SCIENCE FOR EVERYONE

L.V. TARASOV
A.N. TARASOVA

DISCUSSIONS ON REFRACTION OF LIGHT

MIR

Л. В. Тарасов
А. Н. Тарасова

Беседы о преломлении света

Под редакцией В. А. Фабриканта

Издательство «Наука». Москва

L. V. Tarasov
A. N. Tarasova

Discussions on refraction of light

Edited by V. A. Fabrikant

Translated from the Russian
by Vladimir Kolesnikov
and Nicolas Kulanin

Mir
Publishers
Moscow

Л. В
А. Н

Бес

Под

Изда

First published 1984

На английском языке

Preface

Why does a beam of light change its direction when passing through the interface between two media? Why does the setting sun appear oblate on the horizon? What causes mirages? Why does a prism disperse sunlight into different colours? How can one calculate the angular dimensions of a rainbow? Why do distant objects appear close when we view them through a telescope? What is the structure of the human eye? Why does a light ray get broken into two in a crystal? Can the plane of the polarization of a ray be turned? Can light rays be bent at will? Is the refractive index controllable?

This book will give the reader answers to all these questions. He will get to know how the law of refraction was discovered, how Newton's theory of the refraction of light in the atmosphere was nearly lost forever, how Newton's experiments changed radically the old ideas concerning the origin of colours, how the telescope was invented, how it took twenty centuries to understand the anatomy of human vision, and how difficult it was to discover the polarization of light.

To make the historical and the physical aspects of the book more convincing, the authors have introduced a number of problems and their detailed solutions, geometrical constructions, and optical diagrams of some instruments and devices. No doubt, the reader will get a better understanding of some excerpts from the classics of physical optics (for example, Newton's "Optics" or Huygens' "Treatise on Light") after they have been illustrated with the help of diagrams, constructions and concrete problems.

Thus, as he explores the world of refracted rays, the reader will be able to familiarize himself not only with the physics of the topics being considered but also with the evolution of some of the concepts of physics and their practical applications to problems, constructions and optical schemes. It is the authors' hope that this journey will be both instructive and enjoyable.

The authors are greatly obliged to Professor V. A. Fabrikant for his editing and for the many valuable suggestions he made.

L. Tarasov
A. Tarasova

Contents

Preface 5

Chapter One. Light rays at the interface between two media 11

A Ring at the Bottom of a Water-Filled Vessel 11. Ptolemy's Experiments 14. The Discovery of the Law of Refraction by Snell 15. Descartes' Interpretation of the Law of Refraction. Descartes' Error 17. Huygens' Principle 19. Huygens' Principle and the Law of Refraction 21. Fermat's Principle (the Principle of Least Time) 24. Deduction of the Law of Refraction from Fermat's Principle 26. Application of Fermat's Principle 28. Total Internal Reflection of Light. Critical Angle of Reflection 33. Tracing Refracted Rays 36. Cherenkov Radiation and the Laws of Refraction and Reflection of Light 38.

Chapter Two. Refraction of light in the Earth's atmosphere 42

Refraction of Light in the Atmosphere; the Angle of Refraction 42. Early Ideas about Refraction of Light in the Atmosphere 44. Refraction of Light According to Kepler 45. Reconstruction of Newton's Theory of Refraction on the Basis of His Correspondence with Flamsteed 50. The Exponential Lowering of Atmospheric Density with Height 52. Peculiar Sunsets. Appearance of a "Blind Strip" 54. Stellar Scintillation 59. Curvature of a Light Ray in Optically Nonhomogeneous Media 60. Mirages 63

Chapter Three. Passage of light through a prism 67

Refraction of a Light Ray in a Prism. Deflection of a Ray 67. Symmetric and Asymmetric Passage of a Light Ray Through a Prism 68. Refractometers 71. Double Images of Distant Objects Reflected by Window-Panes 73. Reflecting Prisms 75. The Lummer-Brodhun Photometer 78. The Reflecting Prism as a Component of a Laser Resonator 80. Biprism 81.

Chapter Four. Prisms and dispersion of light 83

Dispersion of Light 83. The First Experiments with Prisms. Theories Concerning the Nature of Colours, Which Preceded Newton's Ideas 84. Newton's Experiments with Prisms; Newton's Theory of the Nature of Colour 85. Euler's Work; Correspondence of Colour to Wavelength 96. The Discovery of Anomalous Dispersion; Kundt's Experiments 98. More about Reflecting Prisms 101. Dispersive Prisms. Angular Dispersion 104. Spectral Instruments—Monochromators and Spectrometers. The Fuchs-Wordsworth Optical System 105. Goethe Versus Newton 109.

Chapter Five. The rainbow 113

The Rainbow to the Eyes of a Careful Observer 113. Interpretations of the Origin of Rainbow: from Fleischer to Newton 117. Newton's Explanation of the Origin of Rainbow in His "Lectures on Optics" 118. Passage of a Light Ray through a Raindrop 119. The Greatest Angle between the Incident and the Emergent Rays 122. Sequence of Colours in Primary and Secondary Rainbows 126. Rainbows on Other Planets 128. Factors Causing Halos. Halo and Rainbow 130.

Chapter Six. Formation of optical images 134

The Formation of an Image in a Camera Obscura 134. The Formation of an Image in a Lens System 135. Derivation of the Thin Lens Formula from Fermat's Principle 137. The Spherical and Chromatic Aberration of a Lens 141. Real and Virtual Images 142. Collecting and Diverging Lenses 145. A Lens in an Optically Dense Medium 148. The Early History of Lens Systems 148. Invention of a Telescope 150. Path of Rays through the Galilean Telescope. Angular Magnification 151. Galileo's Astronomical Observations 155. Kepler's "Dioptrics" and Other Books 156. Dollond's Achromatic Lens 158. Fresnel Zone Plate 160.

Chapter Seven. The eye 165

Two Kinds of Optical Instrument 165. The Structure and Optical System of the Human Eye 167. A System Made Up of a Magnifying Glass and an Eye 168. The Development of the Science of Vision from Democritos and Galen to Alhazen and Leonardo da Vinci 170. Leonardo da Vinci's Comparison of the Eye to the Camera Obscura 172. Kepler on the Role of the Crystalline Lens. Young's Explanation of Accommodation 174. Presbyopia and Myopia 175. The Eye as a Perfect Optical Instrument 176. Spectacles 180. The Use of a Lens System to Increase the Angle of View 181. The Compound (Faceted) Eyes of Insects (184).

Chapter Eight. Double refraction in crystals 186

Bartholin's Discovery of Double Refraction in a Crystal of Iceland spar 186. Crystals as Optically Anisotropic Media 188. Huygens' Explanation of Double Refraction in the "Treatise on Light". Ordinary and Extraordinary Waves of Light

189. Huygens' Construction. Wave and Ray Velocities 192. Huygens' Experiments with Two Crystals (On the Verge of Discovering the Polarization of Light) 199. Newton's Interpretation of Huygens' Results 202. Malus' and Brewster's Investigations 203. Polarization of Light 204. Dichroic Plates and Polarization Prisms 209. Rotation of the Plane of Polarization of Light in a Half-Wave Plate 211.

Chapter Nine. What is fibre optics . 214

A Luminous Jet of Water 214. Light Rays in Straight and Curved Cylindrical Fibres 215. Rays in Conic Fibres 218. The Influence of Fibre Bending 221. Gradient Optical Fibres 222. Thin Fibres 222. Transmission of Optical Images over a Fibre Braid 223. Fibre Equalizer of the Light Field 225. Fibre Image Dissector in High-Speed Photography 225. The Retina as a Fibre-Optics Device 227.

Afterword 228

Control of the Refractive Qualities of a Substance 228. Electrooptic Deflector 231. Cosmic Lenses 233. Advice to the Reader Who Has Reached the End 235.

Index 237

Chapter One

Light rays at the interface between two media

A Ring at the Bottom of a Water-Filled Vessel. Take a shallow vessel with opaque walls; a mug, a tin or a pan will be suitable. Place a ring at the bottom of the vessel and look at it at an angle so that you can see a part of the bottom

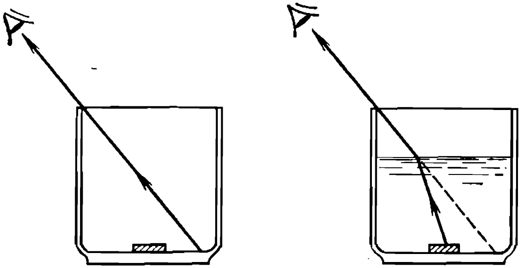


Fig. 1.1.

without seeing the ring. Ask somebody to fill the vessel with water without moving it. When the level of the water has reached a certain height, you will see the ring lying at the bottom. This unsophisticated experiment is an invariable success. It illustrates in a spectacular way the *refraction* of light rays at the interface between water and air (Fig. 1.1).

The experiment described above has been known for a long time. In 1557 a translation of Euclid's "Catoptrics" (3rd century B. C.) was published in Paris. It contains the following statement: "If an object is placed at the bottom of a vessel so that the object cannot be seen, it will come back into view if the vessel is filled with water, the distance remaining unchanged". True, the experiment described has no direct bearing on the question dealt with in Euclid's book. The latter is devoted to catoptrics, which was at that time the name of the branch of optics referring to the reflection of light, whereas the refraction of light was studied by dioptrics. The experiment with a ring at the bottom of a vessel is commonly supposed to have been added by the translator of the book. But still, there is not a shade of doubt that the experiment is about twenty centuries old. It is described in other ancient sources, particularly, in Cleomedes' book (50 A. D.) "The Circular Theory of the Heavenly Bodies". Cleomedes wrote: "Is it not possible that a light ray passing through humid layers of air should curve...? This would be similar to the experiment with a ring placed at the bottom of a vessel, which cannot be seen in an empty vessel, but becomes visible after the vessel is filled with water."

Consider quite a modern problem using the ancient experiment. *In a cylindrical vessel whose height equals the diameter of its bottom, there is a disc in the centre of the bottom whose diameter is half that of the bottom of the vessel. The observer can just see the edge of the bottom (naturally, he cannot see the disc lying at the bottom). How much of the vessel's volume has to be filled with water* *so that the observer can just see the edge of the disc? The refractive index of water $n = 4/3$.*

Designate the diameter of the bottom of the vessel as D, and the level of the water in the vessel at which the observer can see the edge of the disc as H (Fig. 1.2).

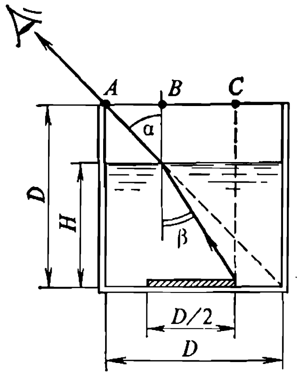


Fig. 1.2.

The law of the refraction of light rays is described by the relation

$$\frac{\sin\alpha}{\sin\beta}=n, \tag{1.1}$$

Rewrite the equation $AB + BC = AC$ as $(D - H)\tan\alpha + H\tan\beta = 3D/4$ or (bearing in mind that $\tan\alpha = 1$ under the conditions of the problem)

$$\frac{D}{H}=4(1-\tan\beta). \tag{1.2}$$

Passing from $\tan\beta$ to $\sin\beta$ and using Eq. (1.1), we have

$$\tan\beta=\frac{\sin\beta}{\sqrt{1-\sin^2\beta}}=\frac{\sin\alpha}{\sqrt{n^2-\sin^2\alpha}}=\frac{1}{\sqrt{2n^2-1}}. \tag{1.3}$$

Substituting (1.3) into (1.2) we find

$$\frac{D}{H}=4\left(1-\frac{1}{\sqrt{2n^2-1}}\right).$$

Since $n = 4/3$, $H/D \doteq 0.67$. Thus, the observer will be able to see the edge of the disc when water fills 0.67 of the vessel's volume.

Ptolemy's Experiments. In the problem considered above the *law of refraction* (1.1) was used. Many investigations conducted over a

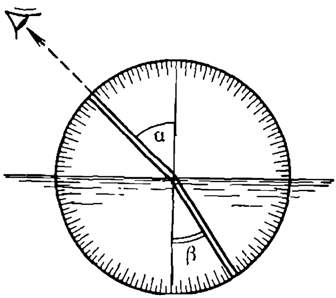


Fig. 1.3.

long period of time preceded the discovery of this law. They date back to the 2nd century A. D., when Ptolemy tried experimentally to determine the relationship between the angles which the incident and the refracted rays make with the normal to the interface between media.

Ptolemy used a disc graduated in degrees. The ends of two rulers were attached to the centre of the disc, so that the rulers could be turned about the fixed axis. The disc was half-submerged in water (Fig. 1.3), and the rulers were positioned in such a way that they both seemed to be in a straight line when viewed from the top. Ptolemy fixed the upper ruler in different positions (corresponding to different values of the angle α) and experimentally found the corresponding position of the lower ruler (the corresponding value of the angle β). It followed from Ptolemy's experiments that the ratio $\sin\alpha/\sin\beta$ laid within the range from 1.25 to 1.34, i.e. it was not quite constant. Thus, Ptolemy failed to discover the exact law of the refraction of light.

The Discovery of the Law of Refraction by Snell. Over four centuries passed before the law of refraction was at last established. In 1626 the Dutch mathematician Snell died. Amidst his papers a work was found, in which, in fact, he was found to have formulated the law of refraction. To illustrate Snell's conclusions, turn to Fig. 1.4. Assume that FO is the interface between the media; the rays are incident on the interface at point O. The figure shows three rays (*1*, *2*, and *3*); α_1, α_2, and α_3 are their angles of incidence, and β_1, β_2, and β_3 are the angles of refraction. Erect the perpendicular FG at a point F chosen at random on the interface between the media. Designate the points at which the refracted rays *1*, *2*, and *3* cut the perpendicular as A_1, A_2, and A_3, and those at which the extensions of the incident rays *1*, *2*, and *3* cut it (in the figure the extensions are represented by dashed lines) as B_1, B_2, and B_3. By experiment Snell established that

$$\frac{OA_1}{OB_1}=\frac{OA_2}{OB_2}=\frac{OA_3}{OB_3}.$$

Thus, the ratio of the length of the refracted ray from the point O to where it crosses FG to the length of the extension of the incident ray from O

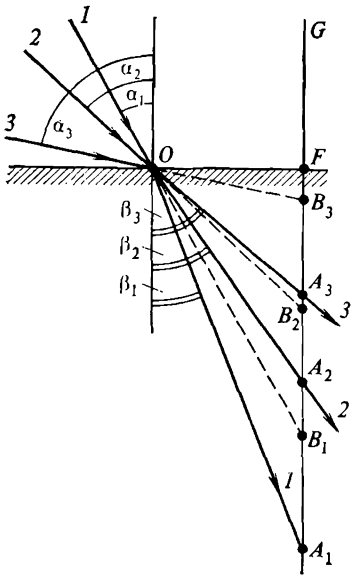


Fig. 1.4.

to where it crosses FG is constant for every ray incident on the interface:

$$\frac{OA_i}{OB_i} = \text{const} \tag{1.4}$$

(the index i indicates different rays).

The commonly accepted formula for the law of refraction follows immediately from (1.4).

Since $OA_i \sin \beta_i = FO$ and $OB_i \sin \alpha_i = FO$, formula (1.4) gives

$$\frac{\sin \alpha_i}{\sin \beta_i} = \text{const.} \qquad (1.5)$$

Thus, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is *constant* for a given pair of substances.

Descartes' Interpretation of the Law of Refraction. Descartes' Error. However, for some unknown reason Snell did not publish his work. The first publication which contains the wording of the law of refraction does not belong to Snell but to the famous French scientist René Descartes (1596-1650).

Descartes was interested in physics, mathematics and philosophy. He had an original and, undoubtedly, vivid personality, and opinions about him were many and controversial. Some of Descartes' contemporaries accused him of making use of Snell's unpublished work on the refraction of light. Whether Descartes did or did not see Snell's work, the accusation is groundless. The fact is that Descartes formulated the law of refraction on the basis of his own ideas about the properties of light. He deduced the law of refraction from the assumption that light travels at different velocities in different media, i.e. his law was arrived at *theoretically*.

Curiously enough, Descartes formulated the law of refraction using the erroneous assumption that the velocity of light increases when it goes from air into a denser medium. Today, we find Descartes' ideas about the nature of light rather confused and naive. He regarded the propagation of light as the transferrence of pressure through ether, a substance which, it was supposed, surrounded and penetrated everything. His work entitled "Dioptrics" reads: "Since there is no vacuum in nature and since each body has pores in it, it is necessary that these pores be

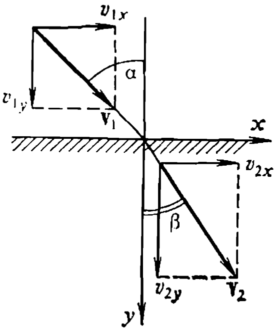


Fig. 1.5.

filled with matter, that is rather very rarefied and fluid, and which propagates incessantly from celestial luminaries towards us.... Light is nothing but a kind of motion or effect produced in the rather rarefied matter filling the pores of the bodies." When analysing the refraction of light, Descartes used an analogy with a ball thrown into water. He claimed that "light rays conform to the same laws as the ball".

Descartes' ideas regarding the refraction of light can be illustrated by Fig. 1.5. Assume that v_1 is the velocity at which light pressure is transferred in the first medium, and v_2 is the velocity in the second medium. Descartes resolved both vectors into two components—one parallel to the media interface (the x-component) and one perpendicular to the interface (the y-component). He supposed that when light leaves one medium and enters the other it is only the y-component of **v** that changes, and in a denser medium this component is greater. Putting it differently, we get

$$v_{1x} = v_{2x}; \quad v_{1y} < v_{2y}. \tag{1.6}$$

The figure shows that

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_{1x}/v_1}{v_{2x}/v_2} = \frac{v_2}{v_1}. \tag{1.7}$$

Descartes' major error was that he supposed that light propagates faster in a denser medium, whereas in reality it is *the other way around.* "The harder the particles of a transparent body", was Descartes' rather obscure reasoning, "the easier they let light pass through, for the light does not need to push any particles out of their place in the way a ball pushes aside particles of water to make its way through...".

Descartes' error was put right by Huygens and Fermat.

Huygens's Principle. The famous Dutch physicist and mathematician Christiaan Huygens (1629-1695) considered the propagation of light to be a wave process. Huygens supposed that light was in fact constituted by waves propagating through ether.

He looked upon the propagation of light waves in the following manner. Assume that the light wave is plane, the cross section of its wavefront being a straight line. Let it be line AA in

Fig. 1.6. Light reaches every point of AA simultaneously and, according to Huygens, all these points start functioning simultaneously as point sources of secondary spherical waves. As Huygens stressed in his "Treatisė on Light", "...light

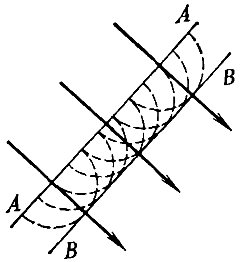


Fig. 1.6.

propagates in consecutive spherical waves". After a certain period of time Δt, these wave-fronts will create the situation shown in Fig. 1.6 by dashed semicircles. Draw the envelope of the fronts, which is actually the line BB. It corresponds to the new position of the plane wave-front. It can be said that within the time Δt the front of the light wave has moved from AA to BB. Naturally, every point on BB can also be regarded as the source of secondary light waves. In the figure, light rays are represented by arrows. At every point in space a light ray is *perpendicular* to the wave-front passing through the point.

This method of representing consecutive position of the wave-front became known as Huygens' method. It is also referred to as *Huygens'* *principle* and is formulated as follows: *every point reached by a light disturbance becomes in its turn the source of secondary waves, the surface enveloping these secondary waves at a given instant indicates the position of the actual propagating wave-front.*

Huygens' Principle and the Law of Refraction. Huygens deduced the law of refraction of light

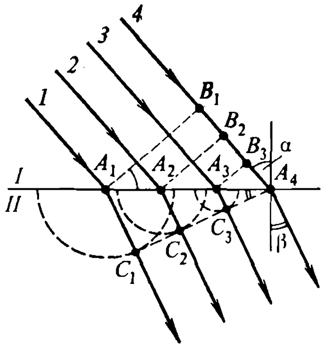


Fig. 1 7.

using his principle (Fig. 1.7). Assume that a plane light wave is incident at an angle α on a surface A_1A_4, which is the interface between two media, for example, water and air. Let the velocity of light in the first medium (air) be v_1, and the velocity in the second medium (water) be v_2. According to Huygens' correct reasoning, $v_1 > v_2$. Four light rays are shown by arrows in the figure; the line A_1B_1 (dotted) shows where the wave-front is at the moment when the ray *1* reaches the interface between the media. According to Huygens, at the same moment, the point A_1 becomes the source of a secondary spherical wave. Note that this wave continues to propagate in both the first and second media, generating reflected and refracted bundles of rays, respectively. We shall confine ourselves to the refracted waves. The dashed semicircle with its centre at A_1 shows the front of the spherical wave under consideration after a period of time Δt_1 during which the ray *4* travels from B_1 to A_4. We can write that

$$\Delta t_1 = \frac{B_1A_4}{v_1} = \frac{A_1C_1}{v_2}. \tag{1.8}$$

When the ray *2* reaches the interface, A_2 becomes the source of a secondary wave. The semicircle with the centre at A_2 (dotted) represents the front of this wave after a certain period of time Δt_2, during which the ray 4 travels from B_2 to A_4. Hence $\Delta t_2 = B_2A_4/v_1 = A_2C_2/v_2$. When the interface is reached by a ray *3*, point A_3 becomes the source of a secondary wave. The dotted semicircle with its centre at A_3 is actually the front of this wave after Δt_3, during which the ray *4* travels from B_3 to A_4, hence $\Delta t_3 = B_3A_4/v_1 = A_3C_3/v_2$. The line C_1A_4 is the envelope of the semicircles shown in the figure; it corresponds to the wave-front of the refracted bundle of rays at the moment the ray *4* reaches the interface. It is clear from the figure that $\sin\alpha = \frac{B_1A_4}{A_1A_4}$, $\sin\beta = \frac{A_1C_1}{A_1A_4}$, and therefore $\frac{\sin\alpha}{\sin\beta} = \frac{B_1A_4}{A_1C_1}$. Using (1.8), we have

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}. \tag{1.9}$$

Unlike (1.7), the correct relation between the velocities is written here.

In this way the constant relation $\sin\alpha/\sin\beta$ discovered by Snell was explained from two opposing theoretical premises: Descartes' erroneous assumption that the velocity of light in a dense medium is greater than it is in air and the correct though opposite assumption made by Huygens. You can thus see how one experiment can be used to substantiate different theories. It stands to reason that a theory is always based on and checked against an experiment. However, one should refrain from putting forward a new theory if it is based upon *insufficient* number of experiments. The history of physics has records of other examples, apart from Descartes' error, when theories formulated on the basis of insufficient experimental data were later proved to be incorrect by further tests. The creation of a new theory calls for a well-considered system of experiments to check it for viability as well as its compliance with other known facts and theories. A brilliant example here is the system of experiments with prisms the great Isaac Newton carried out. He used them to create his famous theory of the origin of colour. This will be given special consideration later (in Chapter Five), but now we should go back to the law of refraction.

We introduce the *refractive index* n for the given medium. According to the present-day views

$$n = \frac{c}{v}, \qquad (1.10)$$

where c is the velocity of light in vacuum (this fundamental physical constant equals $2.9979 \times \times 10^8$ m/s), and v is the velocity of light in the medium under consideration. Using (1.10) and (1.9), we can rewrite the law of refraction as follows:

$$\frac{\sin \alpha}{\sin \beta}=\frac{n_2}{n_1}, \tag{1.11}$$

where n_1 and n_2 are the refractive indices of the first and the second media, respectively. If light passes from air to a denser medium, for example, water or glass, the velocity of light in air can be assumed to be equal to c, i.e. the refractive index of air is unity. Then, we can write

$$\frac{\sin \alpha}{\sin \beta}=n, \tag{1.12}$$

where n is the refractive index of the denser medium.

Fermat's Principle (the Principle of Least Time). However, let us go back to the 17th century to familiarize ourselves with the investigation of Pierre Fermat (1601-1665), a well-known French mathematician. Fermat became interested in the refraction of light some years before Huygens. He came up with a general principle concerning the way light rays travel in different circumstances and, in particular, when light rays pass through an interface between two media. This is known as *Fermat's principle* or the *principle of least time*. The wording of the principle is: *the actual path of the propagation of light (the trajectory of a light ray) is the path which* *can be covered by light within the least time in comparison with all other hypothetical paths between the same points.*

Evidently, Fermat first conceived his idea when considering the statement of Hero of Ale-

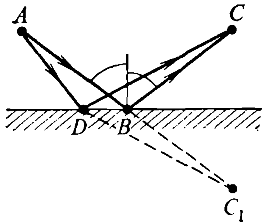


Fig. 1.8.

xandria (2nd century B. C.) that reflected light travels from one point to another along the shortest path. True, it is clear from Fig. 1.8 that ABC which complies with the law of reflection is shorter than any other imaginable path from the point A to C, for example, the path ADC. The length of ABC equals the length of the line AC_1, whereas the length of ADC actually equals the length of the broken line ADC_1 (C_1 is the mirror image of the point C).

It is quite obvious that the refraction of light does not obey the principle of the shortest path. Taking this fact into consideration, Fermat suggested that the *principle of shortest path* be replaced with the *principle of least time*. Fermat's principle explains the reflection of light in a very clear way. Besides, unlike the principle of shortest path, it accounts for the refraction of light as well.

The well-known "Feynman Lectures on Physics" have the following passage: "To illustrate that the best thing to do is not just to go in a straight line, let us imagine that a beautiful girl has fallen out of a boat, and she is screaming for help in the water at point B. The line

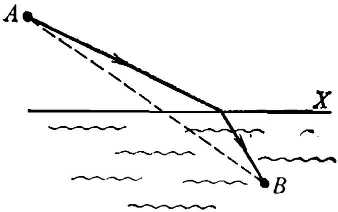


Fig. 1.9.

marked X is the shoreline (Fig. 1.9). We are at point A on land, and we see the accident, and we can run and can also swim. What do we do? Do we go in a straight line?... By using a little intelligence we would realize that it would be advantageous to travel a little greater distance on land in order to decrease the distance in the water, because we go much slower in the water."

Deduction of the Law of Refraction from Fermat's Principle. Now let us reason absolutely rigorously. Let the plane S be the interface between medium *1* and medium *2* with the refractive indices $n_1 = c/v_1$ and $n_2 = c/v_2$ (Fig. 1.10*a*). Assume, as usual, that $n_1 < n_2$. Two points are given—one above the plane S (point A), the other under the plane S (point B). The various distances are: $AA_1 = h_1$, $BB_1 = h_2$, $A_1B_1 = l$. We must find the path from A to B which can be covered by light faster than it can cover any other hypothetical path. Clearly, this path must consist of two straight lines, viz.

AO in medium *1* and OB in medium *2*; the point O in the plane S has to be found.

First of all, it follows from Fermat's principle that the point O must lie on the intersection of S and a plane P, which is perpendicular to S and passes through A and B.

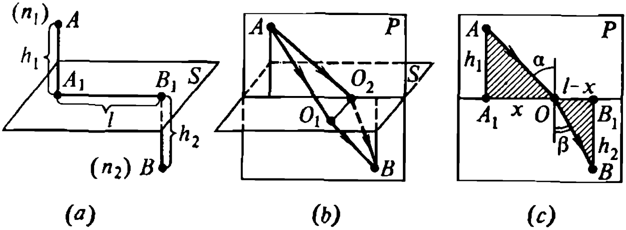


Fig. 1.10.

Indeed, let us assume that this point does not lie in the plane P; let this be point O_1 in Fig. 1.10*b*. Drop the perpendicular O_1O_2 from O_1 onto P. Since $AO_2 < AO_1$ and $BO_2 < BO_1$, it is clear that the time required to traverse AO_2B is less than that needed to cover the path AO_1B. Thus, using Fermat's principle, we see that the first law of refraction is observed: the incident and the refracted rays lie in the same plane as the perpendicular to the interface at the point where the ray is refracted. This plane is the plane P in Fig. 1.10*b*; it is called the plane of incidence.

Now let us consider light rays in the plane of incidence (Fig. 1.10*c*). Designate A_1O as x and $OB_1 = l - x$. The time it takes a ray to travel from A to O and then from O to B is

$$T = \frac{AO}{v_1} + \frac{OB}{v_2} = \frac{\sqrt{h_1^2 + x^2}}{v_1} + \frac{\sqrt{h_2^2 + (l-x)^2}}{v_2}. \qquad (1.13)$$

The time depends on the value of x. According to Fermat's principle, the value of x must minimize the time T. Those familiar with basic mathematical analysis know that at this value of x the derivative dT/dx equals zero:

$$\frac{dT}{dx} = \frac{x}{v_1\sqrt{h_1^2 + x^2}} - \frac{l-x}{v_2\sqrt{h_2^2 + (l-x)^2}} = 0. \qquad (1.14)$$

Now,

$$\frac{x}{\sqrt{h_1^2+x^2}}=\sin\alpha, \quad \text{and} \quad \frac{l-x}{\sqrt{h_2^2+(l-x)^2}}=\sin\beta,$$

consequently,

$$\frac{\sin\alpha}{v_1}-\frac{\sin\beta}{v_2}=0. \tag{1.15}$$

The second law of refraction described by the ratio (1.9) immediately follows from (1.15).

True, Fermat himself could not use (1.14) as mathematical analysis was developed later by Newton and Leibniz. To deduce the law of the refraction of light, Fermat used his own maximum and minimum method of calculus, which, in fact, corresponded to the subsequently developed method of finding the minimum (maximum) of a function by differentiating it and equating the derivative to zero.

Application of Fermat's Principle. Fermat's principle can be illustrated by the following

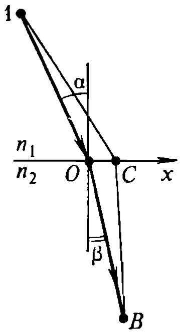


Fig. 1.11.

example. Let a light ray travel from A to B by passing through an interface between media with refractive indices n_1 and n_2 (Fig. 1.11). Let the distance $AO = OB = l$. Assume that the x-axis runs along the interface, and the origin of coordinates is at the point O, where the ray strikes the interface. Draw a broken line ACB (point C must lie on the interface between the media). According to Fermat's principle, the time required to traverse ACB will be greater than the time required to traverse the actual path AOB (for which $\sin\alpha/\sin\beta = n_2/n_1$) at any value of $x = OC$. Make sure this is true considering, for the sake of simplicity, sufficiently small values of x.

Using the law of sines for the triangles AOC and BOC, we have

$$AC = \sqrt{l^2 + x^2 + 2lx\sin\alpha} = l\sqrt{1 + (\eta^2 + 2\eta\sin\alpha)},$$
$$CB = l\sqrt{1 + (\eta^2 - 2\eta\sin\beta)} \quad (\eta = x/l). \tag{1.16}$$

Recall the approximate relation $\sqrt{1+\gamma} = 1 + \gamma/2$ which holds true for $\gamma \ll 1$. Since we are assuming that $x \ll l$ and, consequently, $\eta \ll 1$, we can use the above approximate equation and rewrite (1.16) as

$$AC = l\,(1 + \eta\sin\alpha + \eta^2/2);$$
$$CB = l\,(1 - \eta\sin\beta + \eta^2/2). \tag{1.17}$$

The time T required for the light to traverse AOB is $T = l\,(n_1 + n_2)/c$. Designate the time which it would take the light to traverse the path ACB as T_x, thus $T_x = (AC\cdot n_1 + CB\cdot n_2)/c$.

Substituting (1.17) for AC and CB and using (1.11) (remembering that $\eta = x/l$), we have

$$T_x = \frac{l}{c}(n_1 + n_2) + \frac{1}{2}\frac{l}{c}\left(\frac{x}{l}\right)^2 (n_1 + n_2) = T + Dx^2.$$

Obviously, $T_x > T$ whatever the sign of x, which is what we set out to prove.

Now let us use Fermat's principle to solve the following problem. *There is a coin at the bottom of a reservoir which has a depth H. We view it from above along a vertical line.*

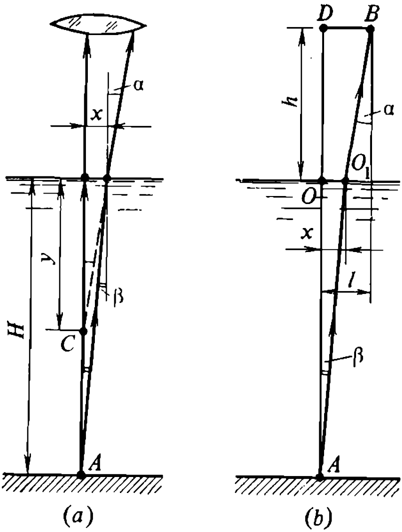


Fig. 1.12.

What is the apparent distance between the water surface and the coin? The refractive index of water n is given.

Figure 1.12*a* shows a greatly magnified crystalline lens of the observer. Two light rays from the coin enter it. One follows a strictly vertical path (it is not refracted), and the other enters it at a very small angle to the vertical (this ray is refracted at the interface between the water and air). The observer sees the coin where the extensions of the diverging rays arriving at the eye converge. The figure shows that this happens at the point C. So, the distance from the water surface to the coin is OC and we designate it as y.

To find the value of y, we have to know the relationship between the angles α and β, which follows from the law of refraction $\sin\alpha/\sin\beta = n$. Since in this instance the angles α and β are *very small*, we can safely use the approximate relations

$$\sin\alpha = \tan\alpha = \alpha, \quad \sin\beta = \tan\beta = \beta. \tag{1.18}$$

(Note that in (1.18) the angles must be measured in radians and not degrees.) Thus, in the problem under consideration, the law of refraction assumes a particularly simple form:

$$\frac{\alpha}{\beta} = n. \tag{1.19}$$

It follows from basic geometry (see Fig. 1.12) that $H\beta = x$ and $y\alpha = x$; so $H\beta = y\alpha$. With regard to (1.19), we get

$$y = \frac{H}{n}. \tag{1.20}$$

Our problem turned out quite simple provided we are familiar with the law of refraction. Now let us assume that we had no knowledge of the law of refraction. Fermat's principle would enable us to deduce (1.19) and through this resolve the problem.

