to deal with oblique rays that are kept in the meridional plane. We have already dealt with some of them in Section 1.5.4, e.g., the ray passing through the center of a lens and the ray directed toward a primary focal point. In this section, we are interested in oblique rays that are directed in any other direction, which can occur in ray tracings that involve a combination of two or more lenses.

For example, in Fig. 1.46, a ray refracted by the lens L_1 is directed to the secondary focal point of L_1. As it hits the lens L_2, how is the ray refracted? With the ray tracing technique illustrated in Fig. 1.40, we do not have a solution. However, the ray tracing shown in Fig. 1.45 gives us a hint as to how to graphically determine the refraction generated by L_2. If we assume that the oblique ray reaching L_2 is part of a bundle of parallel rays, as illustrated in Fig. 1.47(a), such a ray will be refracted (according to Fig. 1.45), diverging

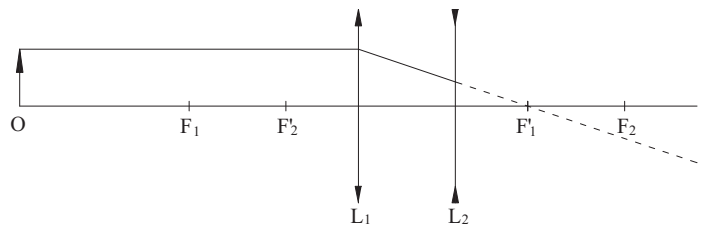


Figure 1.46 Ray tracing for a ray parallel to the optical axis coming out of the object tip. The ray refracted by lens 1 becomes an oblique ray for lens 2.

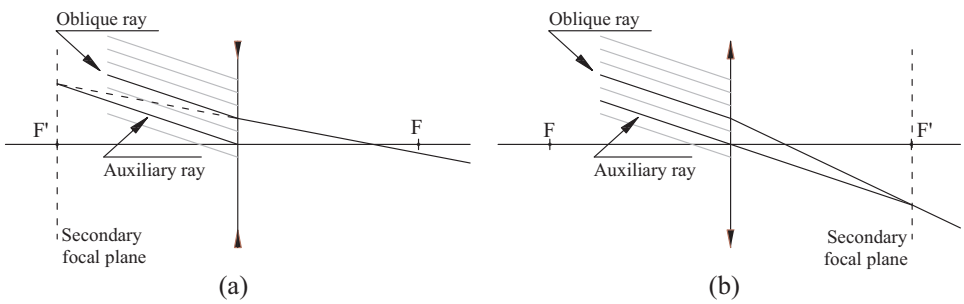


Figure 1.47 Oblique rays incident on thin lenses. (a) A negative lens and (b) a positive lens. To determine the refraction of an oblique ray hitting a lens, an auxiliary ray passing through the center of the lens and parallel to the oblique ray is drawn. The oblique ray is refracted traveling toward, or coming from, the point where the auxiliary ray intersects the secondary focal plane of the lens.

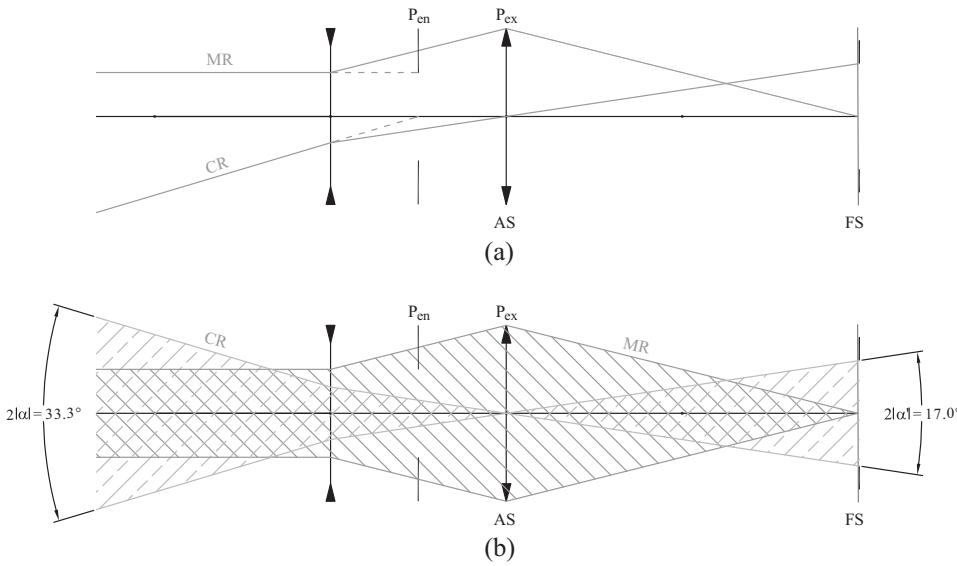


Figure 1.67 Energy and field of view cones for the two-lens optical system from Fig. 1.65. (a) Chief and marginal ray tracing. (b) Whereas the surface boundary of the energy cone is defined by the marginal ray, the surface of the field of view cone is determined by the chief ray.

center of the AS. By resizing the FS, the image extension would be modified, but there would be no effect on the amount of energy reaching each point in the image because the marginal ray passes through the center of the FS.

1.8 Some Optical Instruments

In this section, we will look at the basic setup of some optical instruments and describe their operation in terms of topics covered in the previous sections. First, we will view the human eye as a system of spherical surfaces with symmetry of revolution and simplify it to a thin lens and image plane. Next, we will see the magnifying glass coupled with the thin lens model of the eye. Finally, we will study the telescope and the microscope as extensions of the magnifying glass, i.e., when another lens is added to form the image of very distant objects or very close but small objects.