The light ray travels from A to B; assume that $OD = h$, and $DB = l$ (see Fig. 1.12*b*). Designate the point at which the ray is refracted as O_1; $OO_1 = x$. We must determine the value of x for which the time required to traverse the path AO_1B is the least. The time T of transit over this path is described by the equation

$$T = \frac{n}{c}\frac{H}{\cos\beta} + \frac{1}{c}\frac{h}{\cos\alpha}, \tag{1.21}$$

where c is the velocity of light in vacuum (we hold that the velocity of light in air is the same). Using (1.18), we get

$$\cos\beta = 1 - 2\sin^2\frac{\beta}{2} = 1 - \frac{1}{2}\beta^2; \quad \cos\alpha = 1 - \frac{1}{2}\alpha^2. \tag{1.22}$$

Since $\xi \ll 1$, the following approximate relation holds true:

$$\frac{1}{1-\xi} = 1 + \xi. \tag{1.23}$$

Making use of (1.22) and (1.23), we can write (1.21) as

$$T = \frac{nH}{c}\left(1 + \frac{\beta^2}{2}\right) + \frac{h}{c}\left(1 + \frac{\alpha^2}{2}\right).$$

Since

$$\alpha = \frac{l-x}{h} \text{ and } \beta = \frac{x}{H}, \tag{1.24}$$

we have

$$T = \frac{nH}{c}\left(1 + \frac{x^2}{2H^2}\right) + \frac{h}{c}\left[1 + \frac{(l-x)^2}{2h^2}\right].$$

We must determine the value of x for which T is the least. In other words, we must find the value of x for which the following function reaches its minimum:

$$y(x) = n\,\frac{x^2}{H} + \frac{(l-x)^2}{h} = \frac{nh+H}{hH}\,x^2 - 2\,\frac{l}{h}\,x + \frac{l^2}{h}.$$

It is known that the x-coordinate of the vertex of the parabola $y = ax^2 + bx + c$ is $b/2a$. Consequently, the value of x we are looking for equals

$$x = \frac{lH}{nh+H}. \tag{1.25}$$

Substituting (1.25) into (1.24), we have $\alpha = \dfrac{ln}{nh+H}$, $\beta = \dfrac{l}{nh+H}$, whence $\alpha/\beta = n$.

Total Internal Reflection of Light. Critical Angle of Reflection. Up to now, while examining the refraction of light at an interface, we have virtually disregarded the reflection of light from the interface which occurs *simultaneously* with refraction. Strictly speaking, the two phenomena (refraction and reflection of light) should be considered together. This was proved in a most convincing way by the outstanding French scientist Augustin Jean Fresnel (1788-1827) who obtained the relationships for the intensity of the refracted and reflected beams of light with regard to the incident beam's intensity, the magnitude of the angle of incidence, and the polarization of the light. These relationships are known today as *Fresnel's formulae*. They have preserved their original form in modern optics.

Fresnel's formulae go beyond the confines of this book because we would need to use the electromagnetic theory of light to interpret them. Besides, polarization of light needs to be discussed separately. That is why we shall limit ourselves to a few general remarks concerning the interrelations between the intensities of the refracted and reflected beams of light, and examine the case when light passes from a medium with a higher refractive index to a medium with a lower one (in other words, from a dense to a less dense medium). This case is of special interest to us as it illustrates the phenomenon of *total internal reflection*.

Figure 1.13 shows four cases corresponding to different magnitudes of the angle of incidence α of a light beam. Light falls on an interface between two media with refractive indices n_1 and n_2 respectively, travelling from the medium with n_2 to the medium with n_1, and $n_2 > n_1$. As the value of the angle of incidence increases, the intensity of the refracted beam decreases, while the intensity of the reflected beam

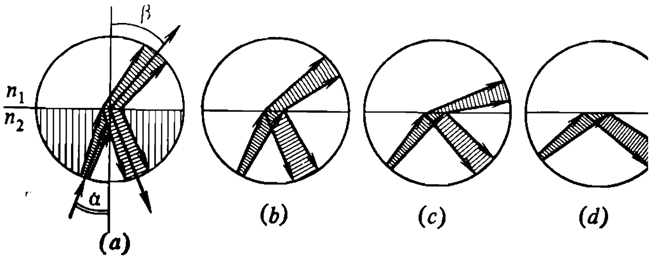


Fig. 1.13.

increases. When the angle of incidence reaches the value

$$\alpha_0 = \arcsin\left(\frac{n_1}{n_2}\right), \tag{1.26}$$

the angle of refraction β assumes the value of 90°, which follows directly from the law of refraction $\sin\alpha/\sin\beta = n_1/n_2$. The angle α_0 is termed as the *critical angle*. As the value of α approaches α_0, the intensity of the refracted beam decreases and equals zero at $\alpha = \alpha_0$. At $\alpha \geqslant \alpha_0$, the beam of light is completely reflected from the interface (Fig. 1.13*d*). We thus have the total internal reflection of light.

It has to be noted that when the critical angle is reached, the refracted ray changes into a reflected one all of a sudden. In fact, there is no sudden change. As the magnitude of the angle α approaches the critical angle, the intensity of the refracted beam *steadily* decreases and ultimately becomes zero, while the intensity of the reflected beam *steadily* increases and eventually equals the intensity of the incident beam. In this connection, let us point out once again the necessity of considering refraction and reflection together.

It is noteworthy that total internal reflection is better (more complete) than reflection from specially manufactured metal mirrors, which invariably absorb some of the energy of the incident beam.

A diver submerging under water should be familiar with the phenomenon of total internal reflection. Consider the following problem. *A diver, whose height is h,*

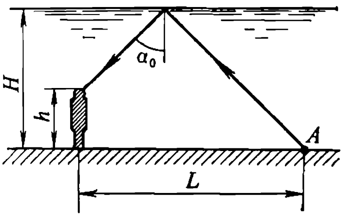


Fig. 1.14.

is standing at the bottom of a reservoir at a depth H. Find the minimum distance from the point where he stands to the points on the bottom he can see due to internal reflection from the water surface. The refractive index of water $n = 4/3$.

Designate the distance we are to find as L; the point A is the nearest point on the bottom the diver can see due to total internal reflection. The path of the light ray from the point A to the diver's eye is shown in Fig. 1.14. Th critical angle α_0 is determined by the equation

$$\sin \alpha_0 = \frac{1}{n}. \qquad (1.27$$

It follows from Fig. 1.14 that

$$L = h \tan \alpha_0 + 2(H - h) \tan \alpha_0. \qquad (1.28$$

Since $\tan \alpha_0 = \sin \alpha_0 / \sqrt{1 - \sin^2 \alpha_0}$ and using (1.27), this takes the form $L = (2H - h)/\sqrt{n^2 - 1}$. Thus, $L = (3/\sqrt{7})(2H - h)$.

Tracing Refracted Rays. There is a relatively simple way of tracing refracted rays. Figure 1.15 illustrates the methods. The upper part of the figure relates to the case when light travels from a medium with the smaller refracting index n_1 to the medium with the greater refractive index n_2, whereas the lower part relates to the opposite case. In the figure $n_1 = 1.4$ and $n_2 = 1.8$.

Look at the upper part of the figure. It shows two circles with a common centre at the point O situated on the interface. The ratio between the radii of the circles is the same as that between the refractive indices of the media. Let circle *1* have a radius proportional to n_1, and the radius of circle *2* be proportional to n_2. Draw in the incident ray with the angle of incidence α; it intersects circle *1* at the point A. Now draw a horizontal line from A and find B, at which the line intersects circle *1*. Then draw a vertical line downward from the point B to cut circle *2* at C. The refracted ray to be found will go through C. To ascer-

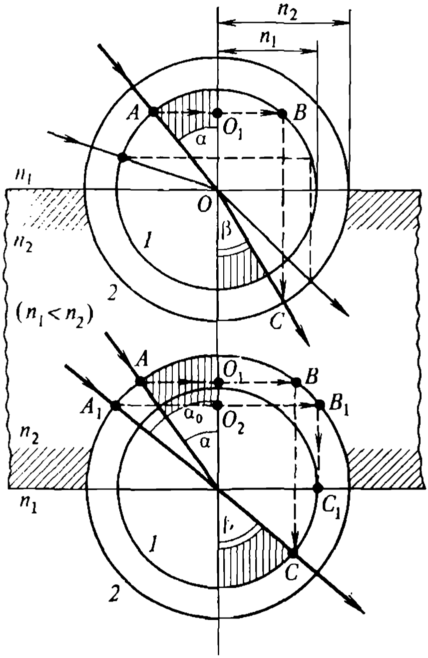


Fig. 1.15.

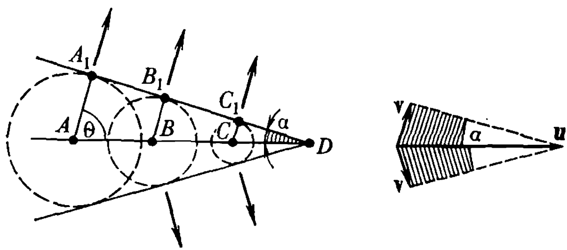


Fig. 1.16.

tain this, look at the following relations:

$$\sin\alpha = \frac{AO_1}{AO},$$

$$\sin\beta = \frac{O_1B}{OC},$$

$$\frac{\sin\alpha}{\sin\beta} = \frac{OC}{AO} = \frac{n_2}{n_1}.$$

Similar reasoning applies in the case depicted in the lower part of the figure. It shows the same circles *1* and *2*. But now A and B lie on circle *2* instead of circle *1*, whereas C is on circle *1*. Assume that the angle of incidence α_0 is such that the distance from the point where the incident ray cuts circle *2* to the vertical OO_1 (the line A_1O_2) equals the radius of circle *1*. In this case the refracted ray should go through the point C_1. The angle α_0 is the critical angle: $\sin\alpha_0 = A_1O_2/A_1O = n_1/n_2$.

Cherenkov Radiation and the Laws of Refraction and Reflection of Light. In conclusion, let us dwell upon a curious *analogy* between the laws of the refraction and reflection of light on the one hand, and the law which governs Cherenkov radiation, on the other hand. The *Cherenkov effect* was discovered in 1934 and occurs if an electron is moving in a medium at a velocity higher than that of light in the same medium. It then generates specific sort of radiation. The wave-front of this radiation can be constructed by using Huygens' principle. Figure 1.16 shows four consecutive positions of a moving electron A, B, C and D ($AB = BC = CD$). The electron moves at a velocity u and covers the distance from A to D in time $\Delta t = AD/u$. Each point of the electron's trajectory can be regarded as the source of a spherical light wave which starts emitting light the moment the electron passes through the point. The light waves generated by the electron propagate at the velocity v. Bear in mind that in this case $v < u$. In Fig. 1.16 the circles represent the fronts of these waves at the moment the electron reaches D. The radii of the circles are $AA_1 = v \cdot \Delta t$, $BB_1 = 2v \cdot \Delta t/3$, and $CC_1 = v \cdot \Delta t/3$; the envelopes encompassing the spherical wave-fronts are the straight lines A_1D and A_2D which make the angle α with the trajectory of the electron. Clearly enough, the following relation holds true for the angle α:

$$\sin \alpha = \frac{AA_1}{AD} = \frac{v}{u}. \tag{1.29}$$

Thus, the radiation dealt with materializes as two plane waves propagating at the angle $\theta = 90° - \alpha$ to the trajectory of the electron.

It is worth mentioning that Cherenkov radiation (1.29) not only arises from a fast-moving electron but also from any other "super-light" source. Let us imagine that a plane light wave falls at an angle α to an interface between two media with the refractive indices $n_1 = c/v_1$ and $n_2' = c/v_2$ (Fig. 1.17). Point A is the trace of the line of intersection of the incident light wave-front and the plane dividing the media. This point moves along the x-axis (along the interface) at

$$v' = \frac{v_1}{\sin \alpha}. \tag{1.30}$$

The velocity v' is greater than the velocity of light in either medium *1* or medium *2* (the latter is only true if $n_1 \leqslant n_2$ or the angle of incidence is less than or equal to the critical angle). This is why Cherenkov radiation will be generated in both medium *1* and medium *2*.

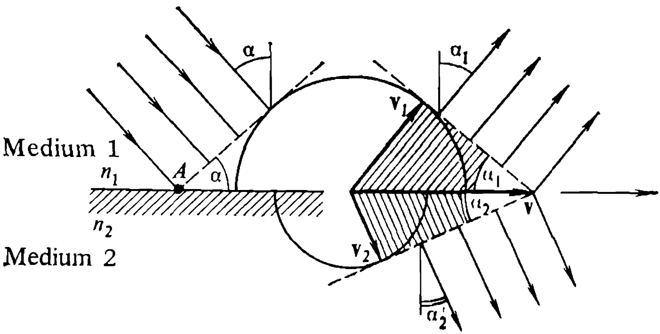


Fig. 1.17.

The fronts of the radiation make the angles α_1 (in medium *1*) and α_2 (in medium *2*) with the x-axis; these angles are different because the velocities of light in the two media are different. In accordance with (1.29), we can write that

$$\sin\alpha_1 = \frac{v_1}{v'}, \tag{1.31}$$

$$\sin\alpha_2 = \frac{v_2}{v'}. \tag{1.32}$$

Comparing (1.30) and (1.31), we get $\sin\alpha_1 = \sin\alpha$, that is, the law of reflection of light, whilst comparing (1.30) with (1.32), we get $\sin\alpha_2 = (v_2/v_1)\sin\alpha$, which corresponds to the law of refraction of light. Thus, there is a curious analogy between Cherenkov radiation and the reflection and refraction of light at the interface between two media (this analogy was noted by Academician I. Frank). Reflected and refracted light waves could be considered to be Cherenkov radiation generated in adjoining media by a "super-light" source which is actually a fast moving line along which the front of the incident light wave intersects the media interface.

If $n_2 < n_1$ and the incidence of the light on the interface is such that $\sin \alpha > n_2/n_1$, the velocity v' determined by (1.30) turns out to be lower than the velocity of light in medium *2*. In this case $v_1 < v' < v_2$. For this reason Cherenkov radiation will be generated only in medium *1*, which, very clearly, corresponds to the total internal reflection.

Chapter Two

Refraction of light in the Earth's atmosphere

Refraction of Light in the Atmosphere; the Angle of Refraction. In the previous chapter we assumed that the velocity of light in air equals the velocity of light in vacuum, i.e. we considered that the refractive index of air is unity. This assumption is, in fact, an approximation. It holds true when we consider the transition of a beam of light from air into water or from air into glass. However, it does not remain valid when the propagation of light through the Earth's atmosphere is being considered. We have to bear in mind not only that the refractive index is a little greater than unity, but also that it varies from point to point in accordance with the varying density of air. The atmosphere is an optically heterogeneous medium, which is why the path of a light ray in the atmosphere is, strictly speaking, always represented by a somewhat curved line. This curvature of the light rays is called the *refraction* of light in the atmosphere.

There are two kinds of refraction: *astronomical* and *terrestrial*. The former deals with the curvature of light rays which travel from celestial bodies (the Sun, the Moon, and the stars) to an observer on the Earth. The latter deals with the curvature of rays which come to the observer from objects positioned on the Earth. In both cases, the curvature of light rays influences the apparent position of the object; the object may look distorted. Sometimes an object can be observed even if it is, in fact, beyond the horizon line. Thus, the refraction of light in the Earth's

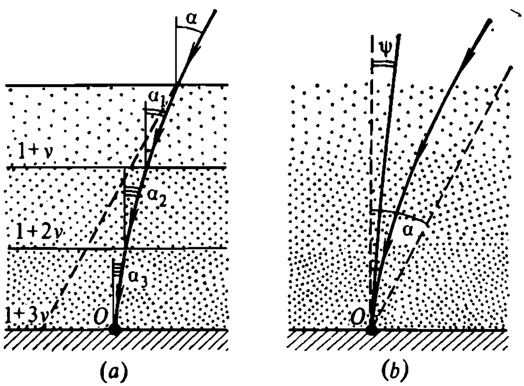


Fig. 2.1.

atmosphere leads to some very interesting *illusions.*

Assume that the Earth's atmosphere consists of a number of optically homogeneous, horizontal layers of the same thickness, the refractive index of each layer being different and increasing gradually from the upper layers to the lower ones. This purely hypothetical situation is shown in Fig. 2.1*a* showing the atmosphere having three layers with the refractive indices $1 + \nu$, $1 + 2\nu$ and $1 + 3\nu$, where $\nu \ll 1$. The path of a ray of light reaching the observer from some celestial body is, in this case, represented by a broken line, and

$$\frac{\sin\alpha}{\sin\alpha_1}=1+\nu;\quad \frac{\sin\alpha_1}{\sin\alpha_2}=\frac{1+2\nu}{1+\nu};\quad \frac{\sin\alpha_2}{\sin\alpha_3}=\frac{1+3\nu}{1+2\nu}.$$

In reality, the density of the atmosphere, and consequently its refractive index, does not change erratically, with increases in height but continuously. This is why a ray of light travels along a curved and not a broken line. Such a path is shown in Fig. 2.1*b*. Imagine that the light ray comes to the observer from some celestial object. If it were not for the refraction of light in the atmosphere, the observer would see the object at the angle α (the angle α is taken in relation to the vertical and termed the zenith distance of the object). Due to refraction the observer does not see the object at the angle α but at the angle ψ. Since $\psi < \alpha$, the object seems to be *higher* above the horizon than it really is. In other words, the apparent zenith distance of the object is less than the actual zenith distance. The difference $\Omega = \alpha - \psi$ is the *angle of refraction*.

Early Ideas about Refraction of Light in the Atmosphere. The first mention of the refraction of light in the atmosphere seems to date back to the first century. A. D. Cleomedes' work *The Circular Theory of the Heavenly Bodies* which was cited earlier, reads: "Is it not possible that a ray of light passing through humid layers of air should curve, which is the reason why the Sun seems to be above the horizon after it has actually gone beyond it?" In the 2nd century A.D., Ptolemy correctly pointed out that refraction should not exist for light travelling from an object at its zenith, and should gradually increase as the object approaches the horizon line (i.e. as the zenith distance grows). The famous Arabian scientist of the 11th century Ibn al-Haytham known in the West as Alhazen was also interested in the refraction of light in the atmosphere. He noted that owing to the refraction of light we get a little more daylight. Using the extension of the daylight hours caused by refraction, Alhazen tried to calculate the height of the Earth's atmosphere.

Refraction of Light According to Kepler. The well-known German scientist, Johannes Kepler (1571-1630), in his modestly entitled "Appendix to Witelo" elaborated a theory of the refraction of light assuming that the atmosphere is a homogeneous layer of thickness H having the *same* density at any height. One should not be surprised at this assumption since in Kepler's time air was considered to be weightless. Almost half a century had to pass before Torricelli proved that the pressure of air is lower at greater altitudes.

Figure 2.2 shows the refraction of light in the atmosphere according to Kepler; R is the radius of the Earth, and H is the thickness of the air layer constituting the atmosphere. The angle $\Omega = \alpha_1 - \alpha_2$ is the angle of refraction. The light ray shown in the figure is refracted only when it enters the atmosphere (at the point A). Applying the theorem of sines to the triangle O_1OA, we have $\frac{O_1A}{\sin(180° - \psi)} = \frac{O_1O}{\sin\alpha_2}$ or, putting it in another way, $\frac{R+H}{\sin\psi}=\frac{R}{\sin\alpha_2}$. Remembering that $\alpha_2 = \alpha_1 - \Omega$, we find

$$\sin(\alpha_1-\Omega)=\frac{R\sin\psi}{R+H}. \tag{2.1}$$

Proceeding from Alhazen's assessments, Kepler took $H/R = 0.014$ and using (2.1) calculated

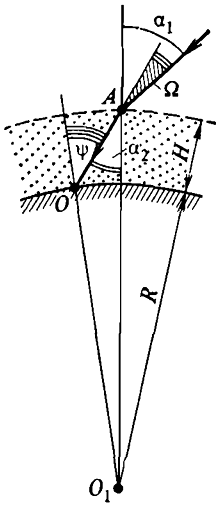


Fig. 2.2.

the angle $\alpha_1 - \Omega$ for $\psi = 90°$. It proved to be $80°29'$, i.e. considerably smaller than one could expect on the basis of the experimental data obtained at the time. To make it agree with the results of observations, a considerably smaller value of H/R (equalling approximately 0.001) had to be used in (2.1). This fact led Kepler to conclude that the light refraction was caused only by that part of the atmosphere which is adjacent to the Earth's surface and is no more than 5 km high. Actually, Kepler nearly found a clue to the discovery that the density of air decreases as

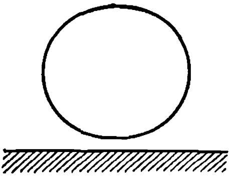


Fig. 2.3.

the altitude increases, but he never took the decisive step.

According to modern data, the *maximum* angle of refraction (the angle of refraction when $\psi = 90°$) equals 35′. When enjoying a sunset at the seaside, we see the lower edge of the sun touch the horizon line but we seldom realize that in fact the edge is 35′ *below* the horizon. Curiously enough, the upper edge of the solar disc is less elevated by the refraction (only by 29′) since refraction decreases as the zenith distance gets smaller. That is why the setting sun looks somewhat oblate (Fig. 2.3).

Imagine for a moment that we are living in Kepler's times, and are aware that when $\psi = 90°$, the angle of refraction $\Omega = 35'$. Suppose that we know the refractive index of air close to the Earth's surface, viz. $n = 1 + \nu$, where $\nu = 2.92 \times 10^{-4}$. This corresponds to a temperature of 15 °C and normal atmospheric pressure.

Proceed further using Kepler's model of homogeneou atmosphere to determine the ratio H/R between the atmosphere's thickness and the radius of the Globe. In other words, *assuming that the atmosphere is optically homogeneous and its refractive index is* $1 + \nu = 1 + 2.92 \times \times 10^{-4}$, *find the ratio between the thickness of the atmo-*

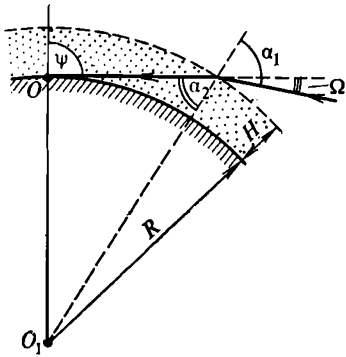


Fig. 2.4.

sphere and the radius of the Earth if it is known that the maximum angle of refraction is $35'$.

Figure 2.4 illustrates the situation. We designate $H/R = \xi$, assume that $\psi = 90°$, and rewrite (2.1) as

$$\sin(\alpha_1 - \Omega) = \frac{1}{1+\xi}. \tag{2.2}$$

Compliment this relation with the law of refraction at the boundary of the atmosphere:

$$\frac{\sin \alpha_1}{\sin(\alpha_1 - \Omega)} = 1 + \nu. \tag{2.3}$$

As a result, we have the following system of equations:

$$\left.\begin{aligned} \sin(\alpha_1 - \Omega) &= \gamma, \\ \sin \alpha_1 &= (1+\nu)\,\gamma, \end{aligned}\right\} \tag{2.4}$$

where $\gamma = 1/(1+\xi)$. Using the formula for the sine of difference of angles, transform the first equation of (2.4) to

$$\sin\alpha_1\cos\Omega - \sin\Omega\sqrt{1-\sin^2\alpha_1} = \gamma.$$

Substituting $\sin\alpha_1$ from the second equation of (2.4), we have

$$(1+\nu)\gamma\cos\Omega - \gamma = \sin\Omega\sqrt{1-(1+\nu)^2\gamma^2}. \tag{2.5}$$

Since the angle Ω is sufficiently small, we can use the approximate formulas (1.18) and (1.22):

$$\sin\Omega = \Omega, \quad \cos\Omega = 1-\frac{\Omega^2}{2}. \tag{2.6}$$

Now (2.5) becomes

$$\left(\nu-\frac{\Omega^2}{2}\right)\gamma = \Omega\sqrt{1-(1+2\nu)\gamma^2} \tag{2.7}$$

(we proceeded from the fact that $\nu \ll 1$ and, therefore $1+\nu)^2 = 1+2\nu$). We stress that in (2.6) the angle Ω must be expressed in radians. To transform $35' = (7/12)^\circ$ into radians, we use the identity

$$\frac{\pi}{180} = \frac{\Omega}{7/12}. \tag{2.8}$$

This gives $\Omega = 1.02 \times 10^{-2}$ rad.

By squaring both sides of (2.7) and performing some simple algebraic transformations, we find

$$\frac{1}{\gamma^2} = 1+\nu+\left(\frac{\nu}{\Omega}\right)^2+\left(\frac{\Omega}{2}\right)^2. \tag{2.9}$$

Since $1/\gamma^2 = (1+\xi)^2 = 1+2\xi$, we finally get

$$2\xi = \nu+\left(\frac{\nu}{\Omega}\right)^2+\left(\frac{\Omega}{2}\right)^2. \tag{2.10}$$

Substituting the values of $\nu = 2.92 \times 10^{-4}$ and $\Omega = 1.02 \times 10^{-2}$ into (2.10), we find $\xi = 5.7 \times 10^{-4}$. We assume that the radius of the Earth is $R = 6380$ km. Consequently, the height of the atmosphere, according to Kepler's model, is only $\xi R = 3.64$ km.

The result we have obtained should not discourage us, for in reality the density of the air, as well as the refractive index, gradually get smaller as the altitude increases. The great English scientist Isaac Newton (1643-1727) was well aware of this.

Reconstruction of Newton's Theory of Refraction on the Basis of His Correspondence with Flamsteed. Isaac Newton made an exceptionally valuable contribution to the development of the theory of astronomical refraction. Unfortunately, he did not include his results in this sphere either in his "Lectures on Optics" or in "Optics". Being extremely scrupulous in matters of scientific publication, Newton evidently underestimated the importance of his tables of the refraction of light. One of his letters dated 1695 reads: "I have no thoughts to write about refractions. The table of refractions I would rather not have yet communicated". It is only due to a stroke of luck that we are today aware of Newton's investigation of the refraction of light. In 1832, more than a hundred years after his death, Newton's twenty-seven letters to Flamsteed were discovered in an attic in London. Flamsteed was an astronomer at the Greenwich observatory, and Charles II had granted him the title "Astronomer Royal". The correspondence between Newton and Flamsteed began in 1680 in connection with the great comet which could be observed that year. It grew more extensive in the early 1690's when Newton was developing his more precise theory of the Moon's movement and was using the results of astronomical observations. In the mid-1690's, in his letters to Flamsteed Isaac Newton expounded a number of theorems dealing with the theory of refraction of light in the atmosphere, and cited his original and accurate tables of refraction which provided the angles of refraction for different values of the zenith distance.

In 1835, Newton's correspondence with Flamsteed was published by the Admiralty and though the book was not for sale, it was sent out to several research institutions and leading astronomers. In 1930's A. Krylov, a prominent Soviet scientist in the sphere of shipbuilding, bought the book quite by chance in a second-hand bookshop in London for two and a half shillings. Academician Krylov was well aware with Newton's work and had translated Newton's "Mathematical Principles of Natural Philosophy". On the basis of Newton's letters to Flamsteed and using only those mathematical means which Newton had at his disposal, Krylov reconstructed the reasoning and conclusions of the great Englishman and published them in his "Newton's Theory of Refraction" in 1935. In conclusion, Krylov wrote: "I went into all these details in order to show how complete and universal Newton's theory of astronomic refraction is. He created it in late 1694, early 1695, but unfortunately failed to publish it. If we develop Newton's theory using the elementary methods of analysis Newton had at his disposal, and compare it with modern theories, it will become clear at once how simple and natural his reasoning was and how little has been added over the past 240 years".

Before investigating Newton's theory of astronomical refraction, let us quote Newton's letter to Flamsteed dated October 24, 1694. Newton wrote: "I am of the opinion also that the refraction ... is varied a little by the different weight of the air discovered by the Baroscope. For when the air is heavier and by consequence denser it must refract something more than when it is lighter and rarer." At first Newton supposed that the density of the air decreases steadily (linearly) from the Earth's surface to the upper borders of the atmosphere. Proceeding from this assumption, he calculated his first table of refraction. Having discovered some differences between the results of his calculations and those of Flamsteed's observations, Newton began to work at a new table of refraction. He gave up the assumption of a linear reduction of air density with increasing height, and assumed that its density decreases in proportion to the lowering of pressure. He wrote, in this connection, that "the density of the air in the atmosphere of the Earth is as the weight of the whole incumbent air". Thus Newton came, in fact, to the conclusion that the density of the atmosphere decreases *exponentially* as the height increases.

The Exponential Lowering of Atmospheric Density with Height. In modern physics, this law is known as the *barometric height formula*:

$$\rho(h) = \rho(0) \exp\left(-\frac{mgh}{kT}\right), \qquad (2.11)$$

where $\rho(h)$ is the density of the air at the height h, T is the absolute temperature of the air, which is assumed to be constant at any height, g is free fall acceleration (9.81 m/s^2), and k is Boltzmann constant (1.38×10^{-23} J/K), since m designates molecular mass, strictly speaking, formula (2.11) only describes the change in density of one particular gas, i.e. one particular component of air such as oxygen, nitrogen, hydrogen, etc. The lighter a gas, the slower its density changes with height. The barometric height formula gives us only a very general description of how the atmosphere's density decreases with height; it does not take either wind, convective flows, or temperature fluctuations into account. Besides, it cannot be used for the heights above 100-200 km so that the dependence of the acceleration g on height can be neglected.

The barometric height formula is associated with Ludwig Boltzmann (1844-1906), a famous Austrian physicist. However, one should never forget that it was Isaac Newton who first pointed out the fact that the density of air decreases exponentially with height. He mentioned the fact in his research papers on the refraction of light in the atmosphere and subsequently used it to make up a more accurate table of refraction.

As atmospheric refraction was investigated further, the ideas concerning the general nature of the change in the atmosphere's refractive index with height were developed and perfected. This is shown in Fig. 2.5. Part (*a*) represents Kepler's theory, part (*b*) is Newton's original theory of refraction, and part (*c*) is Newton's improved and the modern, theory of the refraction of light in the atmosphere.

Peculiar Sunsets. Appearance of a "Blind Strip". When investigating the refraction of light, a number of other factors, many of which are sporadic by nature, must be allowed for along with the systematic changing of the density of air with height. More specifically, it should not be forgotten that the refractive index of air is influenced by convection currents, winds, and

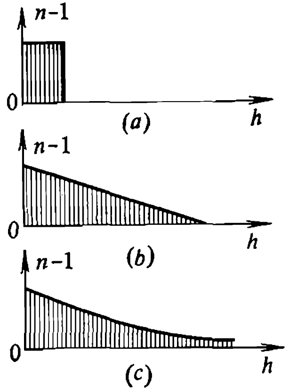


Fig. 2.5.

the different humidities and temperatures of air over different areas of the Earth's surface. In his letters to Flamsteed, Newton pointed out that such factors must always be taken into account when comparing his table of refraction with the results of observations. He wrote: "The reason of the different refractions near the horizon in the altitude, I take to be the different heat of the air in the lower region. For when the air is rarefied by heat it refracts less, when condensed by cold, it refracts more. And this difference must be most sensible when the rays run along in the lower region of the air ... because it is this region only which is rarefied and condensed by heat and cold, the middle and upper regions of air being always cold."

Specific conditions of the atmosphere, and primarily variable heating of its lower region

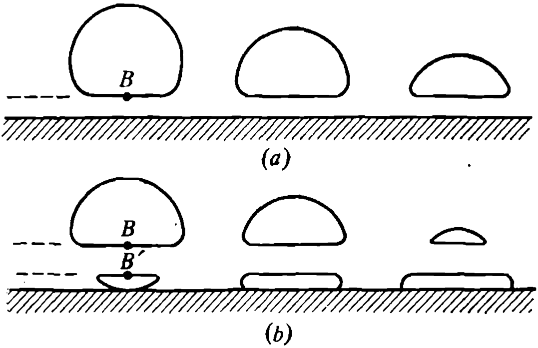


Fig. 2.6.

over different areas of the Earth's surface, result in quite peculiar sunsets. Thus, the sun sometimes does not seem to set behind the horizon but sets behind some invisible line above the horizon (Fig. 2.6*a*). Curiously enough, this phenomenon can be observed when the sky at the horizon is quite cloudless. If you were on the top of a hill in these conditions or the top storey of a building, or the upper deck of a big steamer, you will see something even more peculiar. Now the sun will be setting behind the horizon, but the solar disc will seem to be cut by a horizontal "blind strip" whose position in relation to the horizon remains unchanged (Fig. 2.6*b*).

This is usually observed when the air close to the Earth's surface is cold, while above it there is a layer of relatively warm air. In this case, the refractive index of air changes with height as shown in Fig. 2.7*a*; the transition from the cold lower layer of air to the warmer

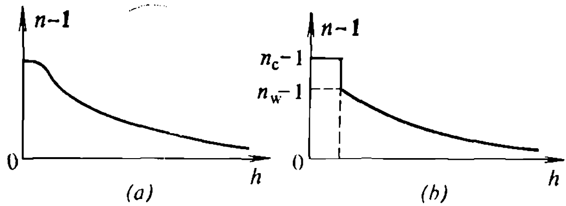


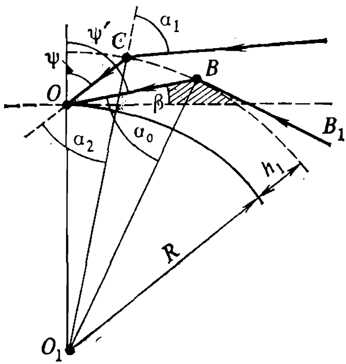
Fig. 2.7.

one above it may lead to a rather abrupt fall in the refractive index. For the sake of simplicity, assume that this fall is *sudden* and therefore there is a distinct interface between the cold and warm layers h_1 meters above the Earth's surface (Fig. 2.7*b*). In the figure, the refractive index of the air in the cold layer is designated n_c, and that in the warm layer close to the cold one is designated n_w.