1.8.1 The human eye (schematic representation)

In Section 1.3.1, an overview of the eye as an optical system with various refracting surfaces is given. Due to the existing variation of the characteristics of the eye from one person to another, to study the formation of an image in the eye, models with ocular parameters that represent the mean values of the population are used. This can be done with varying degrees of precision. If the refracting surfaces of the eye are assumed to be spherical and centered on

the same axis (optical axis), a family of models is obtained, the so-called schematic eyes [9]. In particular, they are used in the paraxial region to obtain information such as refractive power, magnification, illumination on the retina, Purkinje images (reflections on the refracting surfaces of the eye), location of the pupils, location of the focal points (F, F′), location of the principal points (P, P′) and the nodal points (N, N′), and the effects of refractive errors (myopia and hyperopia). When image information is required beyond the paraxial region, more refined models called wide-field schematic eyes are used, including nonspherical refracting surfaces, off-center refracting surfaces, and *Gradient Index* (Grin) models for the refractive index of the crystalline lens [9].

In this section, we will consider the Gullstrand–Emsley schematic eye in the relaxed and fully accommodated eye configurations, assuming the eye has no refractive errors (emmetropic eye). A relaxed eye is understood as the configuration that the eye has when observing an object located at infinity. In this situation, the crystalline lens has the longest focal length it can have. As the object approaches the eye, the crystalline lens changes its geometry to decrease the focal length so that the image remains focused on the retina. This process is called *accommodation*. Accommodation has a limit; thus, there is a distance limit from which, for distances less than this limit, the image on the retina can no longer remain in focus. The distance where the accommodation limit is located is called the *near point distance* (*npd*). The farthest distance, where the eye can see without accommodation, is called the *far point distance* (*fpd*). In an emmetropic eye, we have $fpd = \infty$. The range in diopters between *npd* and *fpd* is called the *accommodation range*. The *npd* varies with age and is lower in children. For example, a 10-year-old child has an *npd* of $\simeq 70$ mm and his accommodation range will be $[1000/(70\text{ mm}) - 1000/\infty] = 14.29$ D, whereas a 50-year-old adult has an *npd* of $\simeq 400$ mm and her accommodation range will be $[1000/(400\text{ mm}) - 1000/\infty] = 2.5$ D.

In Fig. 1.68, the Gullstrand–Emsley schematic eye is shown in the relaxed eye configuration [9]. The parameters R and d are expressed in millimeters, and the optical power is expressed in diopters. The optical system is represented by four surfaces: (1) cornea, (2) iris (AS) and anterior surface of the lens, (3) posterior surface of the lens, and (4) retina. The distances of the cardinal points (focal points F-F′, principal points P-P′, and nodal points N-N′) with respect to surface 1 are: $VF = -14.983$, $VF' = 23.896$, $VP = 1.550$, $VP' = 1.851$, $VN = 7.062$, and $VN' = 7.363$. The secondary focal point F′ coincides with the retina R′. The focal lengths are $f = 16.53$ and $f' = 22.05$. On the other hand, with respect to surface 1, the pupils are located at $VP_{en} = 3.052$ and $VP_{ex} = 3.687$. When examining the refractive power of each element $[P = (n' - n)/R]$, the cornea contributes approximately two-thirds of the total refraction of the eye. All distances are in millimeters.

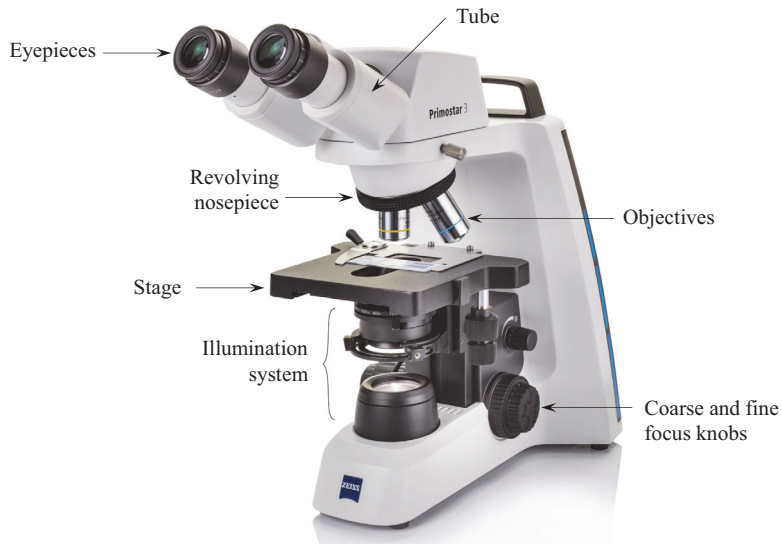


Figure 1.85 Components of a ZEISS Primostar transmission microscope. Image reprinted with permission from ©ZEISS Microscopy (https://www.zeiss.de/).

find the eyepiece(s), in which the eye(s) can be placed to observe the final image. The object can also be illuminated from above (for example, with a concave mirror), in which case what will be seen is the light reflected by the object (reflection microscope).

1.9 Monochromatic Optical Aberrations

In Section 1.3, the Cartesian oval is defined as the surface of revolution that forms a point image of a point object, both on the optical axis. The shape of this surface is derived from Fermat's principle and depends on the object distance and image distance, with respect to the surface vertex, and on the relation between the media (surface and surroundings) refractive indices. Therefore, for different object distances, we will have different Cartesian ovals. This imposes a practical limitation when designing optical imaging systems; the image would be well in focus for only one object position. A practical solution is to use reflecting and/or refracting spherical surfaces and limit the optical region through which the energy flow will pass to a very small region around the optical axis, which can be done with a diaphragm (AS) of small diameter. This condition is known as the paraxial approximation and gives rise to the Gaussian equation [Eq. (1.21)], which implies that the image of a point object will be a point image, so the image of an extended object will be a copy of the object except for a scale and orientation factor.

When the optical system is not limited to the paraxial region, the image of an extended object will no longer be a faithful replica of the object. The

between the lenses seeking the greatest possible symmetry to have the least distortion.

1.9.4 Astigmatism and coma

To see these two aberrations, we will first define the tangential plane and the sagittal plane. In Section 1.3.5, the meridional plane is defined as the plane containing the optical axis. In that section, the extended objects are represented as arrows of a certain height contained in the meridional plane, and the rays emerging from any point on the object are also limited to the meridional plane. In an optical system, the plane that contains the optical axis and the chief ray is called the *tangential plane*, and the plane that contains the chief ray and is orthogonal to the tangential plane is called the *sagittal plane*. The tangential plane is a meridional plane, but the sagittal plane is not. The tangential plane maintains its spatial orientation throughout the optical system, whereas the sagittal plane changes its inclination in the same way as the chief ray. In Fig. 1.92, these planes are illustrated in a system composed of an AS and a convex refracting surface. The height of the object is $h = -\overline{\mathrm{OP}}$, and the height of the image is $h' = \overline{\mathrm{O'P'}}$. In Fig. 1.92, the rays that leave P and are in the tangential plane are called tangential rays (blue); the rays leaving P and contained in the sagittal plane are called sagittal rays (red). The projection of the AS on the refracting surface is shown in the figure and four points are marked: T_1 and T_2 correspond to the intersections of the tangential marginal rays on the refracting surface, and S_1 and S_2 correspond to the intersections of the marginal sagittal rays on the refracting surface.

The meridional plane containing the marginal tangential rays passing through T_1 and T_2 is shown in Fig. 1.93. Sagittal points S_1 and S_2 project into this meridional plane at a single point. Given the effect of spherical aberration

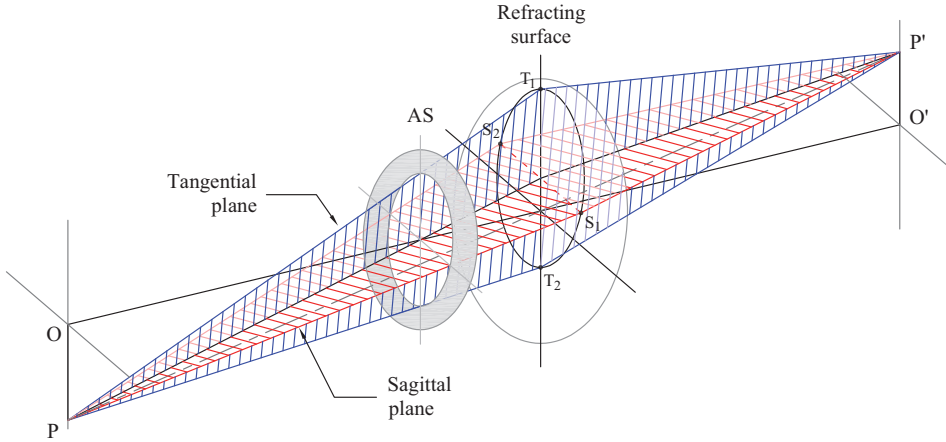
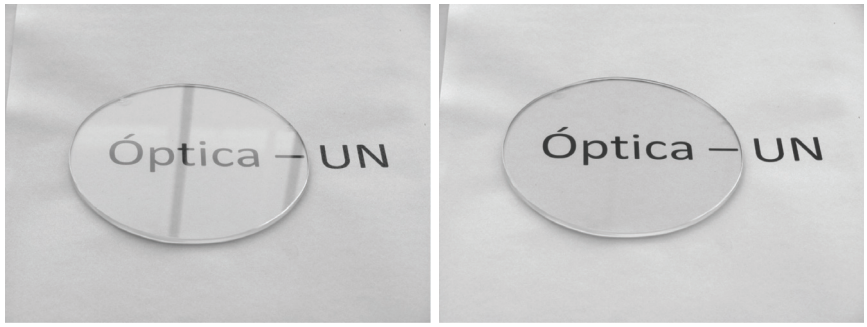


Figure 1.92 Defining the tangential (blue) and sagittal (red) planes.

Chapter 2
Polarization

According to classical physics, light is an electromagnetic wave and its properties are obtained from Maxwell's equations. One of these properties is that light is a transverse wave; i.e., the electric and magnetic vectors (optical field) vibrate orthogonally to the direction of wave propagation. If we assume that the light source is composed of oscillators that emit electromagnetic energy, then in general, the directions of the electric and magnetic vectors are random. However, it is possible to maintain the vibration of the resulting electric (magnetic) vector in a fixed plane or following an elliptical or circular curve. In such a case, the wave is said to be polarized. This chapter defines polarization and shows some of its applications (Fig. 2.1).

Taking into account the linearity of Maxwell's equations, one can limit the study of polarization to plane harmonic waves. Although the emitted or reflected optical field can have any form, Fourier analysis shows that the complex form of the optical field wavefront can be expressed by the sum of



(a) Image with nonpolarized light. (b) Image with polarized light.

Figure 2.1 Polarization by reflection. The light that enters through a window in a room is not polarized. When reflecting off a glass plate (smooth surface), as in (a), part of the window is visible along with the text below the glass plate. If the reflection is viewed at an angle close to Brewster's angle, the light will be linearly polarized, which is verified by placing a linear polarizer between the glass plate and the photographic camera taking the image. This eliminates reflected light, and the text below the page is seen clearly, as shown in (b).

harmonic plane waves. Thus, the results for plane waves can be extended to more complex forms of the optical field.

This chapter begins by developing the algebra to describe linear, elliptical, and circular polarization. Among the polarization mechanisms, dichroism, polarization by total internal and external reflection, and birefringence are discussed in detail, with the latter limited to the case in which the principal directions of the refractive indices coincide with the axes of the crystal glass. The refractive media considered here are dielectrics without absorption. Finally, the Jones formalism to describe polarization states and polarizing elements is presented.