The model represented in Fig. 2.7*b* is used in Fig. 2.8, which shows part of the Earth's surface and the adjacent layer of cold air at h_1. (The scale of the figure is necessarily distorted, for in reality h_1 is about 100 000 (!) times smaller than the Earth's radius R.) The observer is at the point O. The ray of light CO reaching him from a celestial body is refracted at the point C of the interface between the cold and warm layers, and

$$\frac{\sin\alpha_1}{\sin\alpha_2}=\frac{n_C}{n_W}=1+\nu, \tag{2.12}$$

where $\nu=(n_c-n_w)/n_w$ (it is evident that $\nu\ll 1$). The law of sines applied to ΔO_1OC



ig. 2.8.

determines the relationship between the angle ψ (the zenith distance) and the angle of refraction α_2, viz. $O_1O/\sin\alpha_2=O_1C/\sin(180^\circ-\psi)$, or

$$\sin\alpha_2=\frac{\sin\psi}{1+\xi}, \tag{2.13}$$

where $\xi=h_1/R$. (Note that formula (2.13) directly follows from (2.1).) It is clear from (2.13) that as the zenith distance ψ grows from zero to 90°, the angle α_2 increases and reaches its maximum value at $\psi = 90°$.

Now increase ψ gradually beginning with zero; the angle α_2 will increase too. Assume that at a certain value ψ' of the angle ψ the angle α_2 equals the critical angle α_0, characterizing total internal reflection at the interface between the cold and warm layers, in which case $\sin \alpha_1 = 1$. In Fig. 2.8, the ray BO corresponds to the angle α_0; it forms the angle $\beta = 90° - \psi'$ with the horizontal. It is evident that the observer will not see the rays which enter the cold layer at points whose angle of elevation above the horizon is less than that of point B, i.e. less than the angle β. This explains the sunset shown in Fig. 2.6*a*.

The angular width of the "blind strip" in Fig. 2.6*a* (i.e. the angle β in Fig. 2.8) is not difficult to calculate. In this respect, consider the following problem. *Find the angular width of the "blind strip" during the sunset shown in* Fig. 2.6*a if the height of the cold air layer* $h_1 = 50$ m, *and the ratio of the difference between the refractive indices of the cold and warm layers to the refractive index of the warm layer equals* $\nu = 10^{-5}$.

For the solution of the problem, use Fig. 2.8. Since the ray B_1B is perpendicular to O_1B, (2.12) can be written as

$$\frac{1}{\sin \alpha_0} = 1 + \nu. \tag{2.14}$$

The law of sines for O_1OB gives $O_1O/\sin \alpha_0 = O_1B/\sin \psi'$. Assuming that $\beta = 90° - \psi'$ and $h_1/R = \xi$, the previous equality takes the form

$$\cos \beta = (1 + \xi) \sin \alpha_0. \tag{2.15}$$

From (2.14) and (2.15), we have

$$\cos \beta = \frac{1+\xi}{1+\nu}. \tag{2.16}$$

Now make use of the fact that the angle β is rather small, and, by consequence, $\cos\beta = 1 - \beta^2/2$. Besides, ξ and ν are sufficiently small, too, so we can assume that $(1 + \xi)/(1 + \nu) = (1 + \xi)(1 - \nu) = 1 - (\nu - \xi)$. Now relation (2.16) becomes

$$\beta^2 = 2(\nu - \xi). \tag{2.17}$$

Thus,

$$\beta = \pm\sqrt{2(\nu - \xi)}. \tag{2.18}$$

The two signs in the last equation signify that the blind strip exists both above the horizon (the plus sign) and below it (the minus sign). To make sure that the blind strip exists under the horizon, it is sufficient for the observer to climb a hill. This matter will be dealt with a little later, but for the present consider only the blind strip which corresponds to the plus sign in (2.18). For $h_1 = 50$ m and $R = 6380$ km, we have $\xi = 0.78 \times 10^{-5}$. Substituting this value of ξ into (2.18) we find $\beta = 2.1 \times 10^{-3}$ rad $= 7.2'$.

Now, it will be easy to account for the kind of sunsets shown in Fig. 2.6*b*. If the observer is high up he can, in fact, see the rays characterized by the zenith angle which exceeds $90° + \beta$ (Fig. 2.9). In this case, he will see part of the solar disc below the blind strip, the angular width of which is 2β. Assuming that $\beta = 7'$, we come to the conclusion that the width of the blind strip crossing the solar disc is $14'$. The sun is seen at the angle of $32'$, consequently, the width of the blind strip in this case is a little less than half the diameter of the solar disc.

Stellar Scintillation. There is one more phenomenon connected with the astronomical refraction of light. This is *stellar scintillation* or, in common speech, the twinkling of stars. What happens is that air currents in the atmosphere make the angle of refraction change slightly in time for a star observed from the Earth's surface. For this reason, the star appears to a terrestrial observer to scintillate. Stellar scintillation close to the horizon is most perceptible, because then we observe the stars through a

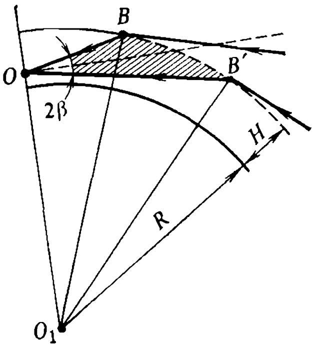


Fig. 2.9.

thick layer of atmosphere. Naturally, astronauts cannot observe the stellar scintillation.

Curvature of a Light Ray in Optically Nonhomogeneous Media. So far we have been dealing with the astronomic refraction of light (the curvature of light rays going from extraterrestrial bodies to an observer on the Earth). No less interesting is the terrestrial refraction of light which consists in the curvature of beams coming to the observer from objects situated on the Earth. This can lead to a most impressive class of phenomena known as *mirages*.

Before we investigate mirages, consider the following simple experiment. Take a vessel with transparent walls. For the sake of convenience, a common fishbowl can be used. Fill it with water and dissolve some sugar in it. The refractive index of the solution will vary continuously decreasing from the bottom of the vessel

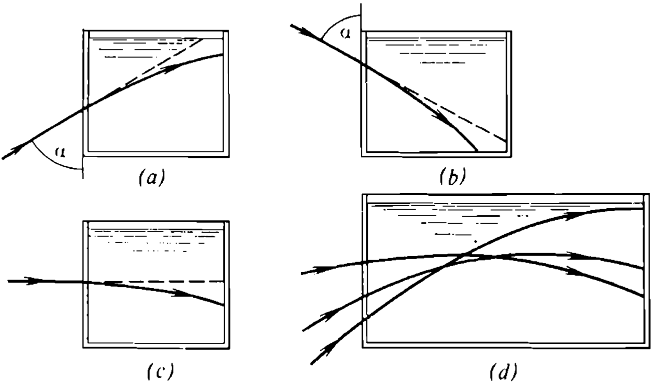


Fig. 2.10.

towards its upper part. Now send a narrow beam of light through the side wall.

First send the beam upwards at the angle α to the vertical (Fig. 2.10*a*). As the beam propagates to layers with a lower refractive index, its angle with the vertical will increase. The beam of light will curve inside the vessel, the direction of the beam approaching the horizontal.

Now send the beam from the top downwards at the angle α to the vertical (Fig. 2.10*b*). The angle between the beam and the vertical will decrease as the former passes through the layers with higher refractive indices. The beam of light will curve, its direction deviating more and more from the horizontal.

The situation described in both cases is easy to account for; suffice it to recall the examples of astronomical refraction discussed above. Now consider a more interesting case. Let a beam of light enter the vessel through the side wall

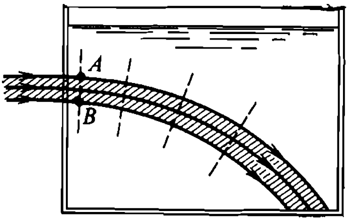


Fig. 2.11.

horizontally (Fig. 2.10*c*). It would seem at first sight as if there should be no change of direction while it travels inside the vessel. In reality, however, the beam of light propagating in the liquid curves more and more, being concave towards the optically denser layers.

This is not very difficult to account for, if we bear in mind that an infinitely narrow light ray is an idealization whereas in reality we have to deal with beams of light that have a finite size (finite aperture). Imagine an ideal plane-parallel beam of light entering the vessel horizontally. This kind of beam is shown in Fig. 2.11; the dotted lines represent cross sections of the wave-front of the beam at different points along its axis while the arrows indicate the light rays. Note that at any point, the wave-front is perpendicular to the light rays. Examine the front of the light beam AB at the moment it enters the liquid. Let v_A be the velocity of light at the point A, and v_B its velocity at the point B. Since the refractive index of the solution at A is smaller than it is at B, we have $v_A > v_B$. It follows immediately that the initially vertical wave-front of the beam (the front AB) will gradually wheel around as the beam propagates through the liquid.

The experiment described above leads to the following conclusion: if light propagates through a medium with a refractive index constantly decreasing from bottom to top, then regardless of the initial direction of the beam, it will invariably curve, its trajectory being convex upwards (see Fig. 2.10*d*). If the refractive index decreased downwards, the curved beam would be convex downwards. The two cases can be generalized into the following rule: *in an optically non-homogeneous medium*, the curved beam of light is always convex in the direction of the lower refractive index.

Mirages. This rule explains certain kinds of mirages. Figure 2.12 shows the appearance of the so-called superior mirage, or looming. It appears when the refractive index of the layer adjacent to the surface decreases fast enough with height, a condition that is possible when, for example, there is a layer of cold air near the ground and a layer of a warmer air above it.

The so-called inferior mirages can appear above heated surfaces (for instance, in deserts

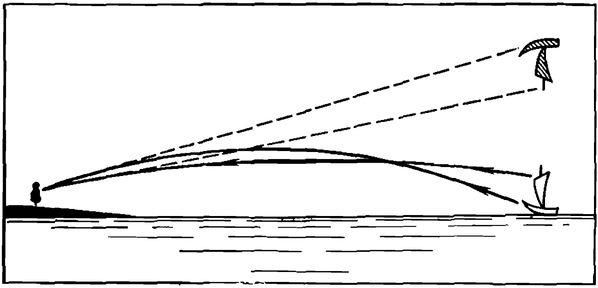


Fig. 2.12.

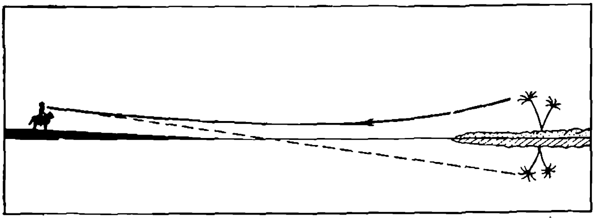


Fig. 2.13.

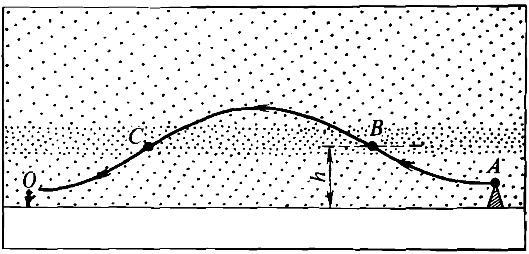


Fig. 2.14.

or above an asphalt road on a hot day). They include the lake mirages—visions of lakes with surrounding greenery reflected in them which appear in deserts (Fig. 2.13). The water of these "lakes" is a reflection of the sky.

There are many different kinds of mirages, and this is due to changes in the character of the locality where mirages are observed and the state of the atmosphere at the time. This can be illustrated by Fig. 2.14. It shows the path a ray takes from the object to the observer in the case when a very hot layer of air close to the surface (with a relatively low refractive index) is covered with a layer of rather cold air (with a much higher index of refraction). Before reaching the observer, the ray travels along a rather complicated path, which may cause peculiar mirages. Note that the path of the ray in Fig. 2.14 is always convex in the direction of the lower refractive index of the air. Divide this path into three sections. The sections AB and CO are convex downwards because within the lower layer whose height is h the refractive index decreases downwards. The BC section of the path is convex upwards as above h the refractive index decreases upwards.

Many kinds of books, both fact and fiction, have descriptions of mirages. The most impressive of them were given names of their own long ago and are associated with various legends and tales. There are legends about the Flying Dutchman (a ghost ship which haunts the seas and appears before seamen about to die), fata morgana (visions of palaces which appear on the horizon and vanish as one gets closer to them) and the Spectre of the Brocken (gigantic figures of people and beasts moving across the sky).

Most mirages, especially superdistant ones, in which case the image travels over thousands of kilometres, are rather complicated optical phenomena. Refraction of light in the atmosphere alone is not sufficient to account for their appearance, they are caused by far more intricate factors. Under certain conditions, enormous air lenses, peculiar lightguides, secondary mirages, i.e. mirages of mirages are formed in the atmosphere. It is quite probable that part of the reason why mirages occur is the presence of the ionosphere (a layer of ionized gases at an altitude of about 100 km) which can reflect light waves.

Chapter Three

Passage of light through a prism

Refraction of a Light Ray in a Prism. Deflection of a Ray. When passing through a prism, a ray of sunlight is not only *refracted* but is also *decomposed* into different colours. The decompo-

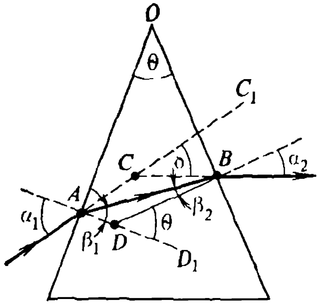


Fig. 3.1.

sition of light into different colours is dealt with in the next chapter. In this one we will only investigate the refraction of a ray in a prism. Strictly speaking, this means that the ray of light is assumed to be *monochromatic* (from the Greek "chromos" meaning colour and "mono" meaning one).

Figure 3.1 shows the passage of a ray of light through a prism with a refracting angle* θ

---

* The refracting angle of a prism is often called simply the prism angle (*transl. note*).

and refractive index n; assume that the refractive index of the medium surrounding the prism (air) equals unity. The figure shows that the angle of incidence of the ray on the left face of the prism is α_1. Using the law of refraction at the points A and B, we have

$$\frac{\sin\alpha_1}{\sin\beta_1}=n; \quad \frac{\sin\alpha_2}{\sin\beta_2}=n. \tag{3.1}$$

When passing through the prism, the ray is deflected from its original path by the angle C_1CB; we will designate it as δ and refer to it as the angle of deviation. Since $\angle C_1CB = \angle CAB + \angle CBA$, we can write

$$\delta = (\alpha_1 - \beta_1) + (\alpha_2 - \beta_2). \tag{3.2}$$

Note that $\angle D_1DB = \angle DAB + \angle ABD = \beta_1 + \beta_2$. Since $\angle D_1DB = \angle AOB$, we conclude that

$$\beta_1 + \beta_2 = \theta. \tag{3.3}$$

Given (3.3), (3.1) and (3.2) become

$$\left.\begin{array}{l} \sin\alpha_1/\sin\beta_1 = n, \\ \sin\alpha_2/\sin(\theta-\beta_1) = n, \\ \delta = \alpha_1 + \alpha_2 - \theta. \end{array}\right\} \tag{3.4}$$

Symmetric and Asymmetric Passage of a Light Ray Through a Prism. Consider the following problem. *Find the angle of deviation δ of the ray passing through a prism with refracting angle θ and refractive index n when the ray inside the prism is perpendicular to the bisector of the refracting angle.* In this case, the ray's path through the prism is symmetric about the bisector of the refracting angle (Fig. 3.2*a*). Consequently, $\alpha_1 = \alpha_2 \equiv \alpha$, $\delta = 2\alpha - \theta$, $\beta_1 = \beta_2 \equiv \beta = \theta/2$. Using these equations, rewrite the law of refraction $\sin\alpha/\sin\beta = n$ as

$$\sin\frac{\delta+\theta}{2} = n \sin\frac{\theta}{2}. \tag{3.5}$$

We find that $\frac{\delta+\theta}{2} = \arcsin\left(n \sin\frac{2}{\theta}\right)$. Finally,

$$\delta = 2\arcsin\left(n \sin\frac{\theta}{2}\right) - \theta. \tag{3.6}$$

Consider yet another problem. *Find the angle of deviation δ′ of the ray in a prism with refracting angle θ and refractive index n, if the ray falls normally on the first*

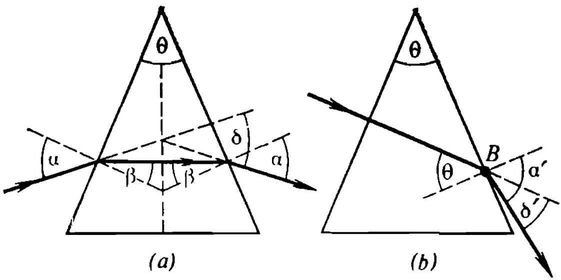


Fig. 3.2

face of the prism. In the previous problem, we dealt with the symmetric passage of the light through a prism (the ray was refracted equally at the two faces of the prism). Now the situation is quite different: the emergent ray is considerably refracted while the incident ray is not refracted at all (Fig. 3.2*b*). The law of refraction at the point B takes the form of $\sin\alpha'/\sin\theta = n$. From here, we have

$$\delta' = \alpha' - \theta = \arcsin(n \sin\theta) - \theta. \tag{3.7}$$

It is worthwhile ascertaining that $\delta' > \delta$. According to (3.6) and (3.7), $\delta' - \delta = \arcsin(n\sin\theta) - 2$

$\arcsin(n \sin(\theta/2)) \equiv \varphi - \psi$. That is why it is sufficient to show that $\sin\varphi - \sin\psi > 0$. Since $\sin(\arcsin\gamma) = \gamma$, $\cos(\arcsin\gamma) = \sqrt{1-\gamma^2}$, we get

$$\sin\varphi - \sin\psi = n\sin\theta - 2n\sin(\theta/2)\sqrt{1-n^2\sin^2(\theta/2)} =$$
$$= 2n\sin(\theta/2)\left[\sqrt{1-\sin^2(\theta/2)} - \sqrt{1-n^2\sin^2(\theta/2)}\right].$$

The first root in the square brackets is greater than the second root, and consequently, we have proved what we set out to. This example illustrates that in the case of asymmetric refraction of the ray in the prism, the angle of deviation increases.

In his "Lectures on Optics" Isaac Newton proves, using geometrical reasoning, that when homogeneous rays are refracted in the prism, the angle between the incident and emergent rays is at a maximum if the refraction is the same at the first and second faces. The rays Newton described as homogeneous are actually what we now call monochromatic, and the angle between the incident and the emergent rays is actually the angle ACB (see Fig. 3.1), i.e. $180° - \delta$. So, *the deflection of the ray is the least when its passage through the prism is symmetric.*

We will prove this statement by differentiating. Let the angle δ be a function of the angle β_1 which is in one-to-one correspondence with the angle α_1 of incidence of the ray on the first face of the prism. According to the third equation of system (3.4), we have $\delta(\beta_1) = \alpha_1(\beta_1) + \alpha_2(\beta_1) - \theta$. To find the value of the angle β_1 at which the angle δ is a minimum, we differentiate δ with respect to β_1 and set the derivative equal to zero

$$\frac{d\delta}{d\beta_1} = \frac{d\alpha_1}{d\beta_1} + \frac{d\alpha_2}{d\beta_1} = 0. \tag{3.8}$$

From the first equation of the system (3.4), we have $\alpha_1(\beta_1) = \arcsin(n \sin \beta_1)$, and the second equation gives $\alpha_2(\beta_1) = \arcsin(n \sin(\theta - \beta_1))$. Recalling that

$$\frac{d}{dx} \arcsin f(x) = [1 - f^2(x)]^{-1/2} \frac{df}{dx},$$

we obtain

$$\frac{d\alpha_1}{d\beta_1} = \frac{n \cos \beta_1}{\sqrt{1 - n^2 \sin^2 \beta_1}}, \quad \frac{d\alpha_2}{d\beta_2} = \frac{-n \cos(\theta - \beta_1)}{\sqrt{1 - n^2 \sin^2(\theta - \beta_1)}}.$$

Substituting these derivatives into (3.8), we get

$$\cos \beta_1 \sqrt{1 - n^2 \sin^2(\theta - \beta_1)}$$
$$= \cos(\theta - \beta_1) \sqrt{1 - n^2 \sin^2 \beta_1}$$

or

$$[(1 - \sin^2 \beta_1)(1 - n^2 \sin^2(\theta - \beta_1))]^{1/2}$$
$$= [(1 - \sin^2(\theta - \beta_1))(1 - n^2 \sin^2 \beta_1)]^{1/2}.$$

This gives $\beta_1 = \theta/2$, which corresponds to the symmetric passage of the ray through the prism.

Refractometers. Refraction by a prism is used practically in some kinds of *refractometers.* The word "refractometer" is used to describe optical devices for determining refractive indices. The sample to be tested must be in the form of a prism with thoroughly polished, refracting surfaces. The liquid is poured into hollow prismatic cell with plane parallel walls. The prism is placed on the rotatable table of a goniometer equipped with a telescope and a collimator (a collimator is a device for obtaining a narrow beam of light). The table is rotated to find the position at which the narrow beam of light incident on the prism is deflected least when passing through the prism. By measuring the angle of deviation δ in this position, the refractive index n can be determined using formula (3.5). Despite its obvious simplicity, the method is quite pre-

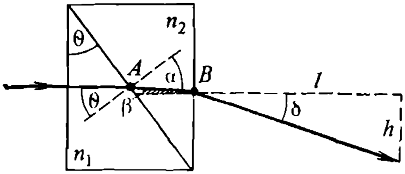


Fig. 3.3.

cise. If the angles δ and θ are measured to an accuracy of several seconds of arc, the refractive index can be determined to within 10^{-5}.

Assume that we have to determine the difference $n_1 - n_2$ of the refractive indices of two liquids when this difference is negligibly small. In this case, we can investigate the consecutive refraction of a ray of light in two prismatic vessels with plane parallel walls containing different liquids. Figure 3.3 shows this system of prisms and the passage of a ray of light through it. The ray is refracted at the points A and B, and $\sin\alpha/\sin\theta = n_1/n_2$, $\sin\delta/\sin\beta = n_2$. It follows from this that $n_1 = n_2 \sin\alpha/\sin\theta = \sin\delta\sin\alpha/\sin\beta\sin\theta$, and thus we obtain

$$n_1 - n_2 = \frac{\sin\delta}{\sin\beta}\left(\frac{\sin\alpha}{\sin\theta} - 1\right). \tag{3.9}$$

Bear in mind that $\alpha = \beta + \theta$, consequently, $\sin\alpha = \sin\beta\cos\theta + \sin\theta\cos\beta$. Since the angle β is very small (remember that the refractive indices n_1 and n_2 are almost equal), assume that in (3.9) $\cos\beta = 1$, which gives

$$n_1 - n_2 = \frac{\sin\delta}{\tan\theta}. \qquad (3.10)$$

The value of $\sin\delta$ is usually determined by measuring l and h (see Fig. 3.3). Since the value of δ is sufficiently small, we can assume that $\sin\delta = \tan\delta = h/l$. Thus,

$$n_1 - n_2 = \frac{h}{l\tan\theta}. \qquad (3.11)$$

This method provides a way of determining the difference of refractive indices to within 10^{-7}.

Double Images of Distant Objects Reflected by Window-Panes. Refraction of light rays in a prism with a small refracting angle makes it possible to account for something that we often observe, though we seldom give it much thought. Look at a distant lamp or the moon reflected in a window. Most likely, you will see *two* images, and as you change your position, one of them will move irregularly in relation to the other. This "ghosting" can be explained by the fact that glass panes are somewhat *wedge-shaped*; one image is produced by the reflection of light from the front plane of the glass, and the other by its reflection from the back plane. Figure 3.4 illustrates this phenomenon. The ray SA from a distant source of light is partially reflected at the point A, and when it reaches the observer, it forms the first image. The same ray partially refracted at A is subsequently partially reflected at the point B and refracted at C. As a result, the observer will see a second image. The greater the angle is between the rays AA_1 and CC_1, the further apart are the images we can observe.

Consider the following problem. *Find the angle between the two images shown in a wedge-shaped glass plate with*

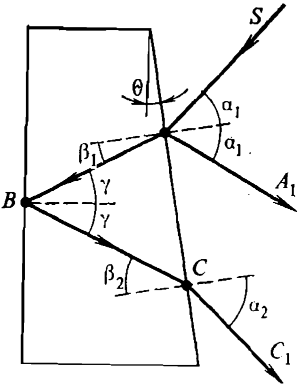


Fig. 3.4.

refracting angle θ and refractive index n if the angle of incidence of the ray on the front plane is $\alpha_1 = 30°$.

Designate the angle to be found by φ. It follows from Fig. 3.4 that $\varphi = \alpha_2 - \alpha_1$, where α_2 is the angle of the emergent ray CC_1 with the plate. The refraction of the ray at A and C is described by formulae (3.1). It is evident that the relation between the angles β_1 and β_2 is such that $\beta_1 + \theta = \gamma = \beta_2 - \theta$; therefore,

$$\beta_2 = \beta_1 + 2\theta. \tag{3.12}$$

Using (3.12), we rewrite (3.1) as

$$\frac{\sin\alpha_1}{\sin\beta_1} = n, \quad \frac{\sin\alpha_2}{\sin(\beta_1 + 2\theta)} = n. \tag{3.13}$$

As the angle θ is small, we have $\sin(\beta_1 + 2\theta) = \sin\beta_1 + 2\theta\cos\beta_1$. Using the first equation of (3.13) which gives $\sin\beta_1 = (\sin\alpha_1)/n$, we have

$$\sin(\beta_1 + 2\theta) = \frac{\sin\alpha_1}{n} + 2\theta\sqrt{1 - \frac{\sin^2\alpha_1}{n^2}}. \tag{3.14}$$

Substituting (3.14) into the second equation of (3.13), we get

$$\sin\alpha_2 = \sin\alpha_1 + 2\theta\sqrt{n^2 - \sin^2\alpha_1}. \tag{3.15}$$

The angle φ to be found is known to be very small, that is why we can assume that $\varphi = \sin\varphi = \sin(\alpha_2 - \alpha_1)$ and, consequently,

$$\varphi = \sin\alpha_2\cos\alpha_1 - \sin\alpha_1\cos\alpha_2. \tag{3.16}$$

For $\alpha_1 = 30°$, the expression (3.15) becomes

$$\sin\alpha_2 = \frac{1 + 2\theta\sqrt{4n^2 - 1}}{2}. \tag{3.17}$$

Bearing in mind that the angle θ is very small, we find step by step

$$\sin^2\alpha_2 = \frac{1 + 4\theta\sqrt{4n^2 - 1}}{4},$$

$$\cos^2\alpha_2 = 1 - \sin^2\alpha_2 = \frac{3}{4}\left(1 - \frac{4\theta\sqrt{4n^2 - 1}}{3}\right), \tag{3.18}$$

$$\cos\alpha_2 = \frac{\sqrt{3}}{2}\left(1 - \frac{2\theta\sqrt{4n^2 - 1}}{3}\right).$$

Substituting (3.17) and (3.18) into (3.16), we finally get

$$\varphi = \frac{2}{\sqrt{3}}\theta\sqrt{4n^2 - 1}. \tag{3.19}$$

Reflecting Prisms. These are a special kind of prism, based on total internal reflection. A light ray enters this kind of prism and is reflected internally one or more times, and then emerges from the prism. The angle between the emergent ray and the exit face of the prism equals the angle of the incident ray to the entry face. Quite often the incident and emergent rays are perpendicular to the respective prism faces. As a esult, reflecting prisms do not decompose white

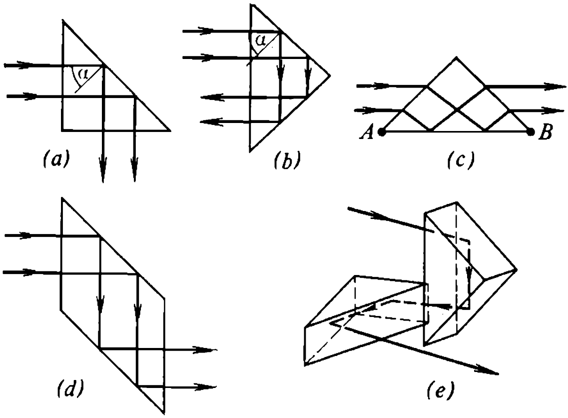


Fig. 3.5.

light into different colours, which is similar to the passage of white light through a plane-parallel plate. That is why the assumption of the original beam of light being monochromatic, which we made at the beginning of the chapter, proves to have been irrelevant when investigating reflecting prisms.

Figure 3.5 shows certain types of reflecting prisms. Clearly, reflecting prisms can be used to change the direction of a beam of light, as well as to achieve parallel shifting of the beam and to turn the image upside down. Figures 3.5*a*, 3.5*b*, and 3.5*c* show the same prism. Its cross section is an isosceles right-angled triangle. In the first two cases the prism changes the direction of the beam of light by 90° and 180° respectively. In the third case it does not change the direction of the beam but instead it turns the image upside down.

Note that in the cases (*a*), (*b*), (*c*), and (*d*), the light is not refracted at all; it only undergoes total internal reflection. That is why in the cases mentioned above white light is not decomposed into colours. In the case (*c*) the light, besides being totally internally reflected, is also refracted. If a beam of white light is incident on the prism, in this case, the emerging rays will be of different colours. It is worth mentioning, however, that the paths of the emergent rays will be parallel to one another and, consequently, the dispersion of the white light will not be noticeable (because the beam of light has a certain width, the coloured rays will overlap). This question will be tackled in the next chapter.

Assume that a prism surrounded by air has a refractive index n. In Fig. 3.5*a* and *b* the rays are incident to the reflecting face at an angle $\alpha = 45°$. For total internal reflection to occur, the condition $\sin \alpha > 1/n$ must be observed or, in other words, it is necessary that $n > \sqrt{2}$. Curiously enough, in the case (*b*) total internal reflection from the face AB occurs at any value of $n > 1$. We prove this by referring to Fig. 3.6.

By using the equation $\angle FCD + \angle CDF = \angle CFA$, we rewrite it as $\beta + (90° - \alpha) = 45°$, or $\alpha = 45° + \beta$. The condition for total internal reflection,

$\sin\alpha > 1/n$, assumes the form $\sin(45° + \beta) > 1/n$, or

$$\sin\beta + \cos\beta > \frac{\sqrt{2}}{n}. \qquad (3.20)$$

The law of refraction for the point C is $\sin 45°/\sin\beta = n$, or

$$\sin\beta = \frac{1}{n\sqrt{2}}. \qquad (3.21)$$

Using (3.21), we can write (3.20) as

$$\frac{1}{n\sqrt{2}} + \sqrt{1 - \frac{1}{2n^2}} > \frac{\sqrt{2}}{n}.$$

Thus, $\sqrt{2n^2 - 1} > 1$, which gives $n > 1$; this is what we set out to prove.

Reflecting prisms are widely used as optical components for turning beams of light by a cer-

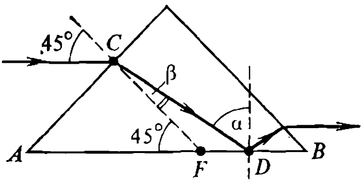


Fig. 3.6.

tain angle, for reflecting them backwards, or for shifting them sideways. They are used in periscopes, binoculars, photometers, cameras, optical radar and communication systems, laser resonators, etc.

The Lummer-Brodhun Photometer. We will consider the *Lummer-Brodhun photometer* by way of example. The essential feature of this photometer is a system of two glass right-angle prisms. The prisms are cemented together to form a glass cube (see Fig. 3.7). The outer portion of the base

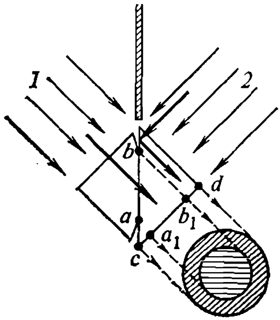


Fig. 3.7.

of one prism is ground away, so it only comes into optical contact with the other prism along the central part of the hypotenuse (section *ab* in Fig. 3.7). When passing through this area of contact, light is neither reflected nor refracted. Where the prisms are not in contact, total internal reflection takes place. The figures *1* and *2* designate beams of light whose intensities have to be compared to each other. It is assumed that the intensities of both beams are constant in cross section. The observer views the plane *cd*. On the central section a_1b_1 of this face, the observer sees the light from the beam *1*, and on the ring-shaped section (between c and a_1 as well as between b_1 and d) he observes the light from the beam *2*. The observer will perceive that the illumination of the central section will be different from that of the ring-shaped area if the intensities of the two beams are not the same.

The Reflecting Prism as a Component of a Laser Resonator. Let us have another example.

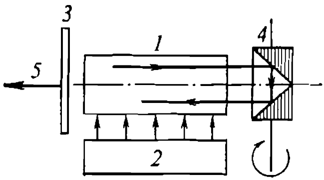


Fig. 3.8.

One of the mirrors of a *laser resonator* can be replaced by a reflecting prism. Figure 3.8 shows a diagram of this kind of laser. In the diagram, *1* designates the active element, *2* is the excitation system of the laser designed to excite the centres in the active element which generate the laser radiation (when the transition from the excited to the nonexcited, i.e. ground, state occurs), *3* is the resonator output mirror, *4* is the reflecting prism acting as the second mirror of the resonator, and *5* is the laser beam. The reflecting prism is used in laser resonators when it is required to obtain radiation in the form of separate short intense light pulses. The prism is quickly rotated around the axis perpendicular to the axis of the resonator. In every position of the prism, except the one shown in the figure, the prism does not return the radiation to the active element. During these periods the prism allows considerable losses to occur which is the reason why no laser radiation is generated. The number of active centres which are excited increases as the excitation energy is supplied to the active element until the instant the rotating prism is oriented as shown in the figure. The generation then avalanches, and the excited active centres are de-excited simultaneously to produce a powerful short-duration light pulse. As the prism is rotating through a new revolution, energy is accumulated in the active element for another light pulse which then occurs when the prism is once again oriented as shown in the figure. When the prism is rotating at the order of 1000 revolutions per second, light pulses 10^{-7} s long are produced. The peak power of such a pulse can reach 10^7 W.