2.1 Plane Waves and Polarized Light

In a vacuum, for a vector point $\mathbf{r} = (x, y, z)$ and time t, the optical field is described by the electric vector $\mathbf{E}$ and the magnetic vector $\mathbf{H}$, which are related to each other according to Maxwell's equations, given by

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \tag{2.1}$$

$$\nabla \times \mathbf{H} = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \tag{2.2}$$

$$\nabla \cdot \mathbf{E} = 0, \tag{2.3}$$

$$\nabla \cdot \mathbf{H} = 0. \tag{2.4}$$

From Eqs. (2.1), (2.2), and (2.3), the wave equation for the electric field is*

$$\nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}, \tag{2.5}$$

with $c^2 = 1/\mu_0 \epsilon_0$. For the magnetic field, an equation analogous to Eq. (2.5) is obtained.

Because $\mathbf{E}(x, y, z) = \{E_x(x, y, z), E_y(x, y, z), E_z(x, y, z)\}$ for a time t, Eq. (2.5) represents a set of three equations, one for each component of the electric field $\mathbf{E}$. If any of these components is represented by $V = V(x, y, z)$, we have a scalar equation of the form

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 V}{\partial t^2}. \tag{2.6}$$

Let $\hat{\mathbf{s}} = (s_x, s_y, s_z)$ be a unit vector in a fixed direction in space. A solution of Eq. (2.6) of the form $V(\mathbf{r}, t) = g(\mathbf{r} \cdot \hat{\mathbf{s}}, t)$ represents a homogeneous plane

*The wave equation is obtained using the identity vector $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2(\mathbf{E})$.

On the other hand, $\alpha = 0$ defines a horizontal linear polarization state ($\mathbf{E}_H$) and $\alpha = 90°$ defines a vertical linear polarization state ($\mathbf{E}_V$).

2.2 Dichroism Polarization

One way to remove one of the **E**-field components is by absorbing that component. This can be achieved by designing a device that performs this task or by using a natural material that has this property [1]. In both cases, the selective absorption of one of the **E**-field components is called *dichroism*. The final effect on the field will be linearly polarized light.

2.2.1 Linear polarizer

To see how a dichroism-based linear polarizer works, suppose that a grid of parallel conducting wires is constructed, as shown in Fig. 2.8, and an unpolarized field **E** (natural light) is incident in a direction orthogonal to the grid plane. Because the electric charges have the possibility of greater displacement in the horizontal direction (along the wires) compared with the vertical direction (cross section of the wires), there will be a greater absorption of electric energy in the direction of the wires; thus, the net component E_x experiences a greater attenuation than the net component E_y. If, ideally, the component E_x is completely attenuated, we will have a linear polarizer and the field will have a vertical linear polarization state, $\mathbf{E}_V$. The direction in which the field is not attenuated is called the *transmission axis of the linear polarizer*.

Using lithographic methods, polarizers based on a grid of conductive wires for the visible spectrum are manufactured, achieving arrangements with a separation of 100 nm between wires. Aluminum microwires are deposited on glass substrates.

The most common dichroic linear polarizers are made of sheets of a special transparent plastic (polyvinyl alcohol). These sheets have been stretched in one direction to align their long molecules, which are then

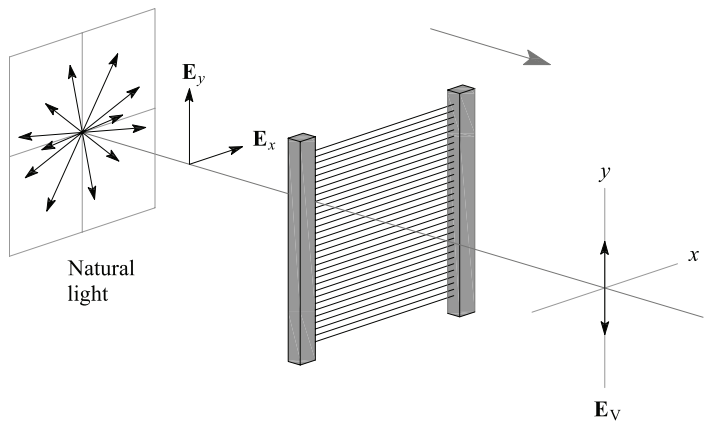


Figure 2.8 Linear polarizer made with a grid of conducting wires.

grazing incidence (at the interface). This even happens with flat opaque surfaces; e.g., if we look at a sheet of paper from a grazing angle, we can see the specular reflection of light very well.

2.4 Polarization by Total Internal Reflection

In the previous section, we show how the orthogonal and parallel components of the reflected field can undergo a phase change with respect to the components of the incident field. The parallel component is reflected in phase, when $0 < \theta_i < \theta_p$, and with a phase shift of $\pm\pi$, when $\theta_p < \theta_i < \pi/2$. On the other hand, the orthogonal component is always reflected with a phase shift of $\pm\pi$. This, together with the change in amplitude of the reflected components, implies that a linearly polarized wave when reflected remains linearly polarized but with a rotation in the plane of vibration. When total internal reflection occurs, the components of the reflected field have phase shifts that vary between 0 and $\pm\pi$, and the phase difference between the components is no longer limited to 0 or $\pm\pi$. Therefore, the reflected wave can have an elliptical polarization state.