Biprism. To conclude this chapter, we shall consider *a biprism.* A biprism is composed of two right-angle prisms

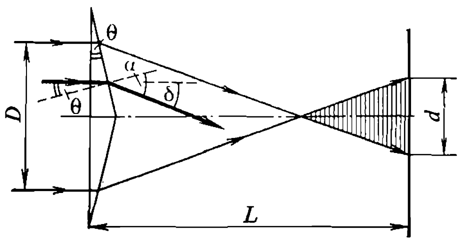


Fig. 3.9.

having small refracting angles and fixed together. More often than not a biprism is made of a single piece of glass. We set a problem as follows. *A beam of light is normally incident on a biprism whose refracting angle is* θ *and whose refractive index is* n*, the aperture of the beam being* D*. Determine the distance* L *from the biprism to the screen on which a shadow band with the width* d *appears in the centre* (see Fig. 3.9).

We denote $(D + d)/2 = a$ and introduce the angle of refraction of the beam δ. It is evident from the figure that $\tan \delta = a/L$. Note that at very small values of θ the angle δ is also small enough for us to use the approximate relation $\tan \delta = \delta$. Therefore,

$$L = \frac{a}{\delta}. \tag{3.22}$$

The law of refraction is written as $\sin \alpha / \sin \theta = n$, or $\sin (\delta + \theta)/\sin \theta = n$. Since δ and θ are very small, we can rewrite the last equation as $(\delta + \theta)/\theta = n$. Therefore,

$$\delta = \theta (n - 1). \tag{3.23}$$

Substituting (3.23) into (3.22), we get $L = a/\theta (n - 1)$. And, finally, we have $L = (D + d)/2\theta (n - 1)$. For $\theta = 0.02$ (which corresponds to $1°10'$), $n = 1.5$, $D = 1$ cm, and $d = 0.5$ cm, we get $L = 75$ cm.

Given biprism, it is quite easy to check the result we obtained for the problem. If a slightly diverging beam of sunlight is incident on a biprism, we would notice that the edges of the shadow band on the screen are rainbowed. The appearance of a coloured strip at the boundary of the dark and illuminated areas on the screen is evidently a result of the dispersion of sunlight. This phenomenon is of particular interest and must be a subject of special discussion.

Chapter Four

Prisms and dispersion of light

Dispersion of Light. On a sunny day admit a narrow beam of sunlight into a darkened room through a small hole in the curtain. The beam will form a bright spot on the opposite wall. If the beam is

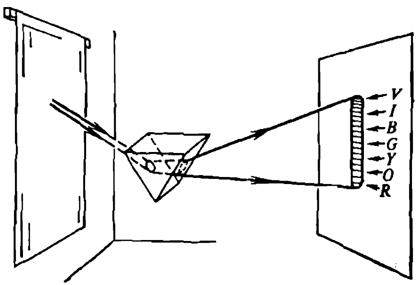


Fig. 4.1.

passed through a glass prism, a coloured band of light will appear on the wall and will be made up of all the rainbow colours from violet to red (see Fig. 4.1: V—violet, I—indigo, B—blue, G—green, Y—yellow, O—orange, R—red). The phenomenon of sunlight being decomposed into different colours is called the *dispersion of light*. The coloured band shown in Fig. 4.1 is termed the *spectrum*.

The First Experiments with Prisms. Theories Concerning the Nature of Colours, Which Preceded Newton's Ideas. The experiment described above is actually an ancient one. As far back as the 1st century A. D., it was known that big single crystals (hexagonal prisms produced by Nature itself) have the property of decomposing light into colours. The English scientist Thomas Harriot (1560-1621) was the first to investigate the dispersion of light experimentally using a triangular glass prism. Similar experiments were conducted by the Bohemian naturalist Marci of Kronland (Johannes Marcus) (1595-1667) who discovered that each colour corresponds to a specific angle of refraction. However, until Newton's time the observations were not properly analysed, while the conclusions that were drawn were not supported by additional experiments. For this reason the ideas that dominated in those times incorrectly interpreted the nature of colour.

Perhaps the most interesting ideas were those in Aristotle's (4th century B.C.) theory of colour. Aristotle claimed that the differences in colour were determined by the different quantities of darkness "admixed" to sunlight. According to Aristotle, violet appears when the most darkness is added to light, while the least darkness produces red. Thus the colours of the rainbow were compounds, whilst white was the basic one. Curiously enough, the appearance of glass prisms and the first experiments on the dispersion of light by the prisms did not shake Aristotle's theory and both Harriot and Marci adhered to it. This is hardly surprising because at first sight the decomposition of light into different colours by a prism seemed to confirm the idea that the colours resulted from a mixture of light and darkness. Hopefully, the readers remembered that in the experiment with a biprism we described at the end of the last chapter the multicoloured strip appears between the dark and illuminated areas, i.e. at the boundary of darkness and white light. The fact that the violet component travels further inside the prism than those of the other colours can easily lead to a false conclusion that violet appears when white light loses most of its "whiteness" as it passes through the prism. In other words, when the path is longest the greatest quantity of darkness is admixed to the white light.

The falsity of these conclusions could have been revealed had the appropriate experiments been carried on using prisms. However, nobody took the trouble to do them before Newton.

Newton's Experiments with Prisms: Newton's Theory of the Nature of Colour. Isaac Newton carried out a *large number* of optical experiments with prisms and described them in detail in his "Optics", "New Theory of Light and Colours" and "Lectures on Optics", the latter first being published after his death. Newton proved convincingly that it was wrong to imagine that the appearance of colour was caused by the admixing of darkness and white light. The experiments he conducted enabled him to assert that a colour hue cannot be made by combining white light and darkness; only various shades of dark occur; the amount of light does not alter the colour. Newton showed that white is not the basic light, it has to be regarded as a combination of different colours (according to Newton, it is "nonhomogeneous", and in modern terms it is "nonmonochromatic"). In reality, different "homogeneous" or "monochromatic" rays are the basic components of white light. The appearance of colours in the experiments with prisms is the result of the decomposition of the compound (white) light into its basic components (into different colours). The decomposition is caused by the fact that each colour is refracted to a different degree. These were the main conclusions drawn by Isaac Newton. They are in full agreement with modern scientific ideas.

Newton's optical research is of considerable value both because it yielded outstanding results and provided an arsenal of methods. Newton's methods of experimenting with prisms (in particular, the method of "crossed prisms") is still used in modern physics.

The aim of Newton's optical research was "not to explain the properties of light by hypotheses, but to propose and prove them by reason and experiments". In order to check an assumption, he usually devised and staged several different experiments. He emphasized that different methods should be used to verify the same idea, for a researcher cannot be too careful he thought.

Consider some of the spectacular experiments with prisms he carried out and the conclusion he arrived at from the results he obtained. Many of the experiments were to verify the relation between the colour of a ray and the degree to which it is refracted (in other words, between colour and refractive index). Here are three experiments oi this kind.

Experiment 1. *Observation of a multicoloured strip through a prism.* Take a paper strip half of which is painted very red and the other painted very blue (Fig. 4.2*a*: *R* is red and *B* is blue). Vfew

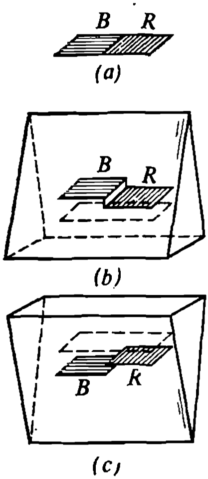


Fig. 4.2.

the strip through a glass prism whose refracting faces are parallel to the strip. Newton wrote: "I found that if the refracting angle of the prism be turned upwards, so that the paper may seem to be lifted upwards by the refraction, its blue half will be lifted higher by the refraction than its red half. But if the refracting angle of the prism be turned downwards, so that the paper may seem to be carried lower by the refraction, its blue half will be carried something lower thereby than its red half". Figure 4.2 illustrates both of the situations described by Newton: Figure 4.2*b* shows what happens when the refracting angle of the prism points up, while in Figure 4.2*c* it points down. Newton goes on to conclude: "... in both cases the light which comes from the blue half of the paper through the prism to the eye does in like circumstances suffer a greater refraction than the light which comes from the red half and by consequence is more refrangible".

Experiment 2. *Passage of light through crossed prisms.* Pass the narrow beam of sunlight coming into a dark room via a narrow opening A through a prism whose refracting edge is horizontal (Fig. 4.3*a*). A spectrum, RV, will appear on the screen, its bottom being red and its top violet. Outline the band on the screen with a pencil. Then set up another prism, its refracting edge vertical (i.e. at right angles to the refracting edge of the first prism), between the first prism and the screen. The beam of light admitted through the slit A passes through the two crossed prisms. A spectrum, $R'V'$, appears on the screen but is displaced with respect to the outline of RV along the x-axis. The violet end of the band is displaced more than the red end is, so the spectrum band turns out bent towards the vertical. Newton concluded that if the experiment with the single prism indicated that rays of light with different refrangibilities correspond to different colours, the experiment with crossed prisms proves that the *reverse* is also true, viz. different colours have different refrangibilities. Hence, the violet component is refracted more than the others in the first prism, and upon its passage through the second prism, its deviation is the greatest. When he analysed the experiment with the crossed prisms, Newton noted that it also proved that the refraction of different light rays obeys the same law, no matter

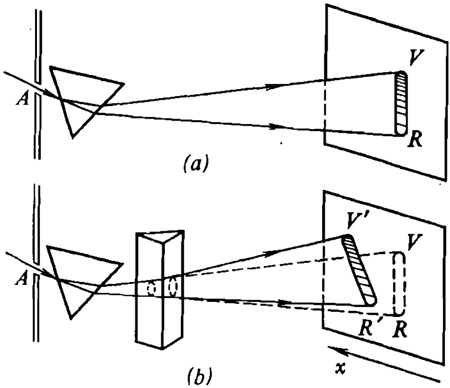


Fig. 4.3.

if they are mixed with rays of different kinds, as in white light, or refracted separately.

Figure 4.4 presents another variant of the crossed prism experiment. Two identical beams of light pass through the prisms. Both beams produce similar spectra on the screen despite the fact that the rays of the same colour (but from different beams) travel along paths of different lengths in the first prism. This actually disproved the assumption that was once made that colour depends on the length of its path inside the prism.

Experiment 3. *Passage of light through a system of two prisms and a reflecting mirror* (Fig. 4.5). A beam of sunlight is admitted through opening

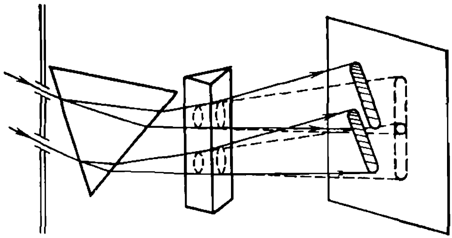


Fig. 4.4.

A, passes through prism *1*, and falls upon mirror *2*. Position the mirror so as to direct only the rays refracted most to prism *3*. After emerging from prism *3*, the refracted rays strike the screen

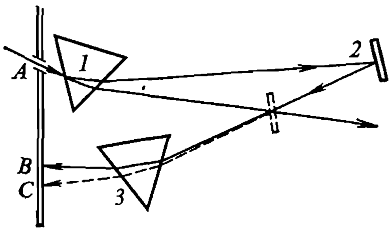


Fig. 4.5.

around the point *B*. Then move mirror *2* so that it will send the least refracted rays to prism *3* (see the dotted line). These rays will be refracted by prism *3* and then will fall upon the screen around the point *C*. It is evident that the rays refracted most of all in the first prism will be most refracted in the second prism, too.

All these experiments led Newton to conclude that the experiments proved that the rays which are refracted differently have different colours, and vice versa, different colours are refracted differently.

Next Newton considered whether it is possible to alter one colour to another by refraction. After conducting a series of convincing experiments, he found he could not. Consider one such experiment.

Experiment 4. *Passage of light through prisms and screens with slits* (Fig. 4.6). A beam of sunlight

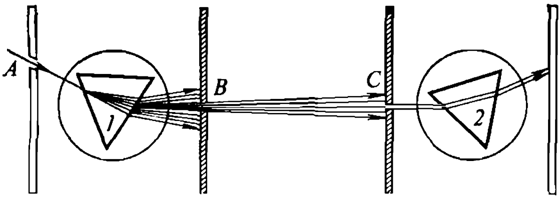


Fig. 4.6.

is decomposed into its constituent colours by prism *1*. Some of the rays of a certain colour pass through slit *B* in a screen positioned behind the prism. These rays subsequently pass through slit *C* in the second screen and fall upon prism *2*. By turning prism *1* and using the screens with the slits, the spectrum can be moved so as to allow any colour to pass through the slits and be refracted by the second prism. The experiment proved that after being refracted by prism *2*, the rays do not alter their colour.

Newton formulated his final conclusion thus: "The species of colour and degree of refrangibility proper to any particular sort of rays, is not mutable by refraction, nor by reflection from natural objects, nor by any other cause, that I could yet observe. When one sort of rays has been well parted from those of other kinds, it has afterwards obstinately retained its colour, notwithstanding my utmost endeavours to change it".

Note that this conclusion can be made on the basis of the crossed prism experiment. If the distance between the prisms is not too small it can be assumed that monochromatic rays whose colour gradually changes along the refraction edge are incident on the second prism. The spectrum observed on the screen shows that the prism does nothing but deflect each of the monochromatic rays without altering its colour.

It is worth mentioning that while investigating the refraction of monochromatic light, Newton designed and built the first *monochromator* of light (a device to isolate a certain wavelength). To collimate the beam of light incident upon the prism, Newton used a collecting lens which he placed between the prism and the perforation admitting the beam of light so that the lens was focused on the perforation. In this case, a slightly diverging (collimated) beam of light fell on the prism (Fig. 4.7). A modern investigator of spectral composition of radiation actually repeats the procedure carried out by Newton. He arranges a prism at an angle of minimum deviation, adjusts and brings the slit of the collimator into focus, just like Newton did in his time, by changing the position of the auxiliary lens.

Newton's experiments on colour mixing are of considerable interest. We shall discuss two of them.

Experiment 5. *Observation of colour mixing with the help of a collecting lens* (Fig. 4.8). As-

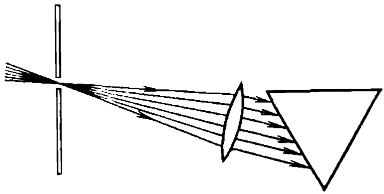


Fig. 4.7.

sume that sunlight from opening *A* passes through a prism and then a collecting lens. The observer puts a sheet of white paper in the way of the rays which have passed through the lens. If he places

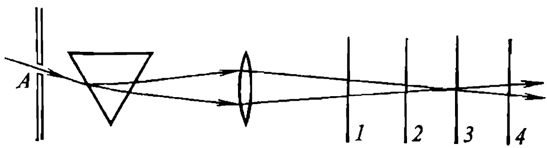


Fig. 4.8.

the sheet of paper in positions *1*, *2*, *3*, and *4*, the observer can see "the colours gradually convene and vanish into whiteness, and afterwards having crossed one another in that place where they compound whiteness, are again dissipated, and severed, and in inverted order retain the same colours, which they had before they entered the composition". Newton also noted that if any colour was intercepted at the lens, the whiteness would be changed into the other colours.

Experiment 6. *Observation of colour mixing when using opposed prisms.* Admit some sunlight

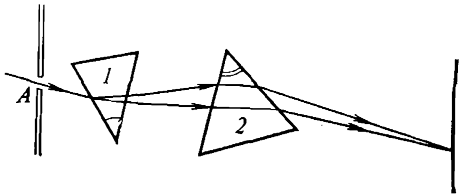


Fig. 4.9.

through two prisms whose refracting edges are opposite each other. In Fig. 4.9 the edge of prism *1* points down, while the edge of prism *2* points up. The refracting angle of prism *2* must be greater than that of prism *1* so that the light will converge. The experiment shows that prism *2* mixes the colours obtained after the sunlight is dispersed into them by prism *1*, and white light appears as a result.

It follows that we can not only decompose sunlight into different colours, we can also carry out the reverse operation and obtain sunlight by mixing the colours. All this led Newton to conclude that "the light of the sun consists of rays differently refractable".

It is appropriate to consider anew the appearance of the multicoloured band at the interface between the light and dark bands in the experiment with the biprism (see the last chapter). The disappearance of the multicoloured band as the screen is moved further away from the prism happens due to the mixture of colours which results in the appearance of white light. A similar explanation applies in modern optics to the appearance of the multicoloured fringe along the outline of an object observed through a prism in polychromatic light.

Summing up, the results of his experiments Newton wrote in "Lectures on Optics" to the effect that sunlight consists of the rays of every colour, not only when it emerges from the prism, but also before it reaches it, before any kind of refraction takes place. Therefore it is no wonder that light is decomposed into colours when a prism is unable to refract rays in a similar way, which can then be collected to produce whiteness with the help of a lens or some other instrument.

The value of Isaac Newton's research can hardly be overestimated. The prominent Soviet physicist Academician Leonid Mandelshtam (1879-1944), who also made a great contribution to the development of optics, noted in his report "Newton's Research in Optics" that "Newton was the first to produce an effective theory of colour on the basis of which he found a great many new interrelated facts, and discovered how to find still more... Geometrical optics before Newton did not deal with any of the problems of colour. Newton proved that the law of refraction holds true for each colour taken separately. Thus, it was due to Newton's discoveries alone that geometrical optics has taken on its present form. A new class of phenomena have now yielded themselves to qualitative investigation". Mandelshtam stressed that "besides their own value, Newton's works have proven to be the turning point of the physical sciences as a whole". It would be no exaggeration to say that before Newton all scientists, Galileo included, proceeded from *a priori* concepts when tackling a physical problem. They used experiment to check or, at best, to correct their concepts. Newton broke free from this tradition. He believed that observation, experiment and generalization were the ways to gain understanding, and that "the best and safest way of philosophising seems to be this: first to search carefully for the properties of things, establishing them by experiments, and then more warily to assert any explanatory hypotheses". Sergei Vavilov (1891-1951), another prominent Soviet physicist, the author of an excellent translation of Newton's "Lectures on Optics" and "Optics" into Russian, wrote: "Unlike his predecessors, even such great ones as Leonardo, Galileo and Gilbert, Newton perceived that rational experiments give answers to certain questions but also pose new questions. In his hands combinations of experiments became as powerful and flexible a method of scientific enquiry as logic and mathematics".

Euler's Work; Correspondence of Colour to Wavelength. Newton's research in optics formed a solid basis upon which the theory of the dispersion of light was developed. It is understood that each colour in the spectrum corresponds to light waves of a definite length. In this connection the work of the famous mathematician Leonhard Euler (1707-1783) is of special interest. According to Vavilov, "in his investigation of the propagation of a light ray Euler was the first to deduce the equation we now take for granted, viz. that of the plane harmonic wave, thus creating the basis of elementary wave optics".

Table

Colour	Wave Length, μm
Violet	0.4–0.45
Indigo	0.45–0.5
Blue	0.5–0.53
Green	0.53–0.57
Yellow	0.57–0.59
Orange	0.59–0.62
Red	0.62–0.75

The table shows the correspondence of different colours to definite wavelengths in air (1 μm = 10^{-6} m). Notice two important features. First, the transition from one colour to another is *continuous* and gradual; each colour corresponds to the wavelengths falling within a certain range rather than one light wavelength. Thus, the range indicated in the table for violet is between 0.4 and 0.45 μm, approximately. We say "approximately" because the boundaries of the colour ranges are not definite. Painters are well aware that a colour can take on various shades, each of which having a different wavelength (or a combination thereof). Strictly speaking, the differentiation between the *seven* colours (violet, indigo, blue, green, yellow, orange and red) is merely conventional and has no scientific foundation.

7—1082

For this reason, when referring to monochromatic light, one should bear in mind *wavelength* (as is done in modern optics) rather than *colour* (as Newton did). After all, the notion of monochromatic light needs clarification. No light ray has a single wavelength; any ray is a combination of waves whose lengths vary within the range from λ to $\lambda + \Delta\lambda$. A ray is considered monochromatic if $\Delta\lambda/\lambda \ll 1$. The smaller $\Delta\lambda/\lambda$ is, the *more monochromatic* the light is. Laser radiation is an example of highly monochromatic light, when $\Delta\lambda/\lambda$ can be of the order of 10^{-6} or smaller.

The second important feature derived from the table is that as we move from the violet part of the spectrum to the red end, the length of light waves gradually increases. The experiments conducted by Newton and other investigators showed that as this transition occurs, the refractive index decreases. This led to the conclusion that the dependence of the refractive index on the length of light waves is described by a *diminishing* function. In other words, *as the wavelength of light increases, the refractive index decreases.*

The Discovery of Anomalous Dispersion; Kundt's Experiments. It was believed until the second half of the 19th century that this conclusion was always true. But in 1860 the French physicist Leroux, while determining the refractive indices of certain substances, suddenly discovered that iodine vapour refractts blue less than it does red. The scientist termed the phenomenon *anomalous dispersion*. In normal dispersion the refractive index decreases with increasing wavelength, whereas in anomalous dispersion the refractive index increases. The German physicist

August Kundt undertook a careful study of this phenomenon between 1871 and 1872. In doing so Kundt used the method of crossed prisms suggested by Isaac Newton.

Figure 4.10*a* shows once again that when light passes through two crossed glass prisms, an inclined

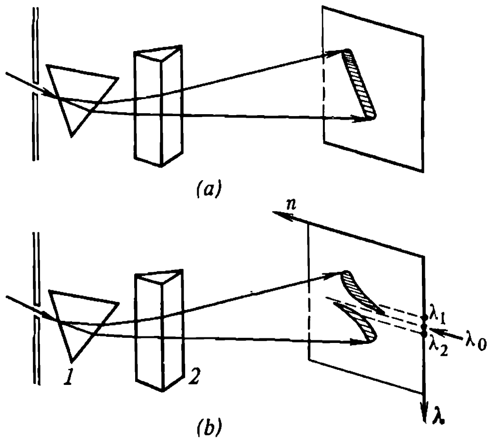


Fig. 4.10.

spectrum appears on the screen. Assume now that one of the glass prisms is replaced by a hollow prismatic cell filled with a solution of the organic compound called cyanine. This is the kind of prism Kundt used in one of his experiments. Figure 4.10*b* shows a diagram of Kundt's experiment; in this figure, *1* is a glass prism and *2* is a prism filled with the solution of cyanine. The glass prism produces normal dispersion. As its refracting edge is pointed down, the axis of wavelengths of the light emerging from the prism also points down (the λ-axis on the screen). The refractive index of the substance filling the second prism is laid off along the n-axis which is perpendicular to the λ-axis. A spectrum appears on the screen which is quite different from the one Newton observed in his experiments. It is clear that $n(\lambda_1) < n(\lambda_2)$ though $\lambda_1 < \lambda_2$. Thus, Kundt

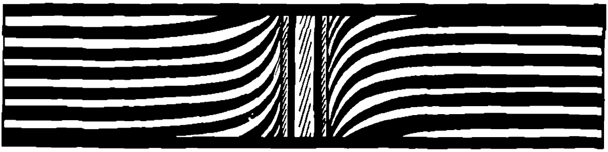


Fig. 4.11.

not only illustrated the phenomenon of anomalous dispersion but also related it to the *absorption* of light in a substance. The wavelength λ_0 indicated in the figure is the one in the vicinity of which light is intensely absorbed by the cyanine solution.

Further investigation of anomalous dispersion showed that the most spectacular results were obtained when a prism and an interferometer are used instead of the two crossed prisms. This method was applied by the Russian physicist Dmitry Rozhdestvensky early in the 20th century. Figure 4.11 shows a photograph he took and illustrates the anomalous dispersion in sodium vapour. Rozhdestvensky improved the experimental technique used at that time and elaborated the "method of hooks" which is common in modern experimental optics.

Our modern ideas are that both normal and anomalous dispersions are the same phenomena and can be described by a *single theory*. This theory is based on the electromagnetic theory of light on the one hand, and on the electronic theory of

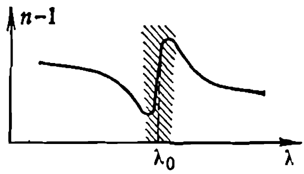


Fig. 4.12.

matter, on the other. In strict terms, "anomalous dispersion" has only a historic meaning. In today's view, normal dispersion is the kind of dispersion which occurs at wavelengths very different from those at which the substance absorbs light, while anomalous dispersion occurs near the absorption bands. Figure 4.12 shows a plot of the refractive index against wavelength for a substance the centre of whose absorption band is at λ_0. Normal dispersion occurs in the nonshaded area, while anomalous dispersion can be observed in the shaded area.

More about Reflecting Prisms. Before dealing with the practical uses made of the dispersion of light when designing prisms and prismatic instruments, we return to the notes we made concerning a reflecting prism we mentioned in the last chapter and which is shown in Fig. 3.5*c*. This kind of prism is called a *Dove prism*. We mentioned that it does not decompose light into colours because the rays leave the prism parallel to each other and that the incident light beam has a definite width. Assume that the light beam is infinitesimally narrow and has only two wavelengths in

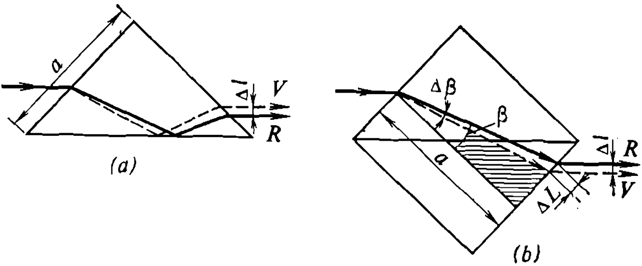


Fig. 4.13.

the red and violet regions of the spectrum. As can be seen from Fig. 4.13*a*, the violet and red rays leaving the prism parallel to each other are displaced with respect to each other by the distance Δl. This displacement is dependent on the length a of the prism edge and on the refractive indices of red and violet rays. It is obvious that in reality a ray always has a certain width which can be designated d. It is clear that an observer can distinguish the red and violet rays that emerge from the prism only if $\Delta l > d$. Otherwise the bundles of rays will overlap and mix.

Consider the following problem. *Find the maximum width d of a bundle of rays that can be resolved at the outlet of a Dove prism, the refractive indices of the different colours* *being* $n = 1.33$ (*red*) *and* $n + \Delta n = 1.34$ (*violet*). *The length of the prism edge* $a = 4$ cm.

The problem can be solved without too much effort if the Dove prism is replaced by a glass cube (see Fig. 4.13*b*). It is fairly clear that this replacement makes the task simpler and at the same time does not alter the problem essentially. We can find the width d from $d = \Delta l = \Delta L/\sqrt{2}$. It follows from the figure that

$$\Delta L = a \tan \beta - a \tan (\beta - \Delta\beta). \tag{4.1}$$

Using the law of refraction, we get the following respective values for the red and violet rays:

$$\sin \beta = \frac{1}{n\sqrt{2}}, \quad \sin (\beta - \Delta\beta) = \frac{1}{(n+\Delta n)\sqrt{2}}. \tag{4.2}$$

Using (4.2) in (4.1), we get

$$\Delta L = \frac{a}{\sqrt{2n^2-1}} - \frac{a}{\sqrt{2(n+\Delta n)^2-1}}$$

$$= \frac{a}{\sqrt{2n^2-1}} \left[1 - \sqrt{\frac{2n^2-1}{2(n+\Delta n)^2-1}}\right].$$

We proceed keeping in mind that $\Delta n \ll n$ and make the following transformations:

$$\sqrt{\frac{2n^2-1}{2(n+\Delta n)^2-1}} = \sqrt{\frac{2n^2-1}{2n^2-1+4n\Delta n}}$$

$$= \sqrt{\frac{1}{1+\Delta n \dfrac{4n}{2n^2-1}}} = \sqrt{1 - \Delta n \frac{4n}{2n^2-1}}$$

$$= 1 - \Delta n \frac{2n}{2n^2-1}.$$

Therefore,

$$\Delta l = \frac{\Delta L}{\sqrt{2}} = \sqrt{2} \quad an \quad \Delta n (2n^2-1)^{-3/2}. \tag{4.3}$$

Substituting the values from the statement of the problem into (4.3), we get $\Delta l = 0.02$ cm. So, the light beam must be narrower than one fifth of a millimetre to enable the observer to resolve the red and the violet emerging from the prism.

Dispersive Prisms. Angular Dispersion. The problem we have just discussed illustrates why we do not observe light dispersion due to refraction in reflecting prisms and plane-parallel plates. Prisms which produce a pronounced dispersion are called *dispersive*. The different colours emerge from dispersive prisms at different angles, which helps to resolve them. Assume that the wavelengths of two columns differ by $\Delta\lambda$, and the angles of deviation of these rays in the prism differ by $\Delta\delta$. The ratio $\Delta\delta/\Delta\lambda$ is termed the *angular dispersion* of a prism. The resolving power of a prism increases as this ratio increases. It can be said that the angular dispersion of reflecting prisms equals zero.

Consider the following problem. *Find an expression for the angular dispersion of a prism whose refracting angle is* θ *if light travels through it symmetrically.* Remember that *when the wavelength changes from* λ to $\lambda - \Delta\lambda$, *the refractive index changes from* n *to* $n + \Delta n$.

Using (3.5), we can write an expression for the refractive index of the ray whose wavelength is $\lambda - \Delta\lambda$, viz.

$$\sin\frac{\delta+\Delta\delta+\theta}{2}=(n+\Delta n)\sin\frac{\theta}{2}, \tag{4.4}$$

and, since $\Delta\delta$ is very small, (4.4) becomes

$$\sin\frac{\delta+\theta}{2}+\frac{\Delta\delta}{2}\cos\frac{\delta+\theta}{2}=n\sin\frac{\theta}{2}+\Delta n\sin\frac{\theta}{2}.$$

With (3.5) this gives us

$$\Delta\delta=\frac{2\sin\frac{\theta}{2}}{\cos\frac{\delta+\theta}{2}}\Delta n=\frac{2\Delta n\sin\frac{\theta}{2}}{\sqrt{1-\sin^2\frac{\delta+\theta}{2}}}$$

$$= \frac{2\Delta n \sin\frac{\theta}{2}}{\sqrt{1-n^2\sin^2\frac{\theta}{2}}}.$$

Therefore,

$$\frac{\Delta\delta}{\Delta\lambda} = \frac{2\sin\frac{\theta}{2}}{\sqrt{1-n^2\sin^2\frac{\theta}{2}}}\,\frac{\Delta n}{\Delta\lambda}. \tag{4.5}$$

More often than not, $\theta = 60°$. In this case the angular dispersion is described by the expression

$$\frac{\Delta\delta}{\Delta\lambda} = \frac{2}{\sqrt{4-n^2}}\,\frac{\Delta n}{\Delta\lambda}. \tag{4.6}$$

Suppose that, like in the previous problem, $n = 1.33$, $\Delta n = 0.01$ and that the refracting angle of the prism is 60°. Using the relation $\Delta\delta = 2\Delta n/\sqrt{4-n^2}$, we find that $\Delta\delta = 0.013$ (or 45′). This means that if the distance between the screen and the prism is 1 m, the relative displacement of the centres of the red and violet rays will be $\Delta l = \Delta\delta\cdot 100$ cm $= 1.3$ cm. It is fairly easy to resolve the colours in this case. The only condition is that the width of the beam of light does not exceed 1 cm.

Spectral Instruments—Monochromators and Spectrometers. The Fuchs-Wordsworth Optical System. Dispersion prisms are widely used in different *special instruments*. These instruments are designed to isolate certain parts of the spectrum (*monochromators*) or to investigate the spectrum of radiation (*spectroscopes*, *spectrographs*, and *spectrometers*). *Spectral analysis* is of considerable scientific and practical value since the spectrum of a gas consists of sets of lines, and each component of the gas has its own spectrum. Hence scientists can determine the chemical composition of the substance from the lines in its spectrum.

Prismatic spectral instruments differ by the type of optical systems used in them. As an example of this, consider the *Fuchs-Wordsworth optical system.* Figure 4.14 shows one of the variants of this system. The base of a dispersion prism having a refracting angle $\theta = 60°$ is fixed rigidly to the reflecting surface of a plane mirror. The prism and the mirror can be rotated about the axis O_1^* so as to change the angle φ between the first surface of the prism and the incident non-monochromatic beam of light travelling along the path MC_1. The beam first passes through the prism and is then reflected by the mirror. Since it is refracted by the prism, the incident beam is split into monochromatic rays that emerge from the prism in different directions. The output collimator N will only let out the monochromatic ray which is reflected by the mirror parallel to the incident beam. We will now verify that this ray passes through the prism *symmetrically* or, to put it differently, we will prove that in the case of symmetric passage of a ray through the prism, the reflected ray GN is parallel to MC_1. Figure 4.14*a* shows the ray in question. Since its passage through the prism is symmetric, $\angle MCO_1 = \angle O_1CG$. As $O_1O \parallel GH$, $\angle O_1CG = \angle CGH$. Besides, $\angle CGH = \angle HGN$ (the angle of incidence equals the angle of reflection). Thus, $\angle MCO_1 = \angle O_1CG = \angle CGH = \angle HGN$. Therefore, $\angle MCG = \angle CGN$, which means that $GN \parallel MC_1$.

* This axis, perpendicular to the plane of the drawing, is labeled by its cross-point O_1 (*transl. note*).