2.4.1 Total internal reflection

If the incident wave goes from a medium with a higher refractive index to one with a lower refractive index, from a certain angle, called the *critical angle*, the reflection and transmission coefficients obtain the values $|r_{\parallel}| = |r_{\perp}| = 1$ and $t_{\parallel} = t_{\perp} = 0$. In other words, the energy of the reflected wave is equal to the energy of the incident wave. In Section 1.1, the condition in which the transmission ray is tangential to the interface is illustrated in Fig. 1.11. The angle of incidence for which this occurs is [Eq. (1.5)]

$$\theta_c = \arcsin\left(\frac{n_t}{n_i}\right), \tag{2.114}$$

with $n_i > n_t$. From this angle the phenomenon of *total internal reflection* occurs.

The Fresnel equations for $n_i > n_t$ and $\theta_i < \theta_c$ apply in the same way as in the external reflection case ($n_i > n_t$), and the only phase changes of the reflected components with respect to the incident components are 0 or $\pm\pi$, as shown in Fig. 2.17 for $n_i = 1.5$ and $n_t = 1.0$. Unlike external reflection (Fig. 2.13), the orthogonal component does not undergo a phase change. In contrast, the parallel component has a phase shift of $\pm\pi$ for $0 < \theta_i < \theta'_p$ and is in phase for $\theta'_p < \theta_i < \theta_c$. The angle θ'_p is the angle at which the polarization of the reflection occurs and is given by $\tan\theta'_p = (n_t/n_i)$. This angle together with the external polarization angle satisfies the relation

$$\theta_p + \theta'_p = \pi/2. \tag{2.115}$$

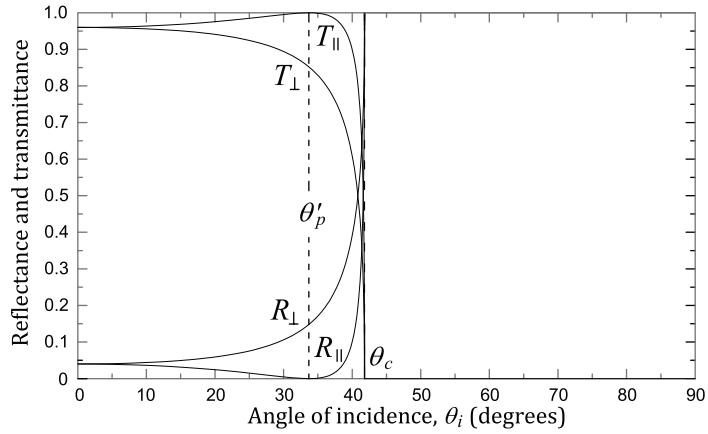


Figure 2.21 Reflectance and transmittance for internal reflection ($0 < \theta_i < \theta_c$) and total internal reflection ($\theta_c < \theta_i < \pi/2$).

we also have two intervals. In the first, $0 < \theta_i < \theta_c$, the transmission coefficients are calculated according to Eqs. (2.98) and (2.99). In the second interval, $\theta_c < \theta_i < \pi/2$, the transmission angle is $\theta_t = 90°$; therefore, according to Eq. (2.110), the transmittance for both the parallel and the orthogonal components becomes equal to 0.* Of course, it is again verified that $R + T = 1$ for the two components of the electric (and magnetic) fields.

2.5 Polarization with Birefringent Materials

The electric polarization vector in dielectric materials is related to the electric field by the electric susceptibility according to Eq. (B.9):

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}.$$

When the material is isotropic, the susceptibility quantity is described by a scalar and the wave equation [Eq. (B.14)] is reduced to $\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 (1 + \chi) \times \partial^2 \mathbf{E}/\partial t^2$, where $\upsilon = c/\sqrt{1 + \chi}$ is the speed of light in the material. When the material is anisotropic, the susceptibility is described by a tensor (3×3 matrix) and the wave equation is a bit more complex.

In general, electric susceptibility can be described as

$$\chi = \begin{pmatrix} \chi_{11} & \chi_{12} & \chi_{13} \\ \chi_{21} & \chi_{22} & \chi_{23} \\ \chi_{31} & \chi_{32} & \chi_{33} \end{pmatrix}, \tag{2.130}$$

*In total internal reflection, the incident energy is completely reflected. However, there is still an electromagnetic wave beyond the interface that is rapidly fading. This wave is known as an evanescent wave.

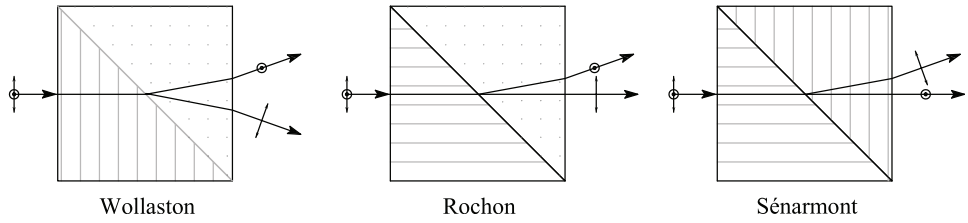


Figure 2.32 Some types of polarizing prisms that separate the components of the optical field into the "s" and "p" states. In all cases, the prisms are quartz [2].

is refracted approaching the normal. On the other hand, $\mathbf{E}_i^{\parallel}$ is the extraordinary ray in the first prism; it changes to an ordinary ray when passing to the second prism, so that on the diagonal it refracts away from the normal. In Rochon prisms, $\mathbf{E}_i^{\perp}$ and $\mathbf{E}_i^{\parallel}$ enter the first prism along the optic axis, so both rays will see the ordinary index. In the second prism, $\mathbf{E}_i^{\perp}$ changes to an extraordinary ray and refracts closer to the normal. In contrast, $\mathbf{E}_i^{\parallel}$ continues as an ordinary ray, so its propagation direction does not change. Finally, in Sénarmont prisms, $\mathbf{E}_i^{\perp}$ and $\mathbf{E}_i^{\parallel}$ also enter the first prism along the optic axis, so both rays will see the ordinary index. In the second prism, $\mathbf{E}_i^{\perp}$ continues as an ordinary ray, so its direction of propagation does not change, and $\mathbf{E}_i^{\parallel}$ changes to an extraordinary ray and refracts closer than normal.

There are also polarizing prisms made of isotropic optical glass. These are beamsplitter cubes with a dielectric film between the diagonals that joins the prisms, allowing the transmission of the p-polarization state and the reflection in the diagonal of the s-polarization state. These are mentioned briefly in Appendix E.

2.6 Vectors and Jones Matrices

To describe the polarization state of a plane wave in Section 2.1, we use the vector [Eq. (2.32)]

$$\mathbf{E}(x, y, z; t) = \{\hat{i}E_{ox} + \hat{j}E_{oy}\}e^{i(kz-\omega t)},$$

where $E_{ox} = |E_{ox}|e^{i\delta_x}$ and $E_{oy} = |E_{oy}|e^{i\delta_y}$. Because the temporal and spatial phase terms are common to the complex amplitudes of the two wave components, it is convenient to represent the state of polarization as a column vector in which its elements determine the relationship between the components of the wave. This representation is known as a *Jones vector*:

$$\begin{bmatrix} E_{ox} \\ E_{oy} \end{bmatrix} = \begin{bmatrix} |E_{ox}|e^{i\delta_x} \\ |E_{oy}|e^{i\delta_y} \end{bmatrix}. \tag{2.154}$$

Chapter 3
Interference

Light wave interference is observed as a modulation of irradiance, usually bright fringes and dark fringes on an observation screen. The geometry of the fringes depends on the shape of the wavefronts and the difference in the optical path traveled by the waves. Differences in the order of the wavelength of light cause changes in irradiance from a bright fringe to a dark fringe, making interference a highly accurate tool for measuring refractive indices, wavefronts, forms of optical surfaces, thicknesses, etc. The physical parameter that determines the quality of the interference (the possibility of generating fringes) is the coherence between the waves. The coherence has its origin in the fluctuations of the optical field emitted by the sources. Natural sources, like the sun, emit spontaneously (randomly), but in artificial sources, like lasers, the emission has a high degree of correlation.

Two interference patterns generated with a He-Ne laser are shown in Fig. 3.1. The laser beam is focused with a positive lens into a small hole in an opaque screen, which is seen as a point of light in the figure (point source). The lens is behind the screen and cannot be seen in the image. The divergent (spherical) wavefront passes through several optical elements. First, it passes through a 1 mm thick microscope slide (flat piece of glass). There, light is reflected from each slide face and interference occurs between the two reflected signals, which is seen on the opaque screen (two-source interference pattern at bottom left). This interference pattern consists of roughly circular fringes, where the thickness of the bright fringes is similar to the thickness of the dark fringes. This is the typical result of the interference of two sources that emit spherical waves. The beam transmitted by the microscope slide is then allowed to enter a Fabry–Pérot interferometer, which consists of two thick plates of highly reflective glass parallel to each other. The separation between the plates is less than a millimeter, and the facing faces have a thin aluminum film that increases reflectance. This generates multiple reflections, with similar amplitude coefficients, so there is now interference from more than two waves. The effect on the reflected interference pattern is a thinning of

3.2.3 Interferogram visibility

The modulation contrast of irradiance in the axial direction was defined by Eq. (3.32). This quantity, based on the irradiances, allows the degree of coherence to be measured. The visibility of the interference fringe pattern can also be calculated from Eq. (3.32), where $I_{\max}$ and $I_{\min}$ are the maximum and minimum irradiance values of the fringe pattern.

The change in visibility depends on the degree of coherence and the relationship between the intensities of the two waves; e.g., if $|\gamma| = 1$, then the visibility

$$C = \frac{2\sqrt{I_1 I_2}}{(I_1 + I_2)} \tag{3.39}$$

only depends on the ratio between I_1 and I_2. In Fig. 3.13, three interferograms are shown along with their profiles in the x direction when the wave amplitudes are $E_1 = E_0$ and $E_2 = E_0$, $E_1 = E_0$ and $E_2 = 0.4E_0$, and $E_1 = E_0$ and $E_2 = 0.1E_0$. In the first case, $I_{\max} = 4I_0$, $I_{\min} = 0$, and $C = 1$; in the second case, $I_{\max} = 1.96I_0$, $I_{\min} = 0.36I_0$, and $C = 0.69$; and in the third case, $I_{\max} = 1.21I_0$, $I_{\min} = 0.81I_0$, and $C = 0.20$. Note that the irradiance oscillates spatially around the mean value $I_1 + I_2$, which in the first case is $2I_0$, in the second case is $1.16I_0$, and in the third case is $1.01I_0$.