Thus, a light ray whose wavelength is such that it travels inside the prism parallel to its base is reflected by the mirror, gets into the collimator

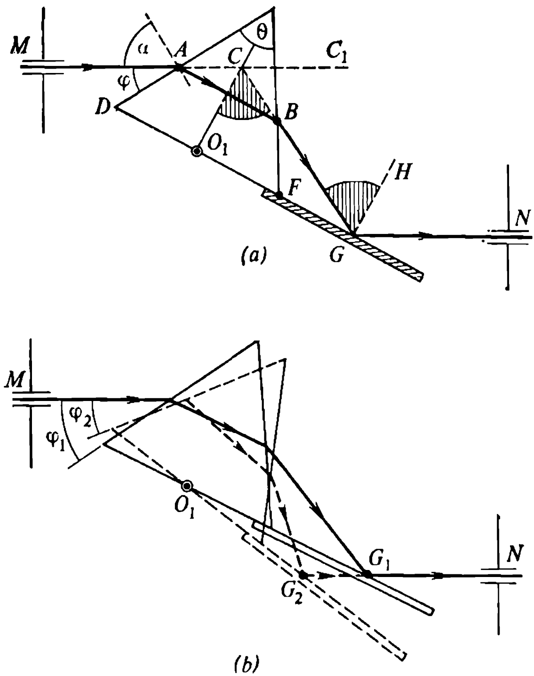


Fig. 4.14.

N and emerges from the instrument. Rays which have different wavelengths will not pass out through the collimator. The wavelength λ of the ray emerging from N is determined by the value

$n(\lambda)$ of the refractive index which satisfies the condition (recall (3.5))

$$n(\lambda) = \frac{\sin\alpha}{\sin(\theta/2)} = 2\cos\varphi. \tag{4.7}$$

If the angle φ is changed by slightly rotating the prism and the mirror about O_1, the slits M and N remaining fixed, condition (4.7) will be true for a different wavelength. This will be the wavelength of the ray which will travel symmetrically through the prism and emerge from the collimator N.

The solid and dotted lines in Fig. 4.14*b* show two positions of the prism and the mirror and the light rays corresponding to each position. The first position (the solid lines) corresponds to the angle $\varphi = \varphi_1$, whereas the second one (the dotted lines) corresponds to the angle $\varphi = \varphi_2$. Since $\varphi_2 < \varphi_1$, according to (4.7) we get $n(\lambda_1) < n(\lambda_2)$, and therefore $\lambda_1 > \lambda_2$. Rotating the prism and the mirror about O_1, we can, at the outlet N, obtain a monochromatic beam having one of the wavelengths in the range of the wavelengths characteristic of the incident light.

Note that in the system described above, as in the other kinds of practical optical trains, the incident beam must be collimated. Another necessary condition is that only a beam reflected in a specific manner by the mirror can leave the instrument. In order to do this, concave spherical mirrors as shown in Fig. 4.15 are used, where *1* is the entrance slit, *2* is a concave spherical mirror whose focus coincides with the entrance slit (this mirror shapes the directed collimated beam which then falls onto the prism), *3* is a prism, *4* is a flat mirror, and *5* is a concave spherical mirror which reflects the beam, *6*, to the exit slit *7* (the mirror focus coincides with the exit slit).

It is quite clear that the system discussed above is only one of the optical trains used in pris-

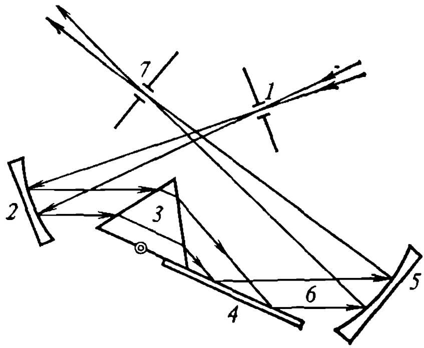


Fig. 4.15.

matic spectral instruments. In fact, multiprismatic optical systems which are systems combining prisms, lenses and other elements are used.

Goethe Versus Newton. As the chapter draws to an end, we return once again to Newton to tell the reader about the opposition of Goethe to Newton. A great German poet, Johann Wolfgang Goethe (1749-1832) took an immense interest in the theory of colour and even wrote a book on the topic. He did not, however, share Newton's scientific views nor did he agree with his conclusions about the dispersion of sunlight into colours. Without taking the trouble to get to the heart of Newton's experiments and doing them himself, he simply brushed Newton's theory of colour aside.

He wrote: "Newton's statements are a formidable assumption. How can any one expect white, the purest and the most transparent of colours, to be a mixture of other colours?" Goethe went on to point out that no artists had yet obtained white paint by mixing all the different coloured pigments; they invariably got only dirty shades. Goethe reproached Newton for his failure to explain the blue of the sky. Everything, Goethe claimed, was evident before Newton's theory: the black night sky becomes blue after the appearance of the sun because of the mixture of the white light of the sun and the black of the sky. "If we assume Newton's point of view," he went on with his reasoning, "and hold that the blue is independent, we should have to explain the blue of the sky by proving that the air itself is blue. But then it becomes impossible to account for the pink of distant mountains peaks and for the red of the setting sun."

Certainly, Goethe and Newton were people with different approaches, they treated Nature in different ways and tried to understand its laws differently. Newton had a brilliant analytical mind; each step forward was based on experiment and careful calculation. He was a meticulous investigator who never allowed either himself or other people "to mingle conjectures with certainties". Unlike Newton, Goethe was far more an inspired dreamer and philosopher rather than a physicist. He perceived the world as a united inseparable whole, preferring imagination, speculation, and inspiration to experiment and calculation.

It is hardly surprising, therefore, that Goethe did not understand Newton or agree with him, and hence Goethe was certainly wrong to criticise Newton's conclusions. In this controversy between these two great personalities, we definitely side with Newton. Nevertheless, we cannot wave Goethe's remarks aside, today more than ever before, because no matter what Goethe's intentions were, his observations have a rational kernel.

Essentially, these observations reduce to the statement that the properties of light Newton discovered in the course of his experiments were not the properties of natural light, but those of light "tortured by all kinds of instruments such as slits, prisms and lenses". This was noted by Academician Mandelshtam who, according to Professor Gorelik of Moscow University, "saw in these words an anticipation, though naive and one-sided, of today's point of view of the role of measuring instruments". Twentieth century physics has a most interesting branch called quantum mechanics. This discipline investigates the behaviour of very small objects and has shown that any measurement conducted in a microcosm inevitably leads to a distortion of the object measured. It appears, therefore, that when Man investigates Nature at the level of microphenomena, he inevitably causes irreversible effects. Thus, when we measure the momentum of an electron we cause distortions which make it impossible to measure the electron's coordinate at the same time. And, the reverse is also true: when an electron's coordinate is being measured, the distortions caused make it impossible to measure its momentum simultaneously.

This anticipation of the problems of modern physics is the rationale behind Goethe's observations, observations we cannot afford to ignore today. However, Goethe's criticism was aimed at the wrong person and cannot diminish the value of Newton's contribution to physics. Nevertheless, the philosophical problems the criticism brings out are, without any doubt, interesting and relevant today. The problems were best formulated by Goethe not in his "Theory of Colour" but in his immortal "Faust":

"Who traces life and seeks to give
Descriptions of the things that live
Begins with "killing to dissect",
He gets the pieces to inspect,
The lifeless limbs beneath his knife
All parts—but link which gave them life!"

As for Goethe's reproaching Newton about his inability to explain why the colour of the sky is blue and that of the setting sun is red, it can hardly be argued with. True, the refraction and dispersion of light do not account for these phenomena. The answer was found much later, when the *scattering of light* by the molecules of air was studied by the branch of physics known as molecular optics which was advanced considerably by Academician Mandelshtam. But this goes beyond the scope of this chapter.

Chapter Five

The rainbow

There is hardly a person who has not admired a rainbow. The beauty of a rainbow has always made a striking impression on observers from ancient times. It was considered to be a good sign having magic powers. According to an old English superstition a crock of gold is to be found at the foot of every rainbow.

People nowadays know that magic powers of the rainbow only exist in fairy tales. In reality, the rainbow is an optical phenomenon connected with the refraction of light in myriads of floating raindrops. However, not many people are aware that it is due to the refraction of light by water droplets that the gigantic many-hued bow appears across the skies. For this reason, it is worthwhile to dwell on a physical interpretation of this spectacular optical phenomenon.

The Rainbow to the Eyes of a Careful Observer. First of all, it has to be noted that a rainbow can be observed only with one's back to the sun. It appears when the sun shines upon a veil of rain. As the rain subsides and then stops, the rainbow fades away. The colours of the rainbow follow one another in the same sequence they do in the spectrum resulting from the passage of a beam of light through a prism. The lower part of the bow is violet and the upper part is red. Sometimes, a larger and somewhat blurred "secondary" rainbow appears above the principal one. The order of the colours in it is reversed, the inner part being red and the lower part violet.

An observer standing on a plane surface on the Earth can see the rainbow if the altitude of the sun, i.e. its angular elevation above the horizon, is under 42°. The lower the sun the higher the rainbow and, consequently, the greater its visible part. A secondary rainbow appears if the sun's altitude is not more than 52°.

The rainbow seems to be a kind of huge wheel on an imaginary axis passing through the sun and the observer's eye. In Fig. 5.1 this is the straight line OO_1; point O designates the observer, and OCD is the plane of the Earth's surface, while $\angle AOO_1 = \psi$ is the altitude of the sun. In order to find $\tan\psi$, it is sufficient to divide the observer's height by the length of his shadow. Point O_1 is termed the anti-solar point and lies below the horizon line CD. The figure shows that the rainbow is the circle confining the base of a cone, OO_1 being its axis and γ being the angle the axis of the cone makes with its generators (the apex of the cone). It stands to reason that the observer cannot see the whole circle but only that part of it which is above the horizon (CBD in Fig. 5.1). Note that $\angle AOB = \Phi$ is the angle at which the top of the bow is visible to the observer, while $\angle AOD = \alpha$ is half the angular distance between the rainbow's feet at one of which, according to the myth, a crock of gold can be found. It is evident that

$$\Phi + \psi = \gamma. \tag{5.1}$$

Thus, the position of the bow with respect to the landscape depends on the position of the observer with respect to the sun, and the angular dimensions of the rainbow are determined by the

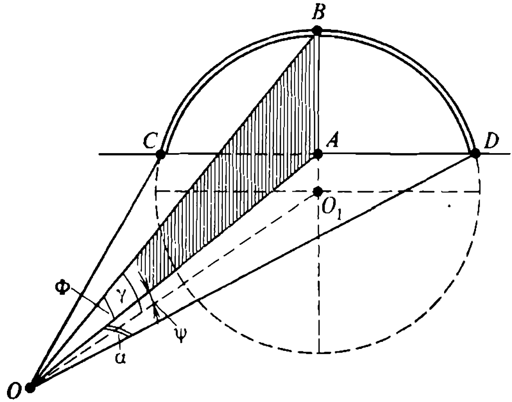


Fig. 5.1.

sun's altitude. The observer is at the vertex of a cone whose axis goes along the line connecting the observer and the sun; the bow is a part of the circle confining the base of the cone which is above the horizon. When the observer moves, the cone and therefore the rainbow itself moves, that is why there is no point in chasing the promised crock of gold.

Two things have to be clarified. Firstly, when we say, the straight line connecting the observer and the sun, we mean the apparent direction of the sun and not the real one. It differs from the real direction by the angle of refraction. Secondly, when we say the rainbow is above the horizon, we mean a relatively distant rainbow, when the veil of rain is several miles away from us. One can also observe a rainbow quite nearby, for example, a rainbow formed by the spray of a big fountain. In this case the ends of the bow seem to go down into the ground. The distance between the rainbow and the observer has no tangible effect on the angular dimensions of the bow.

It follows from (5.1) that $\Phi = \gamma - \psi$. For the primary rainbow the angle γ is about 42° (for the yellow hues of the rainbow), while for the secondary rainbow the respective angle is 52°. This explains why the observer cannot admire the primary rainbow if the sun's altitude exceeds 42°. He will not see the secondary rainbow either when the sun's altitude is over 52°. If an observer is flying in an aircraft, the remarks relating to the sun's altitude have to be reconsidered. It is worth mentioning that an observer in the aircraft can see a rainbow in the shape of a full circle. However, no matter where the observer is positioned (on the Earth's surface or above it) he will invariably be at the vertex of a cone pointed towards the sun and having an opening of 42° in the case of the primary rainbow or 52° in the case of the secondary rainbow. Why must the angles be 42° and 52°? This question will be answered a little later.

Consider the following problem. *Find the angles at which the height and half the rainbow* (the angles Φ and α in Fig. 5.1) *can be seen, if the sun's altitude is* $\psi = 20°$. Using (5.1) we obtain $\Phi = \gamma - \psi = 42° - 20° = 22°$. In order to determine the angle α, we turn to Fig. 5.1. From the triangles BOO_1 and AOO_1, we find that $OO_1/OA = \cos\psi$. It is clear that $OA/OD = \cos\alpha$. Since

$$\frac{OO_1}{OB} = \frac{OO_1}{OA}\cdot\frac{OA}{OB} = \frac{OO_1}{OA}\cdot\frac{OA}{OD},$$

$\cos\gamma = \cos\psi\cos\alpha$. Therefore,

$$\cos\alpha = \frac{\cos\gamma}{\cos\psi} = \frac{\cos 42^\circ}{\cos 20^\circ} = 0.79,$$

which gives us $\alpha = 38^\circ$.

Interpretations of the Origin of Rainbow: from Fleischer to Newton. Ever since people have

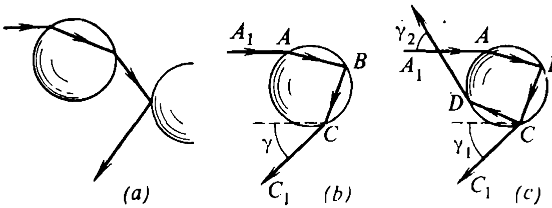


Fig. 5.2.

admitted rainbows, they have tried to understand what caused them. In 1571 Fleischer from Breslau (now Wroclaw, Poland) published a paper in which he asserted that the observer sees the rainbow when he perceives light that has been doubly refracted in one raindrop and subsequently reflected from another (Fig. 5.2*a*). Later, the Italian Marcus Antonius De Dominis (1566-1624) suggested another interpretation of the passage of light to the observer, which was, incidentally, correct. He believed that each of the light rays forming the rainbow is doubly refracted and reflected in the same raindrop (Fig. 5.2*b*). A ray A_1A of sunlight enters the drop, is refracted at A and is then reflected at B, and finally, leaves the drop being refracted at C. The observer sees the ray CC_1. It is angled at γ to the ray A_1A as a result of which the observer sees the rainbow at angle γ to the direction of the incident sunlight.

Rene Descartes continued where Dominis had left it and accounted for the appearance of the secondary rainbow. He held the sunlight is both refracted and reflected at each of the points A, B, and C (see Fig. 5.2*b*). True, the rays reflected at A as well as the one refracted at B do not form the rainbow and are therefore of no interest. As for the ray reflected at C, it may, upon being refracted at the point D, leave the drop and proceed to form another rainbow (Fig. 5.2*c*). While the observer sees the primary rainbow at the angle $\gamma_1 = 42°$, he will see the secondary one at $\gamma_2 = 52°$. Naturally, the secondary rainbow looks paler than the primary one for part of the energy of ray CD is lost when the ray is reflected at D.

However, neither Dominis nor Descartes were able to explain why an observer sees the rainbow at the angle of 42° (or 52°). Furthermore, they failed to account for the colours of the rainbow. Thus Dominis supposed that rays of sunlight which travel along the shortest path inside a raindrop and, therefore, mix with the darkness the least to produce the red. On the other hand, the rays covering the longest path inside a raindrop mix with the darkness the most to produce the violet. The reader knows from the previous chapter about these pre-Newtonian ideas of the origin of colour.

Newton's Explanation of the Origin of Rain-

bow in His "Lectures on Optics". Newton's theory of colour allowed him to explain the origin of the rainbow. His "Lectures on Optics" contain a passage which accounts for the factors causing the rainbow, the idea of the passage being as follows. The rays which enter a raindrop leave it after one reflection, but some of them leave after two reflections, and there are some rays which leave a drop after three or more reflections. Since raindrops are very small with respect to the distance between it and the observer, they can be regarded as points and there is no need to take their sizes into consideration. Only the angles made by the incident and the emergent rays need be considered. *The emergent rays are the most condensed where these angles are the greatest or the least. Since different sorts of rays have different greatest and least angles, the rays which have the greatest density at some place tend to manifest their own colour.*

The gist of the information is expressed in the italics. The information needs certain clarification. As a matter of fact, the rest of the chapter will be devoted to this purpose.

Passage of a Light Ray through a Raindrop. Assume that all the rays falling on a raindrop have the same wavelength. This means that at first we shall consider only the refraction (and reflection) of the rays in the raindrop and disregard the dispersion of the light. Let a parallel beam of monochromatic light fall on a drop with radius R. We will call the ratio $\xi=\rho/R$ the impact parameter. Here ρ is the distance from the ray to the straight line that is parallel to it and passes through the centre of the raindrop. Since the drop is symmetric, every ray with the same impact parameter (they are shown in Fig. 5.3) follows a similar path and leaves the drop at the same angle to the initial direction. The spherical symmetry of the drop causes every ray's path to lie in the same plane, the plane passing through the ray in qu-

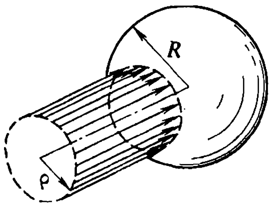


Fig. 5.3.

estion and the centre of the drop. That is why we can consider a two-dimensional problem, tracing the paths of light rays in this plane which will be the same as the plane of the figure.

Figure 5.4 shows the path of a light ray with the impact parameter ρ/R. We designate the angle of incidence of the ray on the drop α. It is obvious that $\sin\alpha = \rho/R = \xi$. Since the triangles AOB and BOC are isosceles, $\angle OAB = \angle ABO$ and $\angle OBC = \angle BCO$. The angle of incidence equals the angle of refraction, therefore, $\angle ABO = \angle OBC$. We call all these angles β (see Fig. 5.4) and designate the angle between the incident and the emergent rays as γ. As the path of the ray is symmetric about the line OO_1, $\angle OO_1C = \gamma/2$. Draw the line MN parallel to OO_1 passing through the point C. It is evident that $\angle MCC_1 = \angle O_1CN = \angle OO_1C = \gamma/2$. Bear in mind that $\angle C_1CP = \alpha$ and $\angle QCP = \beta$.

Since MN is parallel to OO_1, $\angle MCQ = \angle OBC = \beta$. Finally, we get $\angle MCC_1 = \angle MCQ - (\angle C_1CP - \angle QCP) = \beta - (\alpha - \beta)$. Thus, we have $\gamma/2 = \beta - (\alpha - \beta)$ or, put differently,

$$\beta = \frac{\gamma + 2\alpha}{4}. \tag{5.2}$$

We now express the angle γ in terms of the impact parameter ξ of the ray. According to the law

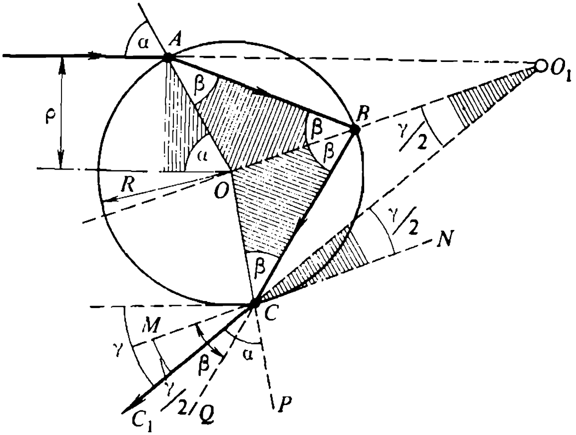


Fig. 5.4.

of refraction, $\sin\alpha/\sin\beta = n$ at A. Using (5.2), we have

$$\sin\frac{\gamma + 2\alpha}{4} = \frac{\sin\alpha}{n}, \tag{5.3}$$

or, in other words,

$$\frac{\gamma + 2\alpha}{4} = \arcsin\left(\frac{\sin\alpha}{n}\right).$$

Thus,

$$\gamma = 4 \arcsin\left(\frac{\sin\alpha}{n}\right) - 2\alpha, \tag{5.4}$$

or bearing in mind that $\sin\alpha = \rho/R = \xi$, we get

$$\gamma = 4 \arcsin\left(\frac{\xi}{n}\right) - 2 \arcsin \xi. \tag{5.5}$$

Consider the following problem. *At which values of the impact parameter will the emergent ray be parallel to a ray incident on the droplet?* In other words, we have to determine the values of ξ at which $\gamma = 0$. Assuming $\gamma = 0$ in (5.5), we get $2 \arcsin(\xi/n) = \arcsin \xi$, or $\sin[2 \arcsin(\xi/n)] = \xi$. Since $\sin 2\nu = 2 \sin\nu \sqrt{1 - \sin^2\nu}$ we can write

$$2\frac{\xi}{n}\sqrt{1-\left(\frac{\xi}{n}\right)^2} = \xi. \tag{5.6}$$

Equation (5.6) has two roots. The first one is obviously $\xi_1 = 0$. The second root is

$$\xi_2 = \frac{n}{2}\sqrt{4-n^2}. \tag{5.7}$$

Substituting $n = 4/3$ into (5.7), we have $\xi = 0.994$. Note that the commonly used value of the refractive index of water $n = 4/3$ corresponds to rays in the yellow part of the spectrum.

The Greatest Angle between the Incident and the Emergent Rays. As the impact parameter increases from 0 to 1, the angle γ increases from 0 to a certain maximum value and then decreases to 0 at $\xi = 0.994$ (for yellow rays). It is most important for us to find the maximum value of the angle γ because, as Newton noted, the emergent rays are at their most intense when these angles are either at a maximum or a minimum.

Consider the following problem. *Find the maximum value of the angle between the rays falling on the droplet and those leaving it. At what value of the impact parameter is this angle realized? The refractive index is 4/3 (yellow rays).* Using (5.5), we differentiate the function $\gamma(\xi)$ and equate the derivative to zero:

$$\frac{d\gamma}{d\xi}=\frac{4}{n\sqrt{1-(\xi/n)^2}}-\frac{2}{\sqrt{1-\xi^2}}=0.$$

From this we have

$$\xi=\sqrt{\frac{4-n^2}{3}}. \tag{5.8}$$

At $n = 4/3$, we get $\xi = 0.861$. Substituting (5.8) into (5.5), we find the expression for the maximum angle between the rays incident on the drop and emerging from it:

$$\gamma_{max}=4\arcsin\left(\frac{1}{n}\sqrt{\frac{4-n^2}{3}}\right)-2\arcsin\sqrt{\frac{4-n^2}{3}}. \tag{5.9}$$

When $n = 4/3$, we have $\gamma_{max} = 42°02'$.

Figure 5.5 shows the dependence of the angle γ on the impact parameter ξ (for yellow rays).

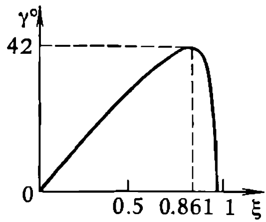


Fig. 5.5.

Rays with impact parameters varying from 0 to 1 are incident on every raindrop. They leave the drop at different angles γ. The observer will see those rays which diverge the *least* as the brightest. These are the rays falling within the area of the apex of the curved line shown in Fig. 5.5, i.e. the rays for which $\gamma = 42°$. According to Newton, these are the rays which are the most intense.

Figure 5.6 illustrates the intensity of the emergent rays when $\gamma = 42°$ or so in a spectacular way. It shows the paths of rays having different impact parameters (the paths corresponding to $n = 4/3$).

It is now clear why the rainbow looks like an arc observed at the angle of 42° to the straight line connecting the observer and the sun. For the sake of simplicity, assume that the sun stands close to the horizon and the veil of rain is a plane wall perpendicular to the direction of the incident rays. Figure 5.7 shows a view of this as cut by the plane of the Earth's surface. Here MN is the rain front, O represents the observer, and O_1 is the anti-solar point. The rays which were reflected once and refracted twice in the drops of rain lie within the shaded area, rays outside this area do not reach the observer. The rays coming to the observer from raindrops to the right of C and to the left of D are much weaker because they greatly diverge. The rays reaching the observer from the boundary of the shaded area are brightest, i.e. from the drops close to C and D, since the divergence of these rays is least. So, if the whole spectrum of the sun consisted of rays of the same wavelength, the rainbow would appear to the observer to be a narrow luminous arc.

Sequence of Colours in Primary and Secon-

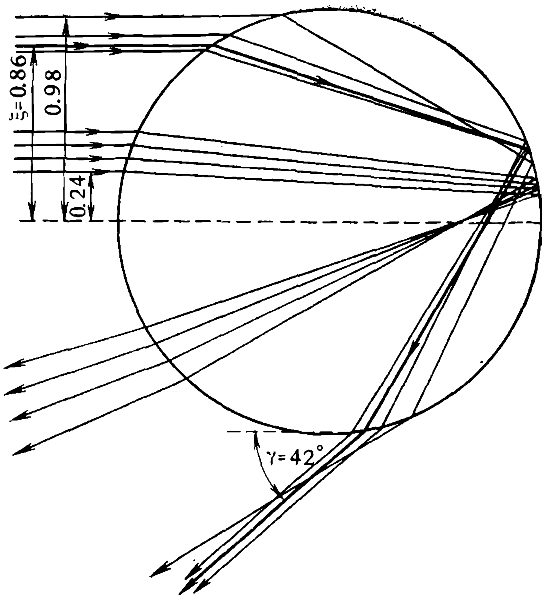


Fig. 5.6.

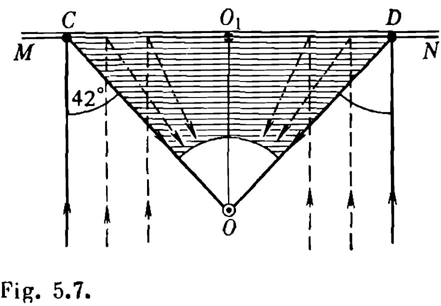


Fig. 5.7.

dary Rainbows. In reality the sun's spectrum consists of light of different wavelengths. That is why the rainbow is so colourful. From now on, we shall take into account that sunlight is non-monochromatic.

For the sake of simplicity, consider only two wavelengths, the respective indices of refraction being $n_r = 1.331$ (red) and $n_v = 1.344$ (violet). Substituting n_r and n_v into (5.8) and (5.9), we have

for the red $\xi_r = 0.862, \quad \gamma_r = 42°22'$;

for the violet $\xi_v = 0.855, \quad \gamma_v = 40°36'$.

Figure 5.8 shows the paths of the red and violet components when each of them makes the great-

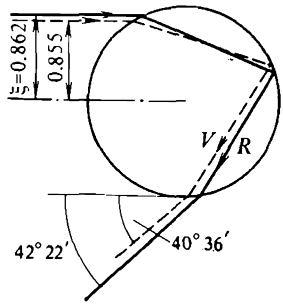


Fig. 5.8.

est angle with the initial direction when emerging from the droplet.

Thus, the values of the maximum angle between the emergent and the incident rays are different for rays of different wavelengths. Newton pointed out in this connection that since different kinds of rays make different maximum angles, the rays most concentrated in certain areas tend to display colours of their own. The observer will see the red arc at an angle of 42°20′, while the violet arc will appear at 40°40′ (recall that these are the angles between the directions from

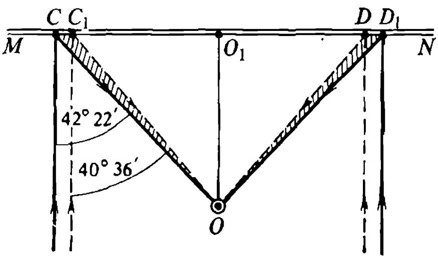


Fig. 5.9.

the observer to the rainbow and to the anti-solar point). This explains why the red is on the outside and the violet on the inside of the bow.

Note another thing connected with the hues of the rainbow. Figure 5.9 shows a picture which is similar to that in Fig. 5.7 (the number of colours under consideration remains two). Intense red light comes to the observer along CO and DO, while violet does not travel along them. Quite intense violet, as well as red light weakened due to divergence, reaches the observer after travelling along C_1O and D_1O. The observer will see the violet with a slight admixture of the red. Weakened (diverging) red and violet rays will come to the observer from the points lying between C_1 and D_1. They will mix together to produce, along with the other colours, white light.

Thus, the appearance of a many-hued arc in the sky is accounted for not only by the fact that each colour corresponds to a specific value of the maximum angle γ, but also by the fact that close to this angle the overlapping, or mixing of colours occurs the least. Hence, another conclusion can be made: the red hues of the rainbow look more vivid and intense whereas the violet, through the reddish tinge, is relatively dull. Note that prisms make it possible to obtain purer colours than those in a rainbow. One could say that the spectrum of the rainbow looks like that of a prism regarded through a transparent glass slightly tinted red.

So far, we have been examining the primary rainbow. Similar reasoning can be applied to the secondary rainbow. We should bear in mind, however, that the secondary rainbow is caused by double refraction and double reflection of light beams in a raindrop (see 5.2*c*). The maximum angle between the paths of the incident and the emergent rays can be proved to equal approximately 52° in this case. It seems calculations can be omitted here. But it will be appropriate to examine Fig. 5.10 which explains why the order of colours in the secondary rainbow is reverse to that in the primary one.

Rainbows on Other Planets. The attentive reader ought to have seen by now how untrue the common misconception that "rainbows are a very simple thing—in fact, nothing but sunlight refracted in raindrops" is. Having understood its real cause, we can give way to our imagination. Try to answer the following question: what would a rainbow look like if the refractive index suddenly increased 1.25 times for all wavelengths?

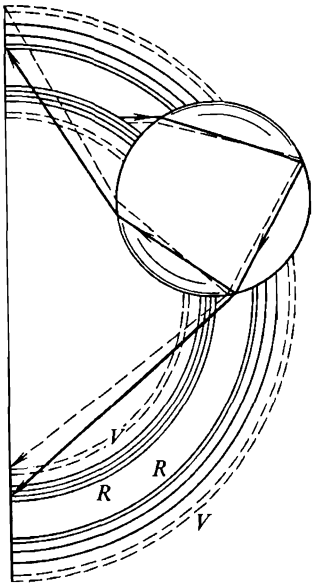


Fig. 5.10.

(Imagine yourself on a planet where the role of the water is performed by another liquid.) This means that now $n_r = 1.66$ for red, while $n_v = 1.68$ for violet. Using (5.9), we have in this case $\gamma_r = 11°$ and $\gamma_v = 10°$. Thus, the angular dimensions of the rainbow have decreased by four times. The rainbow will only be visible when the sun's altitude does not exceed 10°. If the refractive index approaches $n = 2$, the rainbow will turn into a bright spot visible in the direction of the anti-solar point.

Factors Causing Halos. Halo and Rainbow. To conclude this chapter, we shall consider another optical phenomenon—the *halo*. Unlike the rainbow, it is not at all common. Indeed, not many readers will have heard of it. Here is an extract from Minnaert's book "Light and Colour in the Open Air": "After a few days of fine bright spring weather the barometer falls and a south wind begins to blow. High clouds, fragile and feathery, rise out of the west, the sky gradually becomes milky white.... The sun seems to shine through opal glass, its outline no longer sharp but merging into its surroundings. There is a peculiar, uncertain light over the landscape; I feel that there must be a halo round the sun! And as a rule I am right.

A bright ring with a radius of rather more than 22° can be seen surrounding the sun; the best way to see it is to stand in the shade of a house or to hold your hand against the sun to prevent yourself from being dazzled.... It is a grand sight! To anyone seeing it for the first time the ring seems enormous—and yet it is the small halo; the other halo phenomena develop on a still larger scale... You can see a similar ring round the moon, too".

Thus, the halo looks like a luminous ring round the sun or the moon, its angular radius being about 22°. This is the so-called minor halo. There is also the greater halo—a ring with a radius of about 46°.

The reason for halos is close to that of the rainbow. The latter is caused by the refraction and reflection of light in raindrops, while the appearance of the former is accounted for by the refraction of light in the ice crystals that make up clouds high above. These crystals are often hexagonal prisms (see Fig. 5.11*a*). Refraction of a light ray in such a prism follows the pattern of

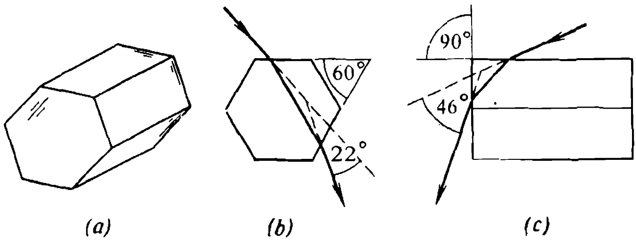


Fig. 5.11.

refraction in a common triangular prism with the refracting angle being either 60° (Fig. 5.11*b*) or 90° (Fig. 5.11*c*).

All these prisms are positioned differently with respect to the incident sun rays. That is why the rays will be deflected by the prisms differently, too. It is worth to mention that the rays which pass through the prism in a *symmetric way* will be the brightest for the observer's eye. Note that this corresponds to the angle of minimum deviation of light in a prism, and, according to Newton, the emergent rays are most condensed where the angles between the incident and the emergent rays are greatest or least. Curiously enough, the observer sees the rainbow when the angle of deviation of light is the maximum and with his back to the sun, whereas the halo is observed if the angle of deviation is least and the observer faces the sun. Figure 5.12 accounts for the appearance of both the minor and the greater haloes. The angle $\delta_1 = 22°$ is the angle

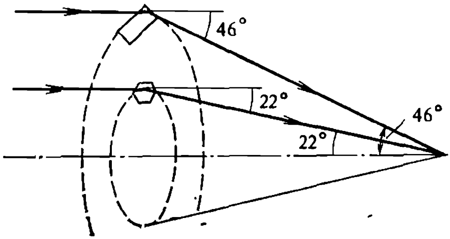


Fig. 5.12.

of minimum deviation of the ray as shown in Fig. 5.11*b* whilst $\delta_2 = 46°$ is the angle of minimum deviation of the ray as shown in Fig. 5.11*c*. These angles can be determined from Eq. (3.6). Assuming that $n = 1.31$ and $\theta = 60°$, we have $\delta = 22°$; at $\theta = 90°$ we get $\delta = 46°$.

Due to the dispersion of light, the halo rings are always many-hued, the inside of the rings assuming the red colour. Since a prism with the refracting angle of 90° is characterised by a greater angular dispersion than a prism with the refracting angle of 60°, a greater halo is richer in colour than a minor one.

If the axes of the hexagonal ice prisms which cause the halo are arranged with no apparent pattern, the ring of the halo will glow with even intensity. If a definite arrangement of the axes is favoured, certain sections of the ring will appear brighter than the others to the observer. In cases like this the halo can assume a specific shape, for example it may resemble a cross. "One could see God's sign: the sun encircled by a ring with a cross right in the middle", these are the words from one Russian chronicle of the 12th century.

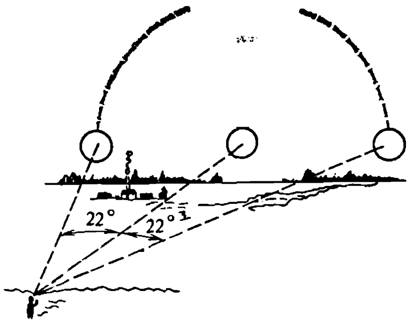


Fig. 5.13.

Religious people trembled at the sight of this phenomenon in the skies, they believed it to be a "divine sign" which would invariably result in multiple death and disasters.

Assume that the axes of the ice hexagons are all vertical. If this is the case, the halo will occur in the form of two bright spots resembling the sun and arranged on the same horizontal line as the genuine sun (see Fig. 5.13). This phenomenon is referred to as *mock suns* (parhelia). The observer of such a phenomenon sees three suns, the angle between the neighbouring suns equaling 22°. The phenomenon can be observed in quiet weather when the sun is close to the horizon.

Chapter Six

Formation of optical images

Assume that we have to get an image of a certain object on the screen or on a photographic plate. It is clear to everyone that in order to achieve this, it is not enough just to place an illuminated object in front of a screen, for light rays reflected from any point of the object's surface will illuminate the whole screen (see Fig. 6.1*a*). To get an image of the object on the screen, we have somehow to arrange the pattern of the light rays getting to the screen from the object. The best thing to do is to make the beams leave the object's surface and arrive at some definite point on the screen.

The Formation of an Image in a Camera Obscura. The easiest thing to do is to place an opaque diaphragm with a small perforation between the object and the screen (see Fig. 6.1*b*). Now the only rays able to pass through the opening will travel from the object as far as the screen; the other rays will be diverted by the diaphragm. This will result in the appearance of the inverted image of the object on the screen. This method of getting an image is the basis of what is called the *camera obscura*. Since only a very small amount of the light reflected by the object helps form the image, we have to place the screen inside a dark space to be able to see the image,

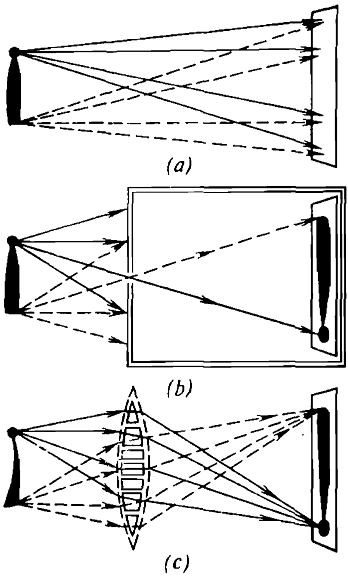


Fig. 6.1.

The Formation of an Image in a Lens System. The principle of forming an image by a camera obscura is to divert all the irrelevant rays. It is better to curve them appropriately (for instance, to refract them) rather than just divert these rays and so let them help the image. In principle, this is achieved by using a system of specially selected and arranged prisms (Fig. 6.1*c*). In practice, however, it is accomplished using lenses instead of a combination of prisms. A lens is a transparent body bounded by two spherical surfaces (see the shaded part of Fig. 6.1c). It is generally known that the combinations of lenses are extensively applied in modern technology to get optical images.

It is possible using a lens to make the light emerging from the point A converge at the point A_1 (Fig. 6.2a). Doesn't this violate *Fermat's principle* (the principle of least time we discussed in the first chapter)? In the case described by Fig. 6.2a, the light passes from A to A_1 following *different* paths when, as it seems, it has to travel along the path to cover which it would take the least time.

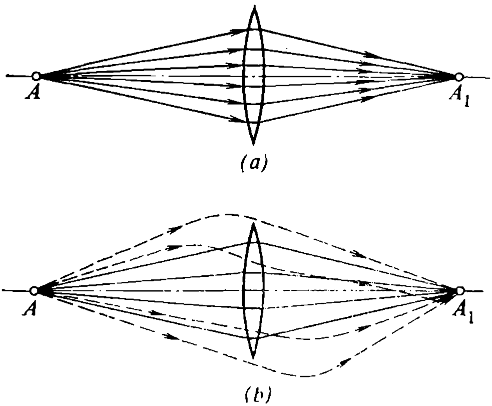


Fig. 6.2.

The case under discussion does not conflict with Fermat's principle in any way. The point is, firstly, that all the paths specified in Fig. 6.2a are travelled by the light in the *same* interval of time, and secondly that this time is actually less than the time required to cover any other path lying outside this set of paths, for example, the ones shown in dotted lines in Fig. 6.2*b*.

Derivation of the Thin Lens Formula from Fermat's Principle. Fermat's principle not only complies with the converging effect of a lens

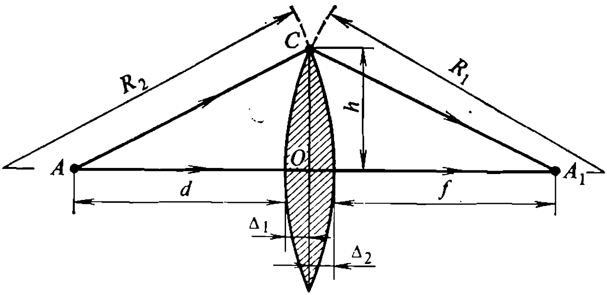


Fig. 6.3.

but also allows us to derive the *thin lens formula* without resorting to the law of refraction. The formula shows the relation between the radii R_1 and R_2 of the spherical surfaces of a lens with a refractive index n and the distance d and f from the lens to the object and to its image, respectively. Note that a thin lens is the one whose thickness is much smaller than R_1 and R_2. Keep in mind that to derive the thin lens formula, it suffices to accept the equality of the periods of time required to cover any two of the paths shown in Fig. 6.2*a*.

Let one of these paths be a straight line connecting the two points A and A_1, and the other be the one touching the edge of the lens (rays AOA_1 and ACA_1 in Fig. 6.3). The time required to cover AOA_1 is $T_1 = [d + n(\Delta_1 + \Delta_2) + f]/c$, and the time to cover ACA_1 is

$$T_2 = [\sqrt{(d+\Delta_1)^2 + h^2} + \sqrt{(f+\Delta_2)^2 + h^2}]/c.$$

Equating T_1 and T_2, we have

$$d + n(\Delta_1 + \Delta_2) + f = \sqrt{(d+\Delta_1)^2 + h^2} + \sqrt{(f+\Delta_2)^2 + h^2}. \tag{6.1}$$

We assume that the so-called *paraxial approximation* holds true, which means that the angles between the rays of light and the optical axis AA_1 of the lens are very small. Thus, $h \ll (d + \Delta_1)$, $h \ll (f + \Delta_2)$ and, consequently,

$$\sqrt{(d+\Delta_1)^2 + h^2} = (d+\Delta_1)\sqrt{1 + \left(\frac{h}{d+\Delta_1}\right)^2}$$
$$= (d+\Delta_1)\left[1 + \frac{1}{2}\left(\frac{h}{d+\Delta_1}\right)^2\right]$$
$$= d + \Delta_1 + \frac{h^2}{2(d+\Delta_1)}.$$

Now we can write

$$\sqrt{(f+\Delta_2)^2 + h^2} = f + \Delta_2 + \frac{h^2}{2(f+\Delta_2)}.$$

Substituting this into (6.1), we have

$$(n-1)(\Delta_1 + \Delta_2) = \frac{h^2}{2}\left(\frac{1}{d+\Delta_1} + \frac{1}{f+\Delta_2}\right). \tag{6.2}$$

Since $\Delta_1 \ll d$ and $\Delta_2 \ll f$ in a thin lens, we rewrite (6.2) as follows:

$$(n-1)(\Delta_1 + \Delta_2) = \frac{h^2}{2}\left(\frac{1}{d} + \frac{1}{f}\right). \tag{6.3}$$

We can write

$$\Delta_1 = R_1 - \sqrt{R_1^2 - h^2} = R_1 - R_1 \sqrt{1 - \left(\frac{h}{R_1}\right)^2}$$
$$= R_1 - R_1 \left[1 - \frac{1}{2}\left(\frac{h}{R_1}\right)^2\right] = \frac{h^2}{2R_1}$$

and it follows that $\Delta_2 = h^2/2R_2$, which immediately gives us the thin lens formula:

$$(n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{1}{d} + \frac{1}{f}. \tag{6.4}$$

The quantity

$$(n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \equiv \frac{1}{F} \tag{6.5}$$

is referred to as the *lens* (or focal) *power*, F is the *focal length* of the lens. The points located on the optical axis on either side of the lens at the distance F from the lens itself are called the *foci* of the lens.

Substituting (6.5) into (6.4), we get the formula the reader is sure to know:

$$\frac{1}{F} = \frac{1}{d} + \frac{1}{f}. \tag{6.6}$$

If a luminous point is brought to the focus of the lens ($d = F$), then, according to (6.6) $1/f = 0$. In this case the lens shapes a bundle of parallel rays as if it were transporting the image of the luminous point to infinity (Fig. 6.4*a*). Since light may also travel in the opposite direction, we can assume that light propagating in a parallel beam along the lens in optical axis will be brought into one point by the lens, the point being the focus (indeed, according to (6.6), if $d = \infty$, we have $f = F$; see Fig. 6.4*b*).

Note that formula (6.4) does not include the value h, which is the distance between the rays under consideration within the plane of the lens.

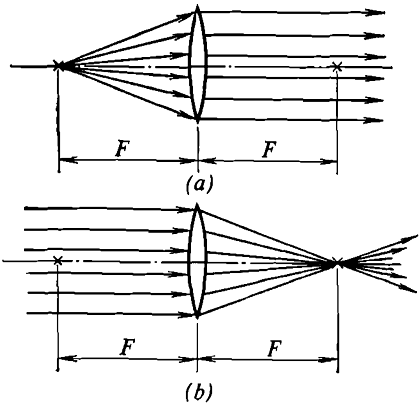


Fig. 6.4.

That is why it is safe to state that any ray which emerged from the point A to pass through the lens is sure to get to A_1. It is clear that h is absent from the thin lens formula because we used the paraxial approximation and we used a thin lens because only in this case we get

$$\sqrt{(d+\Delta_1)^2+h^2}=d+\Delta_1+\frac{h^2}{2(d+\Delta_1)},$$

$$\sqrt{(f+\Delta_2)^2+h^2}=f+\Delta_2+\frac{h^2}{2(f+\Delta_2)},$$

$$\Delta_1=\frac{h^2}{2R_1}, \quad \Delta_2=\frac{h^2}{2R_2}.$$

It is quite clear that the paraxial approximation is only an approximation and the wider the beam of light the worse it fits. It is also clear that any lens has a certain thickness. All this just makes the situation in Fig. 6.4 idealistic.

The Spherical and Chromatic Aberration of a Lens. In reality, a lens is incapable of focusing a bundle of rays into a point. It is impossible,

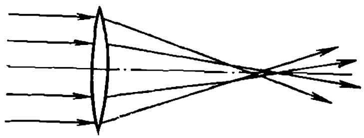


Fig. 6.5.

even if we assume that the incident beam of light is strictly parallel, which in itself is never the case; and neither can we produce a luminous point. It can be shown that the further the path of the ray is from the optical axis of the lens, the shorter the corresponding focal length is (Fig. 6.5). In other words, the edges of a lens deflect light more than does the central part of the lens, the result being blurring of the image. This phenomenon is called *spherical aberration.*

The term "aberration" is generally used to describe all the deteriorations and distortions that occur in optical systems and, particularly, in lenses. Spherical aberration is just one example. There is also what is called *chromatic aberration*. This kind of aberration is due to the dispersion of light. A nonmonochromatic beam of light passing through a lens is decomposed into rays of different colours, violet being deflected more than red. Thus, strictly speaking the focal length of a lens depends on the wavelength of light and increases from the violet region of the spectrum to the red one. Chromatic aberration worsens the quality of an image, bringing about the appearance of a multicoloured band encircling the image.

We cannot afford to dwell any longer upon the aberrations of lenses limited as we are by the confines of this book. Note, however, that proper measures have been developed to cope with their effects and we will discuss them later. For the time being, we shall leave the subject of aberrations and only consider systems of lenses which we shall suppose are *ideal*, i.e. without any aberrations. For this assumption to hold true we shall consider monochromatic and relatively narrow paraxial beams of light.

So let us now return to formulas (6.4)-(6.6).

Real and Virtual Images. Figure 6.6*a* shows the location of the image B_1 of the point B by tracing two rays. The ray BC travels parallel to the optical axis AA_1, is refracted by the lens and then passes through the focus D. The ray BO travels through the centre of the lens and is not refracted at all. The intersection of BO and CD produces the image B_1. Using this diagram, it is quite easy to deduce formula (6.6). The fact that the triangles ABO and OA_1B_1 are similar gives $AO/OA_1 = AB/A_1B_1$. Because OCD and DA_1B_1 are also similar, we have $OD/DA_1 = OC/A_1B_1$. Since $AB = OC$, the right-hand sides of the above equations are equal, which means that $AO/OA_1 = OD/DA_1$, or in other words, $d/f = F/(f - F)$.

It is quite clear that the last equation can be rewritten as formula (6.6).

Assume now that an object AB is placed between the lens and its focus as is shown in

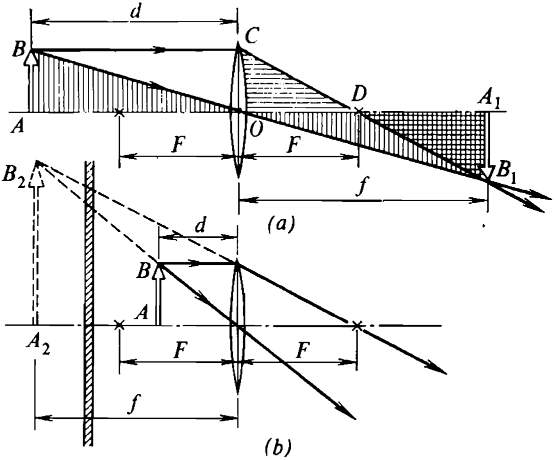


Fig. 6.6.

Fig. 6.6*b*. Constructing the image of the point B as before using two rays, we discover two features. Firstly, the image is now formed as the result of the intersection of the rays' extensions (dotted lines in Fig. 6.6*b*) and not by the rays themselves which passed through the lens. Secondly, the position of the image on the optical axis of the lens is now given by a different formula from (6.6), which the reader can check himself, viz.

$$\frac{1}{F}=\frac{1}{d}-\frac{1}{f}. \tag{6.7}$$

It should be easily understood that the image formed in this way is quite different from the other image we constructed. In the case described by Fig. 6.6*a*, we could, if we wanted, obtain a photograph of the object by placing a photographic plate where the image was formed, that is at the point A_1. It is no use, however, placing a photographic plate where the image is formed in the case shown in Fig. 6.6*b*, that is at the point A_2, since a photograph cannot be obtained there. We can imagine a rather curious situation. Assume that there is an opaque wall between A_2 and A, that is between the image and the object. The image will then be formed behind the wall, so there is no point in placing any photographic plate there.

So can we apply the word "image" to something that is not actually formed by the light rays but by their extensions? Would such an image be anything but a fiction? Despite all the doubts which may arise, the answer to these questions is "yes". The point is that the image can be *seen* even if the opaque wall is there. To see the image, the observer has to be in an appropriate position with respect to the object and the lens. This means that the *observer* must therefore be taken as part of the optical system.

We shall return to this topic in the next chapter which deals with the human eye as an optical system. For the present, just note that an image formed by the intersection of the rays themselves is called a *real image*, whereas an image formed by the extension of the rays is referred to as a *virtual image*. A real image can be photographed or collected on a screen, as well as seen by an observer. A virtual image can be only seen by an eye. In real life we often see virtual images, for instance, this happens every time we look in a mirror.

The formulae (6.7) and (6.6) can be regarded as two variants of the same formula $1/F = 1/d + 1/f$, where d and F have positive signs and f is positive for real images and negative for virtual ones. The designation f in (6.7) and in Fig. 6.6*b* is, therefore, a modulus of a negative quantity, which is why we should rewrite formula (6.7) as follows:

$$\frac{1}{F}=\frac{1}{d}-\frac{f}{|f|}. \tag{6.8}$$

Collecting and Diverging Lenses. The results we obtained above are not only correct for the double-convex lens as shown in Fig. 6.3, but also for all *collecting* lenses. All these lenses are thickest at the centre. A thin beam of light incident on such a lens is focused (if we disregard the effects of the aberration, of course). There are three types of spherical collecting lenses (Fig. 6.7): (a) *double-convex*, (b) *plano-convex*, and (c) *convexo-concave*. They are all described by formula (6.5), if we accept that the radius of a convex surface is positive, and that of a concave surface is negative.

Along with the collecting lenses, use is also made of the *diverging lenses*. A flat bundle of rays incident on such a lens is diverged and the extensions of the refracted rays converge in the lens focus, as shown in Fig. 6.8. Figure 6.9 shows the types of diverging lenses: (a) *double-concave*, (b) *plano-concave*, (c) *concave-convex*. All the

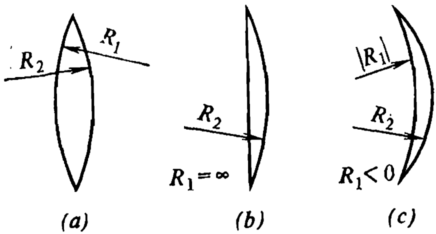


Fig. 6.7.

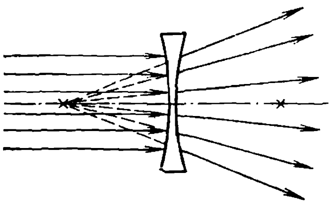


Fig. 6.8.

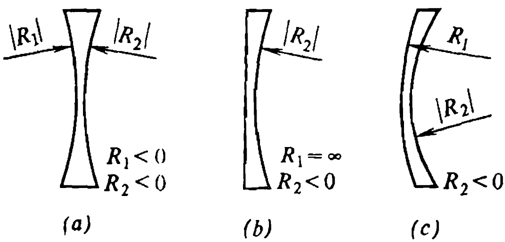


Fig. 6.9.

diverging lenses are thinnest at the centre. If we are to apply formula (6.5) to diverging lenses, the radius of the concave surface must be negative, while the radius of the convex surface must

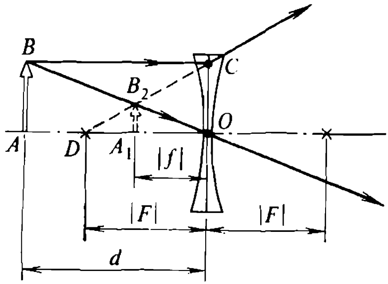


Fig. 6.10.

be positive. It is easy to see that the focal length and, therefore, the focal power of diverging lenses are always negative*. The image formed by a diverging lens is always *virtual* (see Fig. 6.10). Formula (6.6) can be used for diverging lenses if $F < 0$ and $f < 0$. For diverging lenses formula (6.6) can be written as follows:

$$-\frac{1}{|F|}=\frac{1}{d}-\frac{1}{|f|}. \tag{6.9}$$

Using Fig. 6.10, the reader will get formula (6.9) without much effort. To do this use the similarity of the triangles ABO and A_1B_2O and the triangles DCO and DB_2A_1.

* Collecting and diverging lenses are also called positive and negative lenses, respectively, following the sign of their focal lengths (*transl. note*).

A Lens in an Optically Dense Medium. It is worth mentioning that what has been said concerning the collecting and diverging lenses goes for every case when the material of which the lens is made has a greater refractive index than that of the medium in which a lens is placed. Formulas (6.1)-(6.5) were obtained on the understanding that a lens with the refractive index n is placed in the air whose refractive index is taken to be 1 in all such cases. Assume for a change that a lens with a refractive index n_1 is placed in the medium whose refractive index is n_2. Returning to the situation shown in Fig. 6.3, note that now $T_1 = [n_2 d + n_1 (\Delta_1 + \Delta_2) + n_2 f]/c$ and $T_2 = [\sqrt{(d + \Delta_1)^2 + h^2} + \sqrt{(f + \Delta_2)^2 + h^2}]\, n_2/c$, so instead of (6.1) we now have

$$d + \frac{n_1}{n_2}(\Delta_1 + \Delta_2) + f = \sqrt{(d + \Delta_1)^2 + h^2} + \sqrt{(f + \Delta_2)^2 + h^2}.$$

This differs from (6.1) in that n is replaced by the ratio n_1/n_2. Bearing in mind this replacement, we can rewrite (6.4) and (6.5) as follows:

$$\left(\frac{n_1}{n_2} - 1\right)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{1}{d} + \frac{1}{f}, \tag{6.10}$$

$$\left(\frac{n_1}{n_2} - 1\right)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{1}{F}. \tag{6.11}$$

If $n_1 > n_2$, as is the case when a glass lens with $n_1 = 1.5$ is placed in water, $(n_1/n_2 - 1) > 0$ and therefore all the remarks we made concerning diverging and collecting lenses still hold true. If a lens is placed in a medium characterised by greater optical density than that of the lens itself $(n_1 < n_2)$, then $(n_1/n_2 - 1) < 0$, and the sign of the focal length is reversed which causes a lens that is collecting under normal conditions to become diverging (and vice versa). This means that a double-convex lens becomes diverging, while a double-concave one becomes collecting. Thus, a lens can be either collecting or diverging depending on the ratio between the refractive indices of the lens and the medium it is in.

The Early History of Lens Systems. It is rather difficult to state with certainty when the first lenses appeared. We can, however, say that even Alhazen (11th century A. D.) was aware of the ability of a plano-convex lens to magnify an image. The refraction of light by spherical surfaces was first tackled by the well-known English natural scientist Roger Bacon (1214-1292). A graduate of Oxford with an extensive education and good knowledge of classical and Arabic manuscripts, Bacon advocated the experimental method of research. He contributed to experimental optics by carrying out a number of experiments with spherical mirrors, a camera obscura, as well as with plano-convex lenses (referred to as lentils in his time). His work advises people with poor eye-sight to use a *plano-convex lens* when viewing an object. He reasoned that if a person looks at letters or small items using a piece of glass or other transparent object, provided the object is a segment of a sphere with its convex part nearer the eye, the letters would seem bigger and thus be seen better. He went on to claim that a device of this kind was useful for people with poor eye-sight.

At the end of the 13th century, *spectacles* appeared and the production of lenses began to develop quickly (first convex lenses were produced and then concave ones were). We do not know the name of the person who invented spectacles. It is quite probable that the invention was made by artisans whose job was to polish glass. It is not for nothing that the word "lens" (lente) originated from the common people's word "lentil" (lenticula). By the middle of the 14th century spectacles were already widely used. Double-convex lenses were used to make spectacles for "old people" (such spectacles corrected presbyopia), while double-concave lenses were used to produce spectacles for "young people" (to correct myopia).

Invention of a Telescope. The first telescope is considered to have been invented by three Dutch glass polishers, Jansen, Metius and Lippersheim. While carrying out their jobs they must have stumbled on the ability of a system of two lenses to magnify distant objects. However, the telescopes manufactured by Dutch artisans were far from adequate.

The first more or less adequate telescope was invented and manufactured in 1609 by the famous Italian scientist Galileo (1564-1642). It was the first *real telescope* in the world. Galileo gave a detailed account of how he invented the telescope in some of his writings. In "Assay Balance" he wrote: "We are well aware that the Dutchman who first invented the telescope was a common artisan who manufactured spectacles. Manipulating glasses of different types he aligned two glasses, one convex the other concave, placed at different distances from the eye which made him see the effect and in this way he invented the instrument. As for me, I made my instrument through reasoning basing my reasoning on the fact I've just mentioned". Galileo described his telescope thus: "I made a lead tube, and fixed an optical glass at each end. Both of them were flat on one side, while on the other side one lens was convex and the other concave. By looking through the concave glass I saw the objects much bigger and closer; they seemed ten times bigger than they were in reality".

Path of Rays through the Galilean Telescope. Angular Magnification. The telescope Galileo made is now known as a *Galilean telescope.* We shall now discuss its construction in more detail and consider the way light travels in it. A Galilean telescope has two lenses—a collecting and a diverging one. The collecting lens is fixed to the end of the telescope which is pointed at the object, and is called the *objective lens.* The diverging lens is fixed at the other end of the telescope and is the lens through which the observer looks at the object; it is called the *eye-piece.* We shall designate the focal length of the objective lens by F_1, and the focal length of the eye-piece by $|F_2|$. It should be mentioned that $F_1 \gg |F_2|$. The lenses are arranged in such a way that their foci coincide, hence the telescope's length $l = F_1 - |F_2|$. We shall trace the light through a Galilean telescope for two different cases. First we assume that an observer is looking at a small object located nearby; this case is illustrated by Fig. 6.11. We then turn to the situation illustrated by Fig. 6.12, when the observer is looking at a distant object—for example, the Moon—through the telescope.

First consider the observation of the near object. Figure 6.11*a* shows the passage of two rays travelling from the point A to the objective lens. The ray AO passes through the centre O of the objective lens and is incident on the eye-piece at N. In order to determine the further path of this ray draw the ray LO_1 through the centre O_1 of the eye-piece parallel to ON. The ray LO_1 will intersect the focal plane EE of the eye-piece at a point P. It can be proved that after

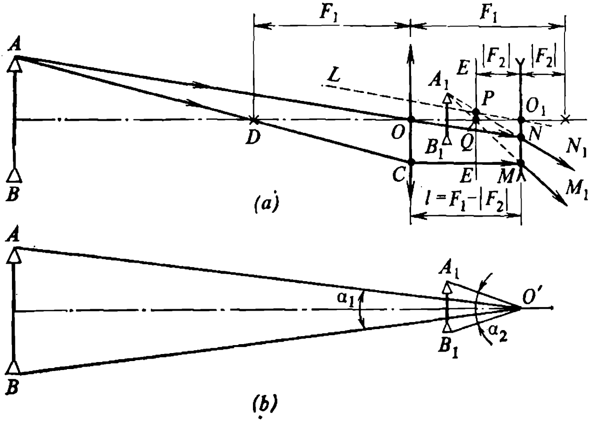


Fig. 6.11.

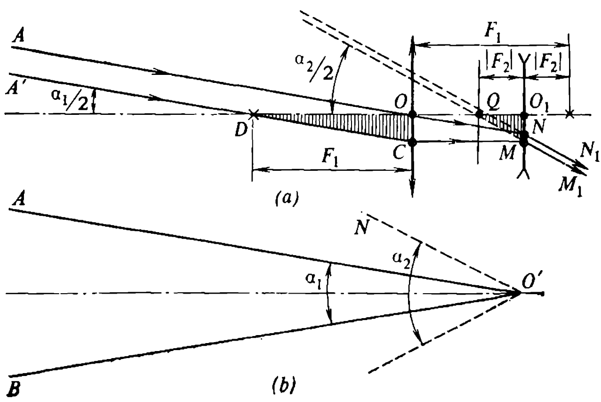


Fig. 6.12.

emerging from the eye-piece, the ray NN_1 will travel along a path whose extension passes through P. This can be shown from the well-known fact that a parallel beam of light is refracted in an ideal lens so that either the rays themselves or their extensions intersect in the focal plane of the lens. In the case under discussion the rays LO_1 and ON are parallel; it is clear that LO_1 passed undeflected through the eye-piece. Consider now the second ray from A to the objective lens. This is the ray AD shown in Fig. 6.11*a*. It passes through the focus D of the objective lens, which is why, having passed through the objective lens, the ray travels parallel to the optical axis of the lens and is refracted in the eye-piece so that its extension passes through the focus Q of the eye-piece. The intersection of the extensions of the rays NN_1 and MM_1 gives the point A_1, which is the image of the point A in a Galilean telescope. The point B_1 is arranged symmetrically to it and is the image of point B.

Thus we see that a Galilean telescope forms an erect virtual image of the object. If the dimensions of the image A_1B_1 and the object AB were compared, a reader might conclude that a Galilean telescope reduces the object instead of magnifying it. This conclusion is false. Although the line A_1B_1 is shorter than AB, it is much *closer* to the observer's eye. As can be seen in Fig. 6.11*b*, the observer at O' sees the image A_1B_1 at an angle α_2 which is greater than the angle α_1 at which the observer would see the object unaided. We stress that the telescope magnifies the *angle* at which the observer sees the object; in other words the telescope magnifies the angular dimensions of the object. The ratio

$$\chi=\frac{\tan\alpha_2}{\tan\alpha_1} \tag{6.12}$$

is called the *angular magnification*. In the next chapter, we shall make sure that it is the angular (not linear) magnification that matters when observing an object.

Now let us assume that the object being observed through the telescope is so distant that the rays arriving at the plane of the objective lens can be regarded as *parallel*. For example, point the telescope at the moon. If observed unaided, the moon's disc subtends an angle $\alpha_1 = 0.5°$. Strictly speaking, the angle α_1 depends on the altitude of the moon above the horizon due to the refraction of light in the atmosphere. However, this is of no importance for us at present. Assume that the parallel rays AO and $A'D$ shown in Fig. 6.12*a* arrive at the plane of the objective lens from the upper edge of the moon's disc; they subtend an angle $\alpha_1/2$ with the optical axis of a telescope pointed at the centre of the moon. The reader can ascertain on his own that after being refracted by the eye-piece, the rays in question will again be *parallel to each other* ($NN_1 \| MM_1$). However, the angle made by the rays with the optical axis of the telescope will be different. Designate it through $\alpha_2/2$. It is evident that $\alpha_2 > \alpha_1$ (see Fig. 6.12*b*).

The ratio α_2/α_1 is easy to determine. It is clear from Fig. 6.12*a* that $\tan(\alpha_1/2) = OC/DO$, and $\tan(\alpha_2/2) = O_1M/QO_1$. Since $OC = O_1M$, we get

$$\frac{\tan(\alpha_2/2)}{\tan(\alpha_1/2)} = \frac{DO}{QO_1} = \frac{F_1}{|F_2|}. \tag{6.13}$$

Proceeding from the fact that the angles α_1 and α_2 are very small (they are drawn very large in Figs. 6.12*a* and *b* in order to make diagram more vivid), we can replace the ratio of the tangents by the ratio of the angles, and remembering (6.12), we rewrite (6.13) as

$$\chi = \frac{\alpha_2}{\alpha_1} = \frac{F_1}{|F_2|}. \tag{6.14}$$

Thus, when we observe objects through a Galilean telescope, it magnifies angular separation $F_1/|F_2|$ times. If, for example, $F_1/|F_2| = 10$, the moon's disc will be seen using the telescope to subtend the angle $\alpha_2 = \alpha_1 F_1/|F_2| = 5°$ instead of $\alpha_1 = 0.5°$. Note that a page of this book subtends the same angle about 1.5 m away.

Galileo's Astronomical Observations. The greatest service of Galileo to science was that he invented the first real telescope and used it to observe the sky. "Having abandoned earthly problems I became engaged with celestial ones", the scientist wrote. Using his telescope, Galileo observed the moon's landscape and later discovered the phases of the Venus and sunspots. Galileo saw with his own eyes (aided by the telescope) what Copernicus believed in—the existence of the four satellites of the Jupiter and later discovered Saturn's satellites. Galileo realized what a blow his discovery of satellites was to the enemies of Copernicus' system and to the dogmas of the Church, for according to its interpretation of the Holy Scriptures, the Earth was believed to be the centre of the Universe. After carrying out careful observations, Galileo wrote the following: "I decided without the least hesitation that there are four luminaries rotating around Jupiter in the same way as Venus and Mercury rotate around the Sun. This is good evidence to disperse the doubts of those people who admitting that the planets rotate around the Sun fail to admit that the Moon rotates around the Earth and both of them complete a yearly revolution around the Sun... . We know that there are planets rotating one around the other and at the same time both of them rotate around the Sun. We also know that there are four moons rotating around Jupiter near it during its twelve-year revolution around the Sun."

It is little wonder that these sensational discoveries caused the servants of the Church to set out on an embittered pursuit of Galileo.

In our time the telescope Galileo invented has been replaced by more sophisticated telescopic systems which achieve much greater angular magnifications and have better quality of images. However, this modest "predecessor" of modern telescopes is still used today. Each time we use a pair of common opera-glasses we are actually using two Galileo's telescopes.

Kepler's "Dioptrics" and Other Books. The outstanding German scientist Johannes Kepler lived in the same era as Galileo. He invented his own version of the telescope in which he used collecting lenses for both the objective and the eye-piece. In 1611 Kepler published his major work on optics called "Dioptrics". It describes the characteristics of different lenses and their combinations, considers how to find their foci, outlines the laws governing the location of image with respect to the position of the object. This work was the first to explain how to locate the image by tracing two rays or their extensions. A description of spherical aberration of lenses can also be found. True, the book does not contain even one precise formula; acting in the spirit of his time, Kepler preferred a description of the uniformities he had noticed more than numerical relations. However, we can hardly criticize this fundamental piece of work, for we should not forget that when Kepler wrote his book the law of refraction had not been formulated.

The first *microscopes* appeared alongside the first telescopes and were used more and more extensively at the end of the 16th and turn of the 17th century. In the middle of the 17th century, the well-known Dutch naturalist Antony van Leeuwenhoek, achieved outstanding results in his manufacture of microscopes and using them, he penetrated the world of microbes.