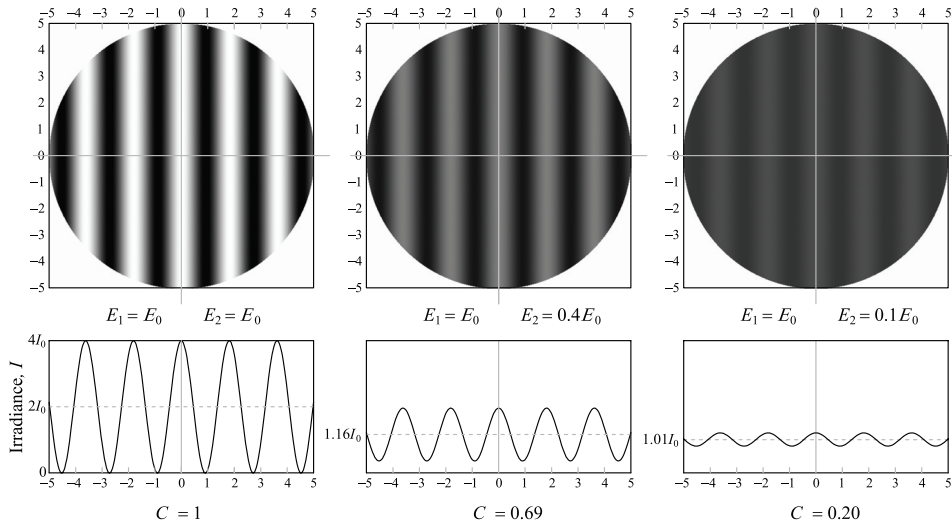


Figure 3.13 Interferograms when the visibility of the fringes depends on the ratio of the wave amplitudes: $C = 1$ ($E_1 = E_0$ and $E_2 = E_0$), $C = 0.69$ ($E_1 = E_0$ and $E_2 = 0.4E_0$), and $C = 0.20$ ($E_1 = E_0$ and $E_2 = 0.1E_0$). The modulus of the degree of coherence is set to $|\gamma| = 1$. The scale of the axes is in millimeters.

$$\Delta x' = \frac{z_0 \lambda}{a}. \tag{3.115}$$

These results [Eqs. (3.114) and (3.115)] were also obtained in Section 3.6 [Eqs. (3.109) and (3.110)], in which spherical wavefronts were approximated by planar wavefronts, an approximation also found in Young's experiment. With this in mind, the irradiance on the observation screen of Young's experiment is given by Eq. (3.104) with $N = 2$, i.e.,

$$I = I_0 \left[\frac{\sin(kx'a/z_0)}{\sin(kx'a/2z_0)} \right]^2, \tag{3.116}$$

with $I_0 = (\epsilon_0 c/2) E_0^{\dagger 2}/z_0^2$, where $E_0^{\dagger}$ is the field amplitude (per unit length) of sources S_1 and S_2. Using the trigonometric identity $\sin(2\alpha) = 2 \sin \alpha \cos \alpha$ in the numerator of Eq. (3.116) leads to

$$I = 4I_0 \left[\cos\left(\frac{\pi a}{z_0 \lambda} x' \right) \right]^2. \tag{3.117}$$

Another approximate way to determine the irradiance is obtained directly from Fig. 3.57(c). Taking into account the condition $z_0 \gg a$, the field at P due to S_1 is $E_1(P) = (E_0^{\dagger}/z_0) e^{iks_1}$ and the field at P due to S_2 is $E_2(P) = (E_0^{\dagger}/z_0) e^{iks_2}$. Therefore, the irradiance at P, given by $I = (\epsilon_0 c/2)|E_1(P) + E_2(P)|^2$, would be

$$I = 2I_0[1 + \cos(k(s_2 - s_1))]. \tag{3.118}$$

Approximating $s_2 - s_1 \approx a\alpha \approx ax'/z_0$ leads to

$$I = 2I_0 \left[1 + \cos\left(\frac{2\pi a}{\lambda z_0} x' \right) \right], \tag{3.119}$$

which is equivalent to Eq. (3.117).

3.8.1 Division of wavefront and division of amplitude

In Young's interferometer there is a remarkable fact as to how to generate the two secondary sources S_1 and S_2 compared with the Michelson interferometer. In the Michelson interferometer (as in the plate with parallel faces), the secondary sources are virtual images of a primary source obtained by the beamsplitter and the mirrors M_1 and M_2 (Fig. 3.18). The beamsplitter divides the amplitude of the incident wave (into a reflected wave and a transmitted wave). Interferometers based on this principle are also called *amplitude-splitting interferometers*. In contrast, in Young's interferometer, the secondary sources S_1 and S_2 are obtained by isolating regions of the wavefront emitted

Chapter 4
Diffraction

Diffraction, like interference, is a wave phenomenon. From a mathematical point of view, the difference between interference and diffraction lies in the number of sources that generate the interference waves. In interference there is a discrete number of sources, whereas in diffraction there is a continuous number of sources. In terms of the behavior of the optical field, diffraction is considered the deviation of the rectilinear path (of light) that is not due to reflection or refraction.

In this chapter, diffraction is limited to the paraxial range, i.e., Fresnel diffraction and Fraunhofer diffraction. Detailed examples of diffraction by a circular aperture and by a rectangular aperture are given. With diffraction through a circular aperture, the formation of the image is analyzed taking into account the wave nature of light; with diffraction through a rectangular aperture, the basic mathematics for one-dimensional diffraction gratings are developed.

The image of a point object (monochromatic) generated by an optical system that models a human eye with myopia, astigmatism, and spherical aberration is shown in Fig. 4.1. The effect of diffraction and aberrations reduces visual acuity in the human eye and generally reduces resolution in imaging systems.

Note on calculated diffraction patterns
Except for Section 4.5.2, which deals with image resolution, calculated diffraction patterns are shown in this chapter as grayscale images that represent the square root of the irradiance distribution. This allows regions of lower intensity to be highlighted. Plots of the irradiance profiles are shown at scale.