After the derivation of the law of refraction, scientists got down to finding formulas for lens systems. In 1646 the Italian Francesco Bonaventura Cavalieri derived the formula for the double-convex lens, viz. $(R_1 + R_2)/R_1 = 2R_2/F$. It is easy to ascertain that Cavalieri's formula follows from (6.5), if $n = 3/2$. In 1693 the British astronomer Edmund Halley derived the general formula for thin lens. Isaac Newton analysed the refraction of light on a spherical surface in "Lectures on Optics". He distinguished between paraxial optics and the optics of rays with significant angles of inclination. Newton also determined the spherical and chromatic aberration due to a spherical surface. In doing this, he made a mistake and assumed that refracting systems cannot in principle be free from chromatic aberration.

Dollond's Achromatic Lens. In 1746 Leonard Euler published "The New Theory of Light and

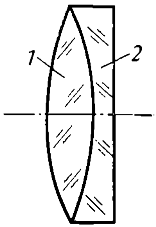


Fig. 6.13.

Colour", in which he discussed the relation between colours and the wavelengths. Euler showed it was possible to eliminate chromatic aberration in lenses. Later, in 1758 the English optician John Dollond managed to produce a lens with no aberration, using Euler's ideas; the lenses were called *achromatic*. The appearance of achromatic lenses increased the sophistication of telescopes, microscopes and other optical instruments.

Dollond's achromatic lens was a combination of two lenses, one made from crown glass, the other from flint glass. A lens of this kind is shown in Fig. 6.13, where *1* is the lens made from crown glass and *2* is the lens made from flint glass. The focal power of the two thin lenses combined equals the sum of the focal powers of the two lenses taken separately. Keeping in mind (6.5) and the fact that the focal length of a plano-concave lens is negative, we can write the following for the focal power of the achromatic lens:

$$\frac{1}{F}=\frac{1}{F_1}+\frac{1}{F_2}=(n_1-1)\frac{2}{R}-(n_2-1)\frac{1}{R}. \qquad (6.15)$$

Indices 1 and 2 are for crown glass and flint glass lenses, respectively; R is the radius of the lens' spherical surface. It is possible to arrange different kinds of glass to make $1/F$ almost independent of the wavelength of light, although $1/F_1$ and $1/F_2$ taken separately depend on the wavelength.

Consider the following problem. *Verify that if two glasses are chosen appropriately the focal power of the achromatic lens shown in Fig. 6.13 will be the same for the wavelength* $\lambda_b = 0.49$ µm *(blue) and* $\lambda_r = 0.66$ µm *(red).*

We rewrite (6.15) as follows:

$$\frac{1}{F}=\frac{1}{R}(2n_1-n_2-1) \qquad (6.16)$$

and indicate the refractive index for λ_b by the superscript "b", and the refractive index for λ_r by the superscript "r". It is clear from (6.16) that we must find appropriate glass so that the following equation is true:

$$2n_1^b-n_2^b=2n_1^r-n_2^r$$

or

$$2(n_1^b-n_1^r)=n_2^b-n_2^r. \qquad (6.17)$$

If we now turn to an appropriate handbook, for example, "Tables of Physical Values" edited by Academician

Kikoin (Atomizdat Moscow, 1976), we can find that for Φ2 flint glass $n_2^b - n_2^r = 16.8 \times 10^{-3}$ and for K19 crown glass $n_1^b - n_1^r = 8.4 \times 10^{-3}$. Thus (6.17) holds true for the kinds of glass we have specified. It is also true for Φ6 flint glass ($n_2^b - n_2^r = 15.9 \times 10^{-3}$) and K5 crown glass ($n_1^b - n_1^r = 7.95 \times 10^{-3}$).

At present lenses are widely used in optics, being incorporated in a great number of instruments such as telescope systems, microscopes, photographic and movie equipment, spectrometers, optical communication lines, and lasers. Modern lenses are often quite sophisticated optical elements for obtaining good quality images, in which aberration is reduced to a minimum.

Fresnel Zone Plate. Along with refinement of lenses made of glass and other materials (for example, transparent polymers), our century has witnessed the development of an altogether *new method* of obtaining optical images. The method was first described at the turn of the 19th century in a book on wave optics written by Agustin Jean Fresnel, a French scientist. It followed from his work that there was no need to manufacture a lens using glass; it can be simply "drawn" on a sheet of transparent material.

Before discussing this unusual lens, let us first meet the idea of a "Fresnel zone", one which is widely used in wave optics. Suppose a parallel beam of monochromatic light is propagating along the line $O'O$; and the plane S is the wavefront of light (Fig. 6.14). We choose a point on $O'O$ and designate it D, while the distance between D and the plane S we will call F ($OD = F$). Imagine a locus in the plane S that lies at a distance of $F_1 = F + \lambda/2$ away from D, where λ is the wavelength of the light; it will be a circle of radius $r_1 = OO_1$. It is clear that $r_1 = \sqrt{F_1^2 - F^2}$. Now draw in the circles which would be the loci of the points lying at the distances $F_2 = F + \lambda$, $F_3 = F + 3\lambda/2$, ..., $F_m = F + m\lambda/2$ away from D, respectively.

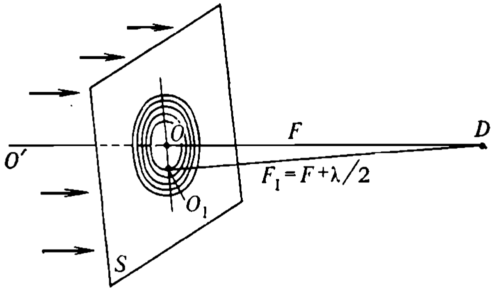


Fig. 6.14.

Having done this, a system of circles appears on the plane S with a common centre at O and with radii $r_m = \sqrt{F_m^2 - F^2}$. It is clear that

$$r_m = \sqrt{m\lambda F + \left(\frac{m\lambda}{2}\right)^2} = \sqrt{m\lambda F},$$

$$m = 1, 2, 3, \dots \tag{6.18}$$

(the $(m\lambda/2)^2$ term can be rejected since $(m\lambda/2)^2 \ll \ll m\lambda F$ at small values of m). Bearing in mind (6.18), we draw in the system of circles accurately (these circles are the shaded ones in Fig. 6.15). Think of these circles as the central lines of alternating ring-shaped zones, each line corresponding to one value of m. The central zone is

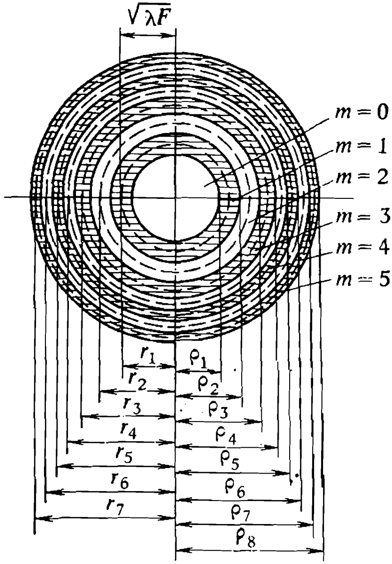


Fig. 6.15.

a disc ($m = 0$). The radius of the circle demarcating the $(m - 1)$th and mth zones can be determined from the formula $\rho_m = \sqrt{\Phi_m^2 - F^2}$, where $\Phi_m = F_m - \lambda/4$. Thus,

$$\rho_m = \sqrt{\left(F + m\frac{\lambda}{2} - \frac{\lambda}{4}\right)^2 - F^2}$$

$$= \sqrt{\left(m - \frac{1}{2}\right)\lambda F}. \qquad (6.19)$$

The zones shown in Fig. 6.15 (shaded and unshaded) are called *Fresnel zones.*

Take a sheet of transparent material and draw Fresnel zones on it. Then, blacken all the zones with odd m's, that is the zones shaded in Fig. 6.15. We get what is called a *Fresnel zone plate*. If $F = 1$ m and $\lambda = 0.64$ μm, we have $\sqrt{\lambda F} = 0.8$ mm. Actually, it is not difficult to make such a plate.

Consider a Fresnel zone plate perpendicular to a parallel beam of light. In order to understand the effect produced by the plate on the light, recall *Huygens' principle* which we discussed in Chapter One. When a plane wave-front of the light beam reaches the surface of the plate all the points in the central circle and in the area of transparent (even) rings instantly become sources of secondary spherical waves. The points of the plate in the nontransparent (odd) rings will not generate secondary waves. Since the zone plate is made so that the distances from the middles of the transparent rings to a point D, which lies on the light beam's axis, differ from each other by integral multiples of the wavelengths, all the secondary waves of light will reach D in the *same phase*; as a result, the intensity of the light at D will increase considerably, as if the original light beam had been focused at the point.

We see that a Fresnel zone plate acts like a collecting lens and can be thought of as a two-dimensional analogue of a lens. A plate whose ring structure corresponds to (6.18) and (6.19) collects a beam of light with wavelength λ into a point on the beam's axis a distance F from the plate.

A Fresnel zone-plate lens does not refract light. A different optical phenomenon, called *diffraction* of light, is occurring. A zone plate is an example of a *diffraction grating*, but any discussion of diffraction is beyond the scope of this book. However, it is worth mentioning that quite recently, to be more exact, after the appearance of lasers, "pictures" similar to Fresnel zone plates have begun to be applied to control light fields and obtain optical images. Such "pictures" are called *holograms*; they are actually diffraction gratings with complicated zone patterns, though they do not resemble in the least the relatively simple Fresnel zone plate we discussed above. A new branch of optics called *optical holography* is now developing very quickly.

Chapter Seven

The eye

Two Kinds of Optical Instrument. Optical image-forming instruments can be divided into two groups. The first group (*projectors*) forms *real* images of an object which is projected onto a screen or photographic plate. This kind of image can be viewed simultaneously by many observers. A motion picture film is a typical example, since everyone in the cinema can simultaneously take in the real image projected on the screen.

The second group forms *virtual* images of an object. As a rule, only one observer can see this kind of image, though in the case of a magnifying glass several people can be looking through it at the same time. It has to be noted that the virtual image itself is something nonexistent, for it is not the rays but their extensions that intersect, and the latter are purely imaginary. However, the image becomes a reality as soon as the *observer's eye* becomes a part of the optical system. The virtual image formed by an instrument is transformed by the eye into a real image projected on the eye's retina. It is no coincidence that optical instruments forming a virtual image are sometimes called *optical aids*. They include magnifying glasses, spectacles, microscopes, and telescopes. Strictly speaking, the observer's eye must always be included in the optical system when tracing the light rays. The necessity of this was mentioned in the previous chapter in connection with the angular magnification produced by a Galilean telescope.

What is the role played by an eye in an optical experiment? Is there a dividing line between the science of vision and the science of light, and what may optics be considered to be after all?

These questions have always been important to scientists. In the early days of optics the role of the eye was definitely exaggerated, and optics was, in fact, a *science of vision.* It has to be remembered that in those days the eye was believed to emit rays, and the expression "light of my eyes" was taken literally. Much later people understood that the role of the eye was to receive the light coming from an object, and still later they began to differentiate between the science of vision and optics which was then regarded as the *science of light.* Today we are in a position to assess the role of the observer's eye in experimental physics. It is curious, however, that while once the role of the eye used to be exaggerated, we now tend to underestimate it. In any case, a few modern textbooks and teaching aids on physics emphasize the role of the observer's eye in perceiving virtual images.

A brief account of how our ideas about the mechanism of vision and the role of the eye have changed with the time will be quite useful. Naturally, they developed as we acquired more knowledge of the structure of the eye. We should therefore get to know what the modern ideas are about the anatomy of the human eye before turning to ancient history.

The Structure and Optical System of the Human Eye. The human eye or rather its eyeball is an almost globular organ. Figure 7.1 shows a sectional view of the eye. It has a tough outer skin called the *sclera* (*1*), and at the front of the eye the sclera is replaced by a convex transparent membrane called the *cornea* (*2*). The sclera is

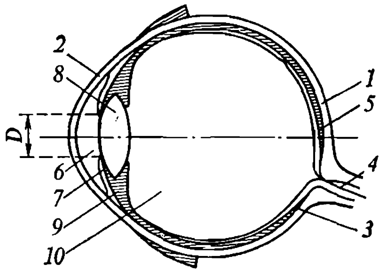


Fig. 7.1.

lined inside with *vascular tissue* (*3*) which consists of blood vessels feeding the eye. The *optic nerve* (*4*) diverges at the back of the eye to form a delicate film of nerve fibres sensitive to light—the *retina* (*5*). The retina consists of several layers of receptor cells of different types; their function is to receive the optical image. At the front of the eye, directly behind the cornea, there is a transparent *aqueous tissue* (*6*), and behind it there is the *iris* (*7*). The iris can be different colours, which explains why people have blue, green, or brown eyes. There is a round opening in the iris called the *pupil*, whose diameter D is variable. The iris, together with the pupil, functions as a diaphragm regulating the amount of light entering the eye. Behind the iris is a *crystalline lens* (*8*), a biconvex lens made by Nature herself. The lens is fastened to the *ciliary muscle* (*9*) which can be relaxed or tensed, and thus make the lens more spherical, changing its focal power. The interior of the eye between the lens and the retina is filled with jelly-like substance called the *hyaloid* (*10*).

Beams of light coming from an object into the eye are refracted by the aqueous tissue, the crystalline lens and the hyaloid. The refractive indices of the aqueous tissue and the hyaloid are about the same as that of water, while the index of refraction of the crystalline lens is approximately 1.4. The real image of an object being observed is formed on the retina and is upside down. The optic nerve sends a signal to the brain which makes the corrections necessary for us to see objects in their natural positions and not the wrong way round.

A System Made Up of a Magnifying Glass and an Eye. Figure 7.2 shows the paths of rays in such an optical system. The eye is close to the collecting lens, and the object AB is placed behind the lens closer to the lens than its focal length F. Examine one of the rays travelling from B, in particular, the one which passes close to the end e of the pupil. Having been refracted in the eye, the ray falls on the retina at B_2. A similar ray travelling from A will reach A_2 of the retina. As the ray from B is refracted at point b of the lens before entering the eye, and the ray from A is refracted at a, the eye will not see the object AB itself but its virtual image A_1B_1 because *be* is on the same line as B_1e, while ac lies on the same line as A_1c (any refraction in the aqueous tissue in front of the crystalline lens is not taken into account). Note that the rays BO_1 and BG-GD, as well as AO_1 and AH-HD, which are shown in the figure, are auxiliary, for they help to locate

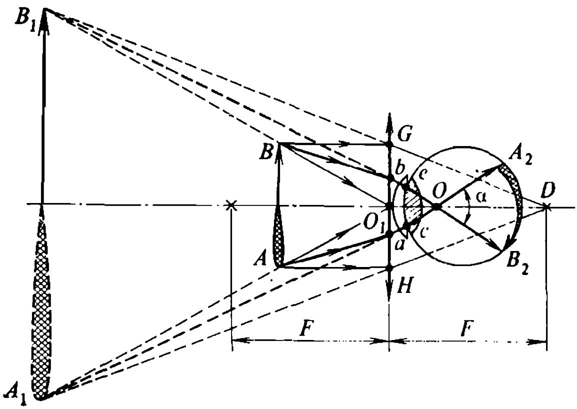


Fig. 7.2.

the image A_1B_1 of the object AB formed by the lens. So, the eye sees the virtual image A_1B_1 and transforms it into the reversed retinal image A_2B_2. Point O is the *optical centre* of the eye (note that in reality it is situated further to the left than the figure shows—on the back surface of the crystalline lens). The angle α at which the observer sees the image is termed the *angle of view*. The greater the angle of view, the greater the *virtual* magnitude of the object.

The nerve fibres in the retina consist of light-sensitive cells and there are two kinds of them, i.e. *rods* and *cones*. In order for an eye to resolve two points, the distance between their images on the retina must exceed the dimensions of the rods and cones. This means that the angle of view must be not less than one minute of arc.

The Development of the Science of Vision from Democritos and Galen to Alhazen and Leonardo da Vinci. Now that we know the general structure of the eye and how light travels inside it, we turn to ancient times.

In the 6th century B.C., the followers of Pythagoras believed that people's eyes generated an invisible emanation which "feels" the objects they are looking at. In the 5th century B.C., Empedocles of Agrigentum supposed that there was besides emanation from eyes an emanation from luminous objects. Democritos (460-370 B.C.), the best exponent of the materialist philosophy in ancient times, refuted the existence of emanations from the eye and explained vision by the effect of atoms emitted by the luminous object and entering the eye. Nevertheless, Euclides wrote: "The beams emanated by the eye travel along direct paths".

Four and a half centuries later Galen (130-200 A.D.) described the anatomy of the human eye for the first time. Although imperfect, the description mentioned the optic nerve, retina, and crystalline lens. However, Galen claimed that the "light of the eyes" produced by the brain travelled along the optic nerve, was dispersed by the hyaloid and then collected by the crystalline lens which Galen believed to be an organ of perception.

About nine centuries later Galen's work attracted the attention of the famous Arabean scientist Alhazen (11th century A.D.). He accepted the description of the anatomy of the eye suggested by Galen but rejected "the light of the eyes" idea outright. Alhazen wrote: "A visual image is formed by the rays emitted by visual bodies

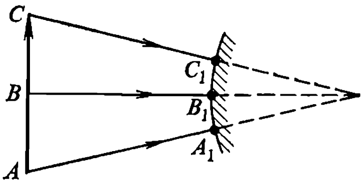


Fig. 7.3.

and entering the eye". It should be emphasized that Alhazen was the first scientist to attempt to understand how a visual image is formed. Before him nobody had troubled to do so, it being taken for granted that a visual image appears immediately in an integral indivisible process. Alhazen made a brilliant guess: he assumed every point on the visible surface of an object corresponds to a point inside the eye, and consequently, image formation was a multitude of elementary processes of forming the images of separate points of the object. True, Alhazen believed that the perception centres were not situated on the retina but were rather on the front surface of the crystalline lens. He wrote: "A visual image is obtained with the help of a pyramid whose vertex is in the eye and whose base is on the visible body". Figure 7.3 illustrates Alhazen's idea.

According to this concept, the rays of light from A, B, and C are concentrated in the middle of the eye.' Points A_1, B_1, and C_1 on the front surface of the crystalline lens correspond to points A, B and C.

Leonardo da Vinci's Comparison of the Eye to the Camera Obscura. The great Italian artist and naturalist Leonardo da Vinci (1452-1519) rectified Alhazen's mistake and transferred the perception points from the surface of the crystalline lens to the *retina*. Furthermore, in a detailed description of the *camera obscura*, he pointed out that the same thing happens inside the eye. Thus, for the first time an apparatus was found which could be considered to be an optical analogue of the eye. Academician Vavilov lays particular emphasis on this: "Before the camera obscura, images were only known to appear in the eye and in man-made pictures. The camera drew a line between light and sight, hence its landmark role in the process of cognition. After the invention of the camera obscura, the structure of the eye which had up to then been a topical question of optics, had to be treated by physiology and medical science. Strictly speaking, in the 16th century optics (which really means "the science of vision") ceased to be such and turned into the science of light."

One has to bear in mind, however, that Leonardo made several errors in his comparison of the human eye and the camera obscura. He supposed that the crystalline lens was spherical and situated in the centre of the eyeball. Figure 7.4 shows the paths of rays in such an "eye". Leonardo saw that the image obtained in the camera obscura is inverted. However, he expected the retinal image to be normal, and believed that the purpose of the spherical crystalline lens inside the eye was to invert the image once again so that the final image would be normal (see

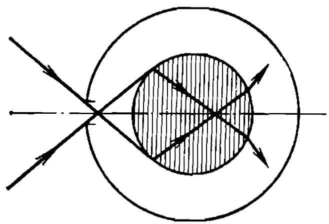


Fig. 7.4.

Fig. 7.4). Leonardo's mistake is explained not only by his wish to have a noninverted image on the retina but also by the fact that his methods for investigating the eye experimentally were far from perfect. He wrote that in order to see the inside of a dissected eye without spilling its liquid it had to be put into the white of an egg and boiled. Leonardo did not know that when boiled the crystalline lens became spherical and its true shape inside the eye was changed.

Withal, the human eye really is like the camera obscura. However, there is not another spherical body inside the hyaloid which would invert the image once again. The crystalline lens is not spherical but has the shape of a biconvex lens and lies right behind the opening of the "camera obscura". Thus the image on the retina, like the one obtained in a camera obscura, is inverted.

Curiously enough, an Italian called Giambattista della Porta demonstrated a true-to-life model of the eye in the late 16th century without realizing it. He made a camera obscura with a collecting lens next to its opening. He observed an inverted image on the back of his improved camera. Porta used his camera obscura for entertainment rather than anything else. He never realized that the crystalline lens in the eye is similar to the lens fitted in the opening of the camera obscura.

Kepler on the Role of the Crystalline Lens. Young's Explanation of Accommodation. Johannes Kepler was the first person (in the early 17th century) to point out that the retinal image is inverted. Kepler also understood that the crystalline lens is necessary for the *accommodation* of the eye, i.e. for its adjustment to close and distant objects in order to obtain sharp images on the retina. However, he misinterpreted the process of accommodation: he supposed that the adjustment of the eye is due to the varying distance between the crystalline lens and the retina.

It was only at the beginning of the 19th century that Thomas Young (1773-1829), an outstanding British physicist, doctor, metallurgist, Egyptologist, oceanographer and botanist, proved that accommodation was not due to movements of the crystalline lens but to changes in the curvature of its surfaces, in other words, to changes in the focal power of the crystalline lens.

We mentioned above that the crystalline lens is held by the ciliary muscle. When the muscle is relaxed the focal power of the lens is the least and the distinct image of a distant object is formed on the retina of a normal eye; the eye is said to be accommodated for infinity. As an observer starts looking at closer objects, the eye accommodates itself automatically: the ciliary muscle tenses, the lens assumes a more spherical form and its focal power increases. Thus, the human eye is self-adjusting, which gives a human being the ability to see objects at different distances clearly.

A man lost in thought is often said to have a blank look. He looks at you but does not see you—he seems to be staring right through you and into nothingness. His eyes are accommodated for infinity, and this shows itself quite clearly. Someone reading a book looks different. His whole aspect is concentration itself, his look intense even if he is not looking you in the face. His eyes are accommodated for a very close object—the book—and the ciliary muscles of the eyes are as tense as they can be. It is only natural that after a great deal of reading people feel their eyes are tired.

Presbyopia and Myopia. The eye's ability to accommodate is not unlimited and to describe this we speak of near points and distant points. A normal eye has no distant point, and the near one is about 20 cm away from it. The minimum distance L at which an eye can see objects distinctly and without getting too tired is called the *normal viewing distance*; for a normal eye $L = 25$ cm. Some people have eyes with abnormal ranges of accommodation, and so there are people who are short-sighted (myopic) and those who are long-sighted (presbyopic). For *short-sighted* persons $L < 25$ cm, and its near point can be only several centimetres away from the eye. Its distant point is not at infinity as it is for normal eyes, but can be relatively close, e.g. only a few metres. Short-sighted people usually have a blurred vision of more or less distant objects. For *long-sighted* people $L > 25$ cm, the near point can be a few metres or more from the eye. Such people can't see things right in front of their eyes, but they have a fairly clear vision of distant objects. The ciliary muscle invariably becomes less elastic with age, which is why elderly people, even those with good sight, find their near point gradually increasing. This is called senile presbyopia.

The Eye as a Perfect Optical Instrument. Despite some possible departures from normal, the human eye is a wonderful optical instrument. We have mentioned that an eye can adjust itself for sharpness by varying the focal power of the crystalline lens. Besides, one has to bear in mind the eye's ability to change the diameter of the pupil which makes it possible, as in the case of the variable aperture diaphragm in the photographic objective, to control the amount of light entering the eye and change the depth of focus. These are just a few of the merits of the eye as an optical instrument.

Aberration does not interfere with the perception of an image. This is, in a way, connected with the structure of the retina. The point is that the resolving power of the retina and its spectral sensitivity are at their optimum only within a relatively small area and are much worse beyond it. This area is called the *macula*. In its centre there is a region called *fovea*—a depression with the greatest number of receptor cells (mostly cones). In Fig. 7.5 the macula is shaded, and the fovea can be clearly seen. The line AA passing through the optical centre of the eye and the centre of the macula (fovea) is termed the *visual axis* of the eye. It is at an angle $\varphi \approx 5°$ to the optical axis OO of the eye. The angular dimensions of

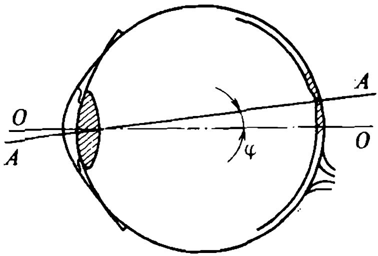


Fig. 7.5.

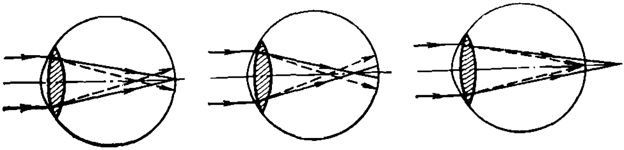
the macula are a little over 5°, while those of the fovea are only 1-1.5°. This peculiarity of the retina results in the eye chiefly seeing the narrow paraxial beam of light travelling along the visual axis. This decreases the possibility of aberration. True, if the object is illuminated more, the possibility of aberration would grow, but then the situation is repaired by a decrease in the diameter of the pupil.

It might appear that the small angular magnitude oft he macula should diminish the *field of vision*, for the angle of view is not much greater than the angular magnitude of the macula. If this were true the eye would be motionless. But Nature has made the eye very *mobile*, which makes up for the small angular dimensions of the most effective area of the retina. When we are viewing an object, we keep changing the direction of the visual axis of the eye without noticing it, thus roaming from one point of the object to another. As a result, different parts of an object's surface will be "seen" on the macula and, in particular, on the fovea at different moments of time. The eye seems to linger on some points of the object, while it slips by other ones without stopping. A complete *visual image* of the object is formed as a result of such consecutive viewings. All this gives us the ability to concentrate on certain details, and at the same time not to notice that the field of distinct vision is limited. The mobility of the eye makes our field of vision appear quite extensive. It seems to be up to 120° along the vertical and 150° along the horizontal.

We must now mention the eye's ability to see a detail for *some time after* it goes out of the field of vision. The period of time is approximately 0.1 s long and this value is optimal. Imagine that this length of time suddenly becomes 100 times greater or smaller. In the first case the details would crowd one upon another without forming a complete visual image. In the second case the visual image of the object would split into disintegrated details. Note that in either case we wouldn't be able to watch a film or a TV program.

We have cited here only five "improvements" that Nature has provided for our organ of vision, viz. the ability to self-adjust for sharpness, variable diameter of the pupil, high retinal sensitivity close to the visual axis of the eye, good mobility of the eye and the optimal duration of visual perception. This shows how perfect our natural optical instrument is. If we tried to make a model of the human eye, we would have to design a quickly rotating camera obscura with a varying diameter of the entrance opening, with a lens of varying focal power by the opening, and a complicated system of light receivers at the back of the camera obscura. Besides, we would have to face the problem of coördinating the movements of the camera with the changes of the diameter of the opening and of the focal power of the lens, as well as the problem of agreement between all these and the luminance of the viewed objects, their shape and the distance between the objects and the eye. Still, this would hardly be an adequate model of human vision, because even this improved camera would be unable to concentrate its attention on certain details. In the final analysis, the seeing is inseparable from the process of thinking.

Academician Vavilov in his book "The Eye and the Sun", which we recommend to the reader, gives a detailed analysis of the characteristics and peculiarities of the human eye. Comparing the characteristics of the eye with those of sunlight, he shows that "the eye is a result of an extremely long process of natural selection, the outcome of the changes in the human body under the influence of environment and struggle for survival, for better adjustment to the outside world". Commenting on the wonderful qualities of the human eye Vavilov emphasises: "All this is the result of the adaptation of the eye to sunlight on the Earth. The eye can't be understood without knowing the Sun. In fact, the characteristics of the Sun can be treated theoretically to



(a) (b) (c)

Fig. 7.6.

yield outline of the properties of the eye even before we know them."

At present mankind cannot create an artificial eye and the project does not seem feasible in the foreseeable future either. However, we learned to correct the defects of eyesight and to increase the capabilities of the living eye rather a long time ago.

Spectacles. Thus spectacles have been used for quite a long time to correct myopia and presbyopia (see Chapter Seven). Figure 7.6 shows the passage of rays coming to the observer's eye from very distant objects. The three cases represent (a) normal sight, (b) myopia, and (c) presbyopia. The rays inside the eye (shown by solid lines) are those that occur when the ciliary muscle is relaxed, while the rays represented by the dotted lines occur when the muscle is tense. We can see that in the case of normal sight the eye is accommodated for infinity when the ciliary muscle is relaxed, in the case of myopia accommodation for infinity is quite impossible, and in the case of presbyopia it is possible if the ciliary muscle is tense. Figure 7.7 shows how spectacles can help correct the defects of myopia and presbyopia. Short-sighted people are prescribed spectacles with diverging lenses (Fig. 7.7*a*), while long-sighted people are prescribed collecting lenses (Fig. 7.7*b*). The figure makes it clear that

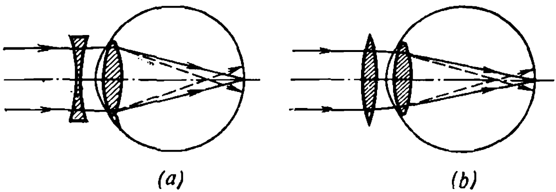


Fig. 7.7.

the spectacles change the paths of the rays inside the eye so that they become the same as they are in normal eyes (compare it with Fig. 7.6*a*). Now both short-sighted and long-sighted people have eyes that are accommodated for infinity when the ciliary muscle is relaxed.

The Use of a Lens System to Increase the Angle of View. At the beginning of the chapters we noted that optical instruments forming the virtual image of an object are termed optical aids. Their chief function is *to increase the angle of view.* Assume that we are viewing a small object with our naked eyes, for instance, a letter of height l in the text, while the book is placed at the normal viewing distance L from the eyes. The angle of view α_1 is determined by the simple for-

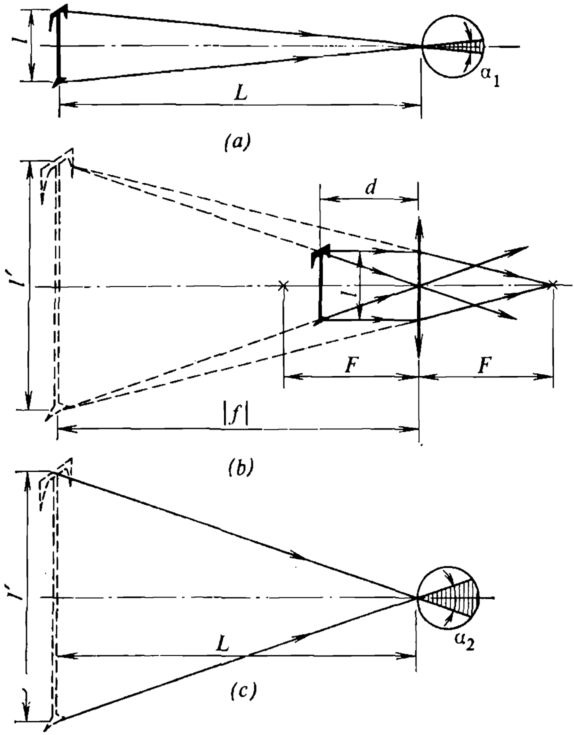


Fig. 7.8.

mula (see Fig. 7.8*a*):

$$\alpha_1 = \frac{l}{L}. \tag{7.1}$$

If $l = 2$ mm and $L = 25$ cm, we get $\alpha_1 = 0.008$, which is 27′. Assume now that there is a collecting lens with the focal length F right in front of our eyes, and we are viewing the magnified virtual image of the page (the lens is being used as a *magnifying glass*). The book must be at a distance d from the lens (or in this case from the eye) and positioned so that the distance from the image to the lens (to the eye) equals the normal viewing distance L. Taking $|f| = L$ and using (6.8), we have

$$d = \frac{FL}{F+L}. \qquad (7.2)$$

Figure 7.8*b* shows the virtual image of the object (a letter) in the collecting lens. It follows from the figure that the linear dimension l' of the letter's image is related to its real height l thus:

$$l' = l\,\frac{F+L}{F}. \qquad (7.3)$$

By moving the eye close to the lens and looking at the image, we can see the letter at an angle of view determined by the formula (see Fig. 7.8*c*)

$$\alpha_2 = \frac{l'}{L}. \qquad (7.4)$$

Substituting (7.3) into (7.4), we find that

$$\alpha_2 = l\,\frac{F+L}{FL},$$

or, taking (7.1) into account, that

$$\alpha_2 = \alpha_1\,\frac{F+L}{F}. \qquad (7.5)$$

Let F equal 10 cm. It then follows from (7.5) that $\alpha_2 = 3.5\,\alpha_1 = 0.028$, or $1.5°$. Thus, the use of a magnifying glass makes it possible to increase by 3.5 times the angle of view at which the retinal image of the letter is formed. According to (7.2), the book must be $d = 7$ cm from the eyes.

The magnification of the angular dimensions of the image (and therefore the angle of view)

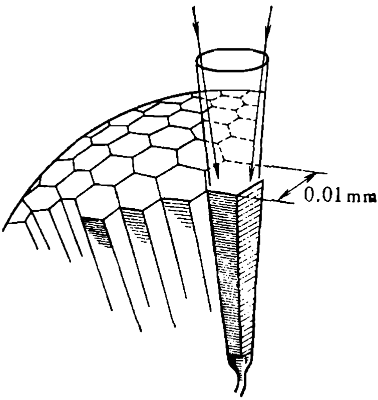


Fig. 7.9.

was mentioned in the previous chapter when we discussed the paths of rays through a Galilean telescope.