4.1 Huygens–Fresnel Principle

Huygens' principle, discussed in Section 1.1.3, states that *every point on a wavefront can be considered as a source of secondary spherical waves that propagate with the same speed as the wavefront. After a while, the propagated*

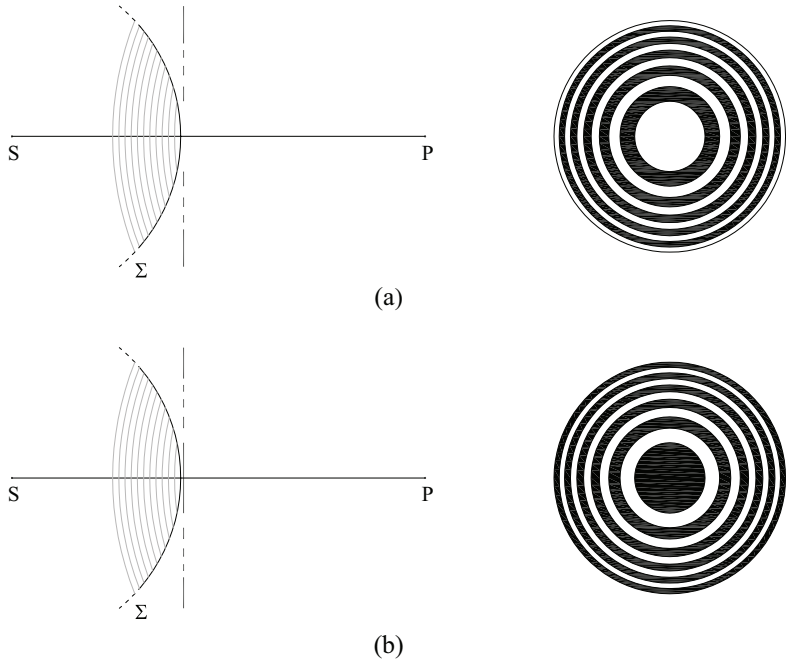


Figure 4.7 Fresnel zonal plates to block (a) even and (b) odd zones.

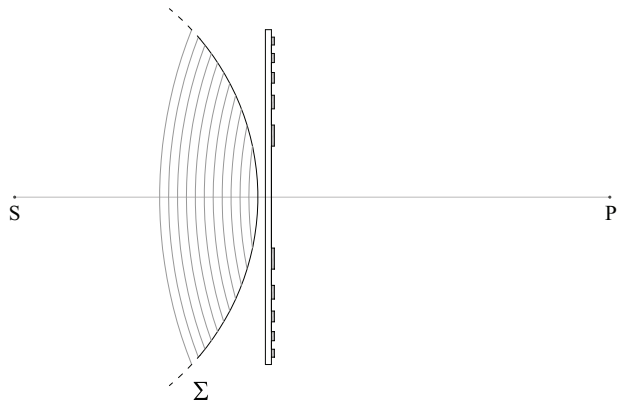


Figure 4.8 The phase zone plate takes advantage of the entire optical field Σ and constructively interferes with the fields contained in the odd and even Fresnel zones.

4.2 Diffraction Integral

Diffraction involves finding the optical field at any point in space generated by a source with boundary conditions. The typical geometry in diffraction is illustrated in Fig. 4.9. In region I, the source S (point-like or extended) is located; in region II, the volume is limited by the closed surface $\Sigma = \Sigma_1 + \Sigma_2$, in which the optical field is measured. Region II is called the *diffraction region*.

Gaussian profile in millimeters and is taken as the radius of the coherence region. In the experiment, the opening gap varies from 2 to 12 mm in steps of 2 mm. By examining the profile of the interferograms in the horizontal direction (x'), curves similar to those shown in Fig. 4.25 are obtained. When $a = 12$ mm, there is no more interference. Note that in these images no interference rings are observed [as shown in the simulated pattern in Fig. 4.23(a)]. This occurs because in practice, the maximum irradiance of the first ring is very small with respect to the maximum of the central region, as can be seen in the profile of Fig. 4.23(b).

From the lessons learned in this section, one can imagine how careful Thomas Young was to look at the interference fringes, which must be colored if the primary source is the sun.

4.5 Image Formation with Diffraction

According to geometrical optics, the image of a point formed by an optical system free of optical aberrations is also a point. Suppose the point object is located on the optical axis. The spherical wavefront that diverges from the object when passing through the optical system will be truncated by the aperture diaphragm; i.e., the diaphragm plays the role of the aperture that diffracts the light. This implies that the image cannot be a point. On the other hand, the image of a large object will depend on the spatial coherence of the optical field in the object. This section briefly deals with the topic of imaging by taking diffraction into account in the paraxial approximation (Fresnel/Fraunhofer diffraction).*

4.5.1 Image of a point (source) object

Let us consider the system shown in Fig. 4.27. The thin lens represents the imaging optics, and the edge of the lens is the aperture diaphragm. The lens introduces a phase delay in the wavefront as it passes through the diaphragm. With this in mind, the lens can be modeled as a complex variable transmittance that changes the phase of the incident wavefront at the diaphragm. Thus, the process of image formation of a point object can be described as follows: a spherical wavefront that diverges from the point object is truncated by the aperture diaphragm and undergoes a phase shift due to the transmittance of the lens, then converges as a diffraction pattern in the Gaussian image plane.

In Fig. 4.27, $-s_o$ and s_i are the object and image distances from the thin lens in the plane of the aperture (diaphragm). The phase of the optical field (diverging from the point object) just before the aperture would be

*For a detailed discussion of the problem of imaging, by taking diffraction into account, see Goodman [6].

Yobani Mejía-Barbosa is a Professor in the Department of Physics at the Universidad Nacional de Colombia, where he has taught courses in optics for more than 12 years. He received his B.S. and M.S. degrees in Physics from the Universidad Nacional de Colombia in 1991 and 1995, respectively, and his Ph.D. in Optics from the Centro de Investigaciones en Óptica, Mexico, in 2001. His current research interests include optical design, interferometry, visual optics, and classical coherence. He is a senior member of Optica.

Herminso Villarraga-Gómez obtained his B.S. in Physics in 2007 from the Universidad Nacional de Colombia, under the guidance of Prof. Mejía-Barbosa. He also holds an M.S. in Physics (University of Puerto Rico, 2010), an M.S. in Optics (University of Central Florida, 2012), and a Ph.D. in Optical Science and Engineering (University of North Carolina, 2018). Herminso has worked for world-renowned optical companies, including Nikon (2015–2019) and most recently ZEISS (since 2019). He has been a member of SPIE and Optica since 2012.