The Compound (Faceted) Eyes of Insects. To conclude this chapter, we shall consider the peculiarities of *insects' eyes.* They have a very complicated structure and are made up of very many minute hexagons or "*facets*". The number of facets in one eye is quite large—in an ant's eye it is 100, while in the eye of a dragonfly it can be 20,000. The linear dimension of a facet on the surface of the eye is about 0.01 mm. Each facet acts as a lens (a crystalline lens) for a separate light-sensitive element, an *ommatidium*. Figure 7.9 shows a diagram of the structure of a compound eye. We have singled out one facet. The cone defines the confines within which the rays of light can reach the ommatidium of this facet. On the whole, the eye of an insect has a very large angle of view, but the images of different objects are formed in different ommatidia. As a result, the general image is rough and mosaic-like. Still, this structure has its own advantages for it gives a good view of fast-moving objects. A moving object is viewed consecutively by the different ommatidia of the insect's eye and this also gives the insect the ability to estimate instinctively the speed of the object.

Chapter Eight

Double refraction in crystals

Bartholin's Discovery of Double Refraction in a Crystal of Iceland Spar. "There is in Iceland, the island in the North sea located at 66° of northern latitude, a special sort of crystal, or transparent stone", which is quite peculiar in its shape and other properties. It is particularly famous for the strange manner in which it refracts light". These words are taken from Christiaan Huygens' book "Treatise on Light" published in 1690. Twenty years before, in 1669, the Danish scientist Erasmus Bartholin published a book entitled "Experiments With Crystals of Iceland Spar Which Refract Light in a Peculiar and Strange Manner" and discussed the discovery of a new physical phenomenon—the *double refraction of light* (the term *birefringence* is also used).

As he was observing the refraction of light in a crystal of Iceland spar (calcite, $CaCO_3$), Bartholin discovered to his amazement that a light ray was divided into two rays inside the crystal. One of the rays obeyed the law of refraction, while the other did not. A ray like the first one was referred to as the *ordinary ray*, while the other ray became known as the *extraordinary ray* (Bartholin also called it the "variable" ray). Figure 8.1*a* shows double refraction when the angle of incidence of the light upon the crystal is

α. The angles of refraction for the ordinary and the extraordinary rays are designated β_o and β_e, respectively. The ratio $\sin\alpha/\sin\beta_o$ is a constant which for Iceland spar equals 5/3. As for the ratio $\sin\alpha/\sin\beta_e$, it is extraordinarily inconstant. First of all, it depends on the angle of incidence. Secondly, if the angle of incidence

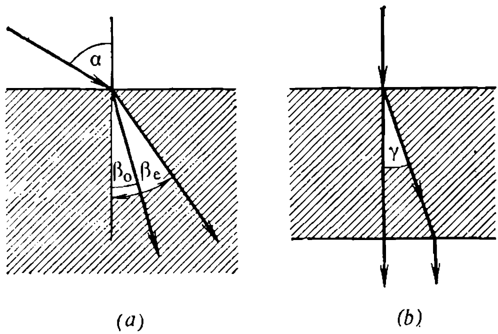


Fig. 8.1.

remains the same, it depends on the choice of the refracting edge of the crystal. It was not for nothing that Bartholin called this ray "variable".

Bartholin discovered that the double refraction occurred even when the incident ray of light was perpendicular to the crystal surface (see Fig. 8.1*b*). In this case, the ordinary ray was not refracted, while the extraordinary ray makes an angle γ with it. Curiously, the two rays are parallel to each other after emerging from the crystal. Bartholin also discovered that there is a direction in the crystal in which a ray of light may be propagated without birefringence.

Double refraction seemed extremely mysterious to many of Bartholin's contemporaries. The behaviour of the extraordinary ray and its apparent contradiction to the law of refraction seemed inexplicable. However, it was soon explained by Bartholin's contemporary, Christiaan Huygens. He became interested in Bartholin's discovery and carried out independent investigation of double refraction in Iceland spar and quartz. The explanation Huygens came up with is still accepted by modern text-books on optics.

Crystals as Optically Anisotropic Media. Before going into detail, we remind the reader of a generally acknowledged truth: a crystal is *an anisotropic medium*. The word "anisotropic" means that the properties of the crystal depend on the direction in which the light is propagated. We have until now assumed that all media are isotropic and this is right when we deal with glass, water, or air, but is not true for transparent crystals. Different crystals manifest anisotropic properties to different degrees. There is a large group of crystals, Iceland spar being just one of this numerous group, that have a direction along which the crystals display no anisotropy. This direction is termed the *optic axis* of a crystal and this kind of crystals are referred to as *uniaxial*. The optical characteristics of a uniaxial crystal are the same along all the directions which make similar angles with the optic axis. These directions are shown with arrows in Fig. 8.2*a*, where OO_1 is the optic axis and θ is this angle. The optical characteristics of the crystal change as θ changes; they are different in the directions shown in Fig. 8.2*b*. The concept of and term "optic axis" was introduced by Huygens. It was Huygens who introduced the term "the principal section" which is widely used in modern optics. This is the plane that passes through the optic axis. We usually consider the *principal section* passing through the optic axis and a light ray.

Huygens' Explanation of Double Refraction in the "Treatise on Light". Ordinary and Extraordinary Waves of Light. The explanation of double refraction is contained in the following extract from Huygens' "Treatise on Light": "As there were two different refractions, I conceived that there were also two different emanations of waves of light... I attributed to this emanation of waves the regular refraction which is observed in this stone, by supposing these waves to be ordinarily of spherical form. As to the other emanation which should produce the irregular refraction, I wished to try what elliptical waves, or rather spheroidal waves would do... It seemed to me that disposition or regular arrangement of these particles could contribute to form spheroidal waves, nothing more being required for this than that the successive movement of light should spread a little more quickly in one direction than in the other..."

Assume that a point light source is located at a point O of the crystal. According to Huygens and our modern theories, it will generate two different light waves. These waves will differ in the forms of their *wave-fronts*. Recall that a wave-front is a locus which is reached by the light emanating from a point light source within a certain period of time. The surface of one wave is *spherical* (in the case of the ordinary wave), while the surface of the other wave is an *ellipsoid of revolution* around the optic axis of the crystal drawn through the point O (in the case of the extraordinary wave)*. The principal sections of these wave-fronts look like a *circle* and an *ellipse*, respe-

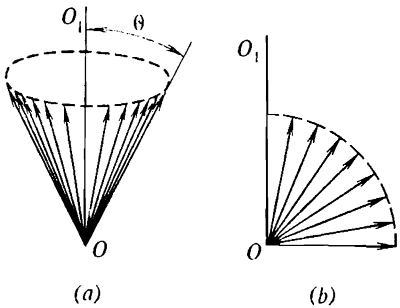


Fig. 8.2.

ctively (see Fig. 8.3). For the ordinary wave, the crystal can be considered an isotropic medium; the wave propagates with the same velocity in all directions. We shall call v_0 the velocity of the ordinary wave; it is determined by the radius of the circle shown in Fig. 8.3. We can conclude from the figure that when the light propagates along the optic axis OO_1, the extraordinary wave travels with the same velocity v_0 as the ordinary wave, whereas when the light travels in the

* Note that the optic axis is not to be understood to be a straight line passing through particular points in the crystal. It is rather a certain direction; therefore an imaginary optic axis can be drawn through any point.

direction lying at right angles to the optic axis, the velocity of the extraordinary wave is different. It is measured by the length of the straight line OA; we designate it v_e.

Note that in this case $v_e > v_o$. These uniaxial crystals are called *negative* (Iceland spar is a

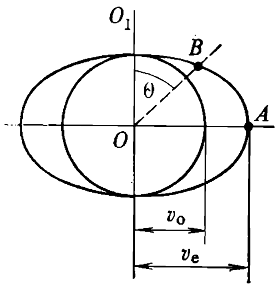


Fig. 8.3.

negative uniaxial crystal). There are also uniaxial crystals for which $v_e < v_o$, and these are called *positive*.

The reader should not think that the straight line OB (see Fig. 8.3) is the velocity of the extraordinary wave when light propagates at the angle θ to the optic axis. The point is that if the angle is neither $\theta = 0$ (180°) nor $\theta = 90°$, the notion "the direction of light wave propagation" becomes ambiguous and needs clarification. It is double refraction that is directly responsible for it. The necessary clarifications will be given further on and it will be shown that at $\theta = 0$ (180°) and 90° double refraction is not manifested.

Huygens' Construction. Wave and Ray Velocities. Let us return again to Huygens' work. "Supposing, then, these spheroidal waves besides the spherical ones, I began to examine whether they could serve to explain the phenomena of the irregular refraction, and how by these same phenomena I could determine the figure and position of the spheroids: as to which I obtained at last the desired success...". We shall now discuss the use of *Huygens' principle* (see Chapter One) proceeding from the assumption that there are *two types* of light waves. Let a parallel beam of light of the width d fall normally on the surface MN of the crystal (see Fig. 8.4). We assume that the optic axis OO_1 of the crystal is at an angle θ with the direction of the incident ray. The moment the plane wave-front touches the surface of the crystal all the points of AB become sources of two types of secondary wavelets, i.e. spherical and elliptical wavelets. The wave-fronts of the elliptical wavelets are appropriately arranged with respect to the refracting face MN of the crystal. Line *1* is the cross section of the envelope of the spherical wave-fronts after a certain period of time Δt, and this envelope is actually a plane wave-front of the ordinary wave which is propagated from MN within the crystal. Line *4* is the envelope of the elliptical wave-fronts after the same time period Δt, and this envelope is the plane front of the extraordinary wave. All the points of the section A_1B_1 of line *1* can be regarded as sources of secondary spherical wavelets. On the other hand, all the points of the segment C_1D_1 of line *4* are the sources of secondary elliptical wavelets. Successive positions

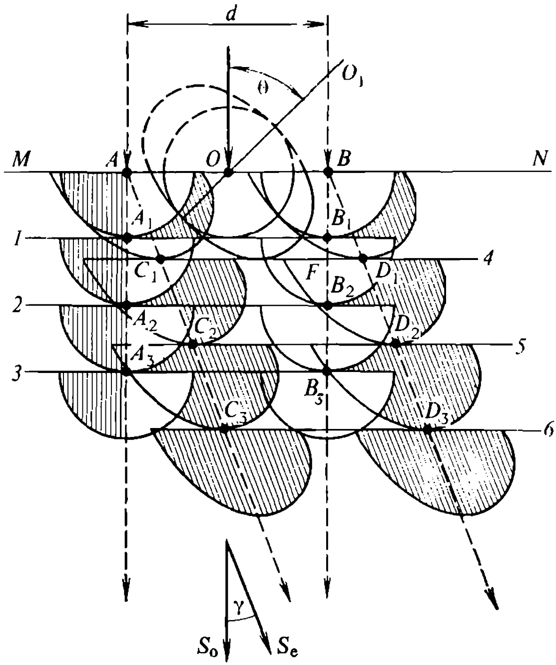


Fig. 8.4.

of the ordinary wave-front are shown in the figure by the segments A_1B_1, A_2B_2, and A_3B_3; the positions of the extraordinary wave-fronts are shown by C_1D_1, C_2D_2, and C_3D_3.

Considering the figure, we can arrive at a number of important conclusions. Firstly, we see that a light ray really splits into two rays, ordinary and extraordinary. An ordinary ray is propagated normally to the MN boundary, while the extraordinary ray travels at an angle to it. The directions of these rays are shown in the figure by the vectors S_0 and S_e; the vectors are at an angle γ to each other. It is quite obvious that the value of γ depends on the degree of elongation of the ellipse which describes the surface of the secondary wavelets generating the extraordinary ray (in other words, it depends on the ratio v_e/v_0), as well as on the position of the ellipse with respect to the refracting surface of the crystal.

Secondly, we see that the fronts of both the ordinary and extraordinary waves always remain parallel to the MN boundary (the segments A_1B_1, A_2B_2, A_3B_3, C_1D_1, C_2D_2, and C_3D_3). It follows from this that on reaching the exit face of the crystal, which we shall consider to be parallel to the entry face, each ray will simultaneously generate secondary wavelets along it. If we keep in mind that secondary wavelets are generated in the air and that is why they are spherical for both rays, we can easily see why both rays are propagated perpendicularly to the surface of the crystal after leaving it.

Thirdly, we begin to realize that the notion "the direction of light propagation within a crystal" certainly needs clarification. When we speak about the velocity of light in a medium, we usually have in mind the velocity of the propagation of a wave of light. It is essentially the velocity of the wave-front and the vector of the velocity lies at right angles to the front surface, at each point. In the case shown in Fig. 8.4 the velocities of the ordinary and extraordinary waves will have the same direction at right angles to the surface MN of the crystal; however, the velocities will be unequal, the former being determined by the segment AA_1 (v_0), and the latter by the segment BF. Along with the velocities of the ordinary and extraordinary waves, we should consider in this case the "ray velocities", which characterise the propagation of light. The directions of these velocities are the same as the directions of the respective light rays (the vectors $\mathbf{S}_0$ and $\mathbf{S}_e$). The ray velocity of the extraordinary ray is determined by the segment BD_1. As for the ray velocity of the ordinary ray, it coincides with the velocity v_0 of the ordinary wave.

Thus, when talking about the propagation of light in a medium, it is necessary to distinguish between the *wave velocity* and the *ray velocity*. The former is the velocity of the propagation of the wave-front within a crystal, i.e. the propagation of the constant phase (this explains why the term the *phase velocity* can also be applied here). The latter is the velocity of the propagation of the light field's energy within the crystal. In the isotropic medium, these velocities are the same. In a uniaxial crystal these velocities are the same for the ordinary ray. As for the extraordinary ray, the velocities are the same only if light is propagated along the optic axis or at right angles to it. Now the reader should understand better the reservations we made concerning Fig. 8.3. The segment OB represented in this figure is actually the ray velocity of the extraordinary ray of light and not the velocity of the extraordinary wave. Only if $\theta = 0$ (180°) or 90°, the velocity of the extraordinary ray equals the velocity of the extraordinary wave.

Using Fig. 8.4, we considered the case when the axis OO_1 was at a certain angle to the refracting face and a ray of light is incident on it normally. Assume now that a light ray is incident on the refracting face of the crystal at a

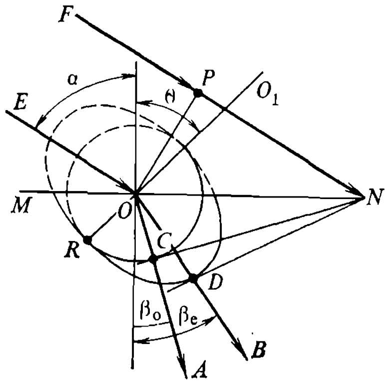


Fig. 8.5.

certain angle α and the optic axis remains in the same place with respect to the refracting face. Figure 8.5 shows the way to trace the ordinary and the extraordinary rays in this case. The procedure of making up the drawing is analogous to the one used to create Fig. 1.7 (in Chapter One). The only difference will be that now O is the source of both a spherical and an elliptical wave. The wave-front of the elliptical wave is arranged in a certain way with respect to the refracting surface MN. The figure shows two parallel rays (*EO* and *FN*) incident on the boundary *MN* of the crystal, and *OP* is the plane front of the incident light beam. The centre *N* is positioned so that the relation $PN/OR = c/v_0$ is satisfied, where c is the velocity of light in the air. We draw a tangent *NC* from *N* to the spherical wave-front and a tangent *ND* from *N* to the elliptical surface. The line *NC* is the front of the ordinary wave, while *ND* is the front of the extraordinary wave. The ray drawn from the point *O* through the point of tangency *C* is the ordinary ray, whereas the ray drawn from the point *O* through the point of tangency *D* is the extraordinary ray.

Comparing Figs. 8.4 and 8.5 with Fig. 8.1, we see that Huygens' idea that there are two types of waves (spherical and elliptical) within a crystal, combined with the principle of drawing the resulting wave-fronts as the envelopes of the wave-fronts of the secondary wavelets, really does account for the double refraction discovered by Bartholin.

Figure 8.6 shows two important cases. In both cases the light is normally incident on the refracting surface of the crystal. The optic axis of the crystal is in one case perpendicular to the *MN* surface of the crystal (Fig. 8.6*a*), and in the other case runs parallel to it (Fig. 8.6*b*). When the optic axis is perpendicular to *MN*, the light propagates along the axis and both the ordinary and the extraordinary waves travel with the same velocity v_0. When the optic axis is parallel to *MN*, the light propagates perpendicularly to the axis. It is easy to see that in this case the light beam is not decomposed but the velocities of the ordinary and extraordinary waves are different (v_o and v_e, respectively).

If a beam of light is incident on the refracting face of the crystal at a certain angle, double refraction can be observed even if the optic axis is perpendicular to this face or runs parallel to it (Fig. 8.7).

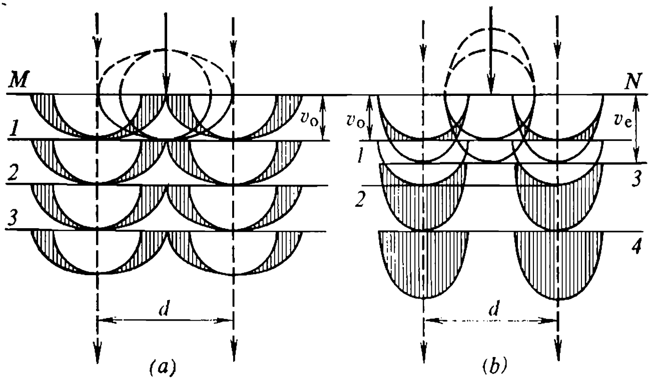


Fig. 8.6.

Pay attention to the fact that in all the three figures (Figs. 8.5, 8.7*a*, and 8.7*b*) the refracting angle β_0 of the ordinary ray is the same if the angle of incidence α remains the same in all the three cases. This is quite natural, because the position of the optic axis does not influence the ordinary ray because it travels through a crystal as if it were an isotropic medium. In other words, unlike the extraordinary ray, the ordinary one obeys the law of refraction: the ratio of the sines of the angles of incidence and refraction in this case is constant.

Huygens' Experiments with Two Crystals (On the Verge of Discovering the Polarization of Light). No matter how obvious his success in explaining double refraction, Huygens deemed it necessary to carry out further experiments. *This resulted in his being on the verge of yet another*

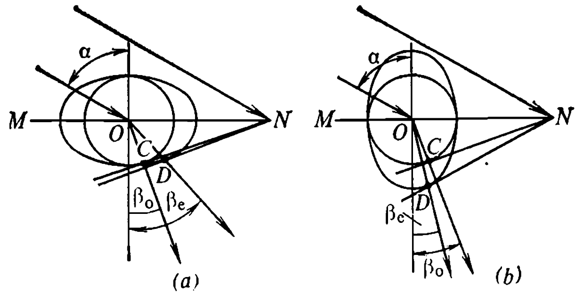


Fig. 8.7.

discovery. In this "Treatise on Light" Huygens wrote: "I will add one more marvellous phenomenon which I discovered after having written all the foregoing. For though I have not been able till now to find its cause, I do not for that reason wish to desist from describing it, in order to give opportunity to others to investigate it. It seems that it will be necessary to make still further suppositions, besides those which I have made..."

Huygens had become interested in what would happen if the two light beams resulting from the double refraction within a crystal of Iceland spar are passed through a second crystal. Huygens knew that light incident on the crystal makes the latter a source of both secondary spherical and elliptical wavelets which cause the ordinary and the extraordinary refraction, respectively. Thus, he reasoned that each of the two rays after emerging from the first crystal and incident on the second crystal must also make it a

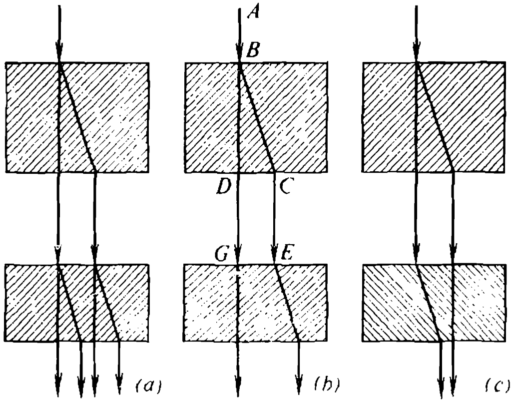


Fig. 8.8.

source of spherical and elliptical secondary wavelets. Hence after leaving the second crystal, there must be four bundles of rays instead of two. This means that we would see the picture shown in Fig. 8.8*a* (for the sake of simplicity, we consider the light normally incident on the crystal; we can thus tell immediately the ordinary ray from the extraordinary one). The experiment Huygens staged using two crystals of Iceland spar showed that what is shown in Fig. 8.8*a* is actually observed, but not always!

Huygens discovered that if the principal sections of the crystals are parallel to each other, we observe the picture shown in Fig. 8.8*b*. It was amazing. Huygens wrote: "Now it is marvellous why the rays *CE* and *DG*, incident from the air on the lower crystal, do not divide themselves the same as the first ray *AB*. One would say that it must be that the ray *DG* in passing through the upper piece has lost something which is necessary to move the matter which serves for the irregular refraction; and that likewise *CE* has lost that which was necessary to move the matter for regular refraction: but there is yet another thing which upsets this reasoning".

This thing was the result of the experiments when the principal sections of the crystals were at right angles. Then the refraction in Fig. 8.8*c* can be seen. Huygens' reasoning concerning this was that if we place two crystals so that the planes forming the principal sections cut each other at right angles, the ray obtained from regular refraction will only undergo irregular refraction and vice versa, the ray leaving the first crystal after irregular refraction will only undergo regular refraction. Rotating the lower crystal with respect to the upper one, Huygens discovered that even when both rays are divided in the second crystal after leaving the first, the situation remains unclear. It turned out that the relation between the intensities of the ordinary and extraordinary rays appearing in the second crystal changed depending on the angle of rotation of the lower crystal.

Turning the results of the experiment over in his mind, Huygens arrived at an important conclusion. If the pattern of refraction depends on the orientation of the crystals with respect to each other, "... it seems that one is obliged to conclude that the waves of light, after having passed through the first crystal, acquire a certain form or disposition in virtue of which when meeting the texture of the second crystal, in certain positions, they can move the two different kinds of matter which serve for the two species of refraction; and when meeting the second crystal in another position are able to move only one of these kinds of matter...". Today we do not accept Huygens' reasoning concerning the different kinds of matter able to move since it has no scientific value. However, we can extract from his words that after passing through the crystal the light waves "acquired a certain form or disposition", for this is an anticipation of the *polarization of light.*

Thus, Christiaan Huygens had been on the verge of discovering the polarization of light. However, he failed to make the discovery itself. This can be accounted for by Huygens' abiding by the wave theory of light and his belief that light waves were longitudinal like sound waves. Polarization of light as it is now known is a characteristic of transverse waves. That is why Huygens, who was an extremely conscientious scientist, had to conclude the quote above thus: "... but to tell how this occurs, I have hitherto found nothing which satisfies me".

Newton's Interpretation of Huygens' Results. It was Isaac Newton who actually made the discovery. After analysing Huygens' experiments on double refraction in two crystals, Newton came to the decisive conclusion that if the ordinary ray is symmetrical around the direction of propagation, the rays which passed through a crystal *lack this symmetry*. In his work "Optics" Newton says that "every ray can be considered as having four sides, or quarters," and when the ray is rotated around its axis, these sides change their position with respect to the crystal. Newton went on to define: "One and the same ray is here refracted, sometimes after the usual, and sometimes after the unusual manner, according to the position which its sides have to the crystals". How should we account for the lack of axial symmetry in a ray of light? Newton answered the question using his *corpuscular theory* of light. He held that the corpuscles had "different sides".

Malus' and Brewster's Investigations. Newton's *idea of polarization of light* remained unnoticed for about a century. In 1808 the Paris Academy of Sciences held a contest for the best mathematical theory of double refraction. The prize was awarded to the French engineer Etienne Malus (1775-1812) for his work "The Theory of Double Refraction of Rays of Light in Crystalline Substances".

Malus had become interested in double refraction after he observed the reflection of the sun in the windows of the Luxemburg Palace through a crystal of Iceland spar and noticed that there was only one image of the sun instead of two. That was similar to the results of Huygens' famous experiments with two crystals of spar. Malus staged a special experiment and found out that sunlight reflected from water surface at 53° has the same property as light passed through a crystal of Iceland spar, the water surface behaving like the principal section of the crystal. To account for his discovery and double refraction, Malus fell back on Newton's corpuscular concept of light. He held that the corpuscles in sunlight were in disarray, whereas they are oriented in a specific manner as they pass through a crystal or in the case of a proper reflective angle. Malus referred to a light ray which was oriented in a specific manner as *polarized*. Since that time the term "polarization of light" has been used in optics.

The English scientist David Brewster (1781-1868) carried Malus' investigations further. He established the law which bears his name and which reads as follows: if the reflected ray is at 90° to the refracted ray, the refracted ray is completely polarized and the reflected ray has maximum polarization (note that the polarization of the rays will be opposite). The angle of incidence at which this law holds true is called the *Brewster angle*.

Polarization of Light. Today polarization of light is explained both by the corpuscular (quantum) theory of light and by the wave theory. It is now generally accepted that electromagnetic waves, light waves being a special case of electromagnetic waves, are *transverse*. That is why the ideas concerning the polarization of these waves are quite natural.

Note that light waves are *electromagnetic waves* whose lengths fall within the optical range. The oscillations in an electromagnetic wave are effected by the vector of electric intensity **E** (the electric vector) and the vector of magnetic induction **B** (the magnetic vector). These vectors are at right angles to each other and to the direction of the ray. In modern optics the polarization of light is considered to be dependent on the direction of the electric vector **E**. If the **E** oscil-

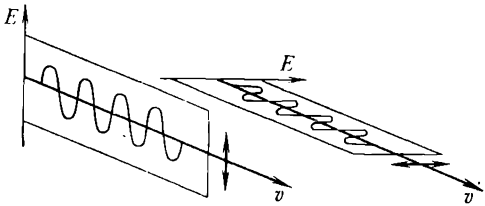


Fig. 8.9.

lations only occur in one plane, we have *plane-polarized* light and the plane is referred to as the *plane of polarization*. Figure 8.9 shows two cases corresponding to this type of polarization together with their planes of polarization. These planes are perpendicular to each other. There are more complicated kinds of polarization of light but we have no intention of dwelling on them in this book. However, note that polarization always means that there is a certain *ordering* in the direction of the **E** vector and, consequently, in that of the **B** vector too. If there is no such ordering we have *nonpolarized* light.

By taking the polarization of light into consideration, we can understand the phenomenon of double refraction of light in a crystal in all its totality. A nonpolarized light ray is divided into two plane-polarized rays by the crystal, i.e. the ordinary and the extraordinary rays. *The* *ordinary ray is polarized perpendicularly to the plane of the crystal's principal section, whereas the extraordinary ray is polarized in the plane of the principal section*, and these rays leave the crystal having acquired these properties. Thus having passed through the crystal, there are two plane-polarized rays whose planes of polarization are perpendicular to each other. Using Huygens' expression, we can say that after emerging from the crystal the rays had "acquired a certain form or disposition". Applying Newton's way of describing the phenomenon, we shall say that the rays are characterized by different "positions of sides" with respect to the principal section of the crystal. Figure 8.10 is used to explain what has just been said, where OO_1 is the optic axis of the crystal, S is the plane of the principal section, AB is the incident light ray, BD is the ordinary light ray, BC is the extraordinary ray, and, DG and CE are the plane-polarized rays that emerge from the crystal. Short arrows oriented in an appropriate way are used to show the directions of the polarization of the rays.

There should be no difficulty now in interpreting the results of Huygens' experiments with two crystals. In the case of Fig. 8.8*b* the principal cross section of the second crystal is parallel to the principal section of the first one. That is why the ray, having been ordinarily refracted by the first crystal, enters the second crystal remaining polarized at right angles to the plane of the principal section. It is obvious that this ray will be the ordinary one in the second crystal. The same applies to the extraordinary ray. It is polarized to the plane of the principal section of the first crystal and remains such with respect to the plane of the principal section of the second crystal.

In the case in Fig. 8.8*c* the principal section of the second crystal lies at right angles to the

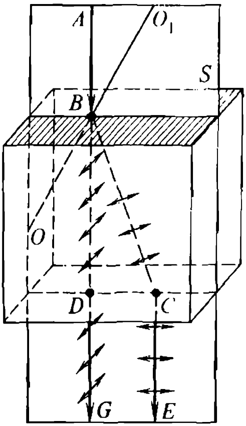


Fig. 8.10.

principal section of the first one. That is why the ray polarized at right angles to the plane of the principal section of the first crystal is polarized in the plane of the principal section of the second crystal, and vice versa. This means that the ray which exhibits the properties of the ordinary one in the first crystal becomes extraordinary in the second crystal, while the extraordinary ray which emerged from the first crystal becomes the ordinary one in the second crystal.

Consider a general case when the planes of the principal sections of the crystals are at some angle φ to each other. Looking along the path of the incident light ray, we see the planes of the principal sections as straight lines. In Fig. 8.11 the line AB shows the principal section of the first crystal, CD shows the principal section of the second crystal. The electric vectors $\mathbf{E}_{o1}$ and $\mathbf{E}_{e1}$ describe the o-

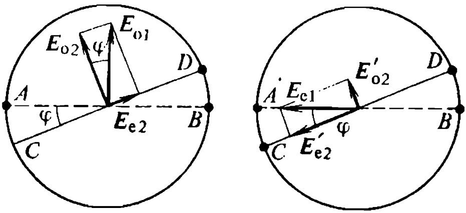


Fig. 8.11.

rdinary and the extraordinary rays leaving the first crystal, respectively. Note that $\mathbf{E}_{o1}$ cuts AB at right angles, whereas $\mathbf{E}_{e1}$ is parallel to it. Upon entering the second crystal, each ray is again divided into an ordinary ray and an extraordinary ray. To determine the electric vectors of the new rays, we have to resolve the vectors $\mathbf{E}_{o1}$ and $\mathbf{E}_{e1}$ into the direction CD and the direction normal to CD (see Fig. 8.11). Designate the electric vectors of the ordinary and the extraordinary ray which appear after the ray with $\mathbf{E}_{o1}$ has passed through the second crystal as $\mathbf{E}_{o2}$ and $\mathbf{E}_{e2}$, respectively, and designate the same vectors for the ray with $\mathbf{E}_{e1}$ as $\mathbf{E}'_{o2}$, and $\mathbf{E}'_{e2}$. As can be seen from the figure,

$$\left.\begin{aligned} \mathbf{E}_{o2} &= \mathbf{E}_{o1}\cos\varphi, \quad & \mathbf{E}_{e2} &= \mathbf{E}_{o1}\sin\varphi, \\ \mathbf{E}'_{o2} &= \mathbf{E}_{e1}\sin\varphi, \quad & \mathbf{E}'_{e2} &= \mathbf{E}_{e1}\cos\varphi. \end{aligned}\right\} \tag{8.1}$$

The relation between the electric vectors of the rays leaving the second crystal and the relation between the intensities of these rays will change depending on the value of the angle φ. In this way we can account for the change in the intensities of the rays Huygens had observed when he rotated one crystal with respect to the other. It is clear from (8.1) that if $\varphi \to 0$ and $\varphi \to 90°$ the electric vectors and, consequently, the intensity of rays change in such a way that only two rays instead of four emerge from the second crystal. If $\varphi = 0$, relations (8.1) will look as follows:

$$\mathbf{E}_{o2} = \mathbf{E}_{o1}; \quad \mathbf{E}_{e2} = 0; \quad \mathbf{E}'_{o2} = 0, \quad \mathbf{E}'_{e2} = \mathbf{E}_{e1}, \qquad (8.2)$$

and if $\varphi = 90°$, we have

$$\mathbf{E}_{o2} = 0, \quad \mathbf{E}_{e2} = \mathbf{E}_{o1}; \quad \mathbf{E}'_{o2} = \mathbf{E}_{e1}, \quad \mathbf{E}'_{e2} = 0. \qquad (8.3)$$

These two cases were dealt with thoroughly above.

Dichroic Plates and Polarization Prisms. Double refraction of light in crystals has been widely applied in practice *to obtain polarized beams of light*. Polarizers of light can be of several types. We shall discuss just two—*dichroic plates*, also called polaroids, and *polarization prisms*.

A dichroic plate is made of a double-refracting crystal which absorbs one of the rays, the ordinary one, for example, more intensely than the other. The dependence of the absorption of light on its polarization is called *aichroism*, which gives the plate its name. A plate made of *tourmalin* is an example of a dichroic plate. If the plate is 1 mm thick, the ordinary ray is completely absorbed. A nonpolarized light beam incident on a dichroic plate becomes plane-polarized by passing through it and its plane of polarization coincides with the principal section of the plate.

We will discuss the *Nicol prism* as an example of a polarization prism. It was designed by the English physicist William Nicol in 1820. It is cut from a crystal of Iceland spar in the manner shown in Fig. 8.12 (in this case OO_1 is the optic axis of the crystal). The crystal is cut along the line AA_1 and then the two pieces are cemented together along the same line using Canada balsam. The geometry of the prism and the glue are chosen so that the extraordinary ray passes through the prism, while the ordinary one is totally reflected from the interface between the crystal and the Canada balsam.

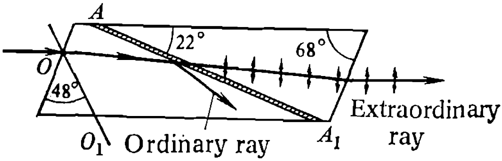


Fig. 8.12.

There are quite a number of different polarization prisms. Another example is the *Glan-Foucault prism* shown in Fig. 8.13. It is combined of two prisms made of Iceland spar and separated from

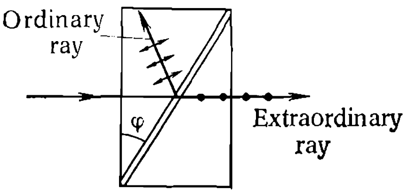


Fig. 8.13.

each other by an air gap. The optic axes of the two prisms are at right angles to the incident ray and to the plane of the figure, and $\varphi = 38°30'$. The extraordinary ray is transmitted through the prism while the ordinary one is reflected at the interface between the crystal and the air.

Rotation of the Plane of Polarization of Light in a Half-Wave Plate. Using the phenomenon of double refraction, it is not only possible to obtain polarized light but also to *control its polarization.* The easiest way of controlling polar-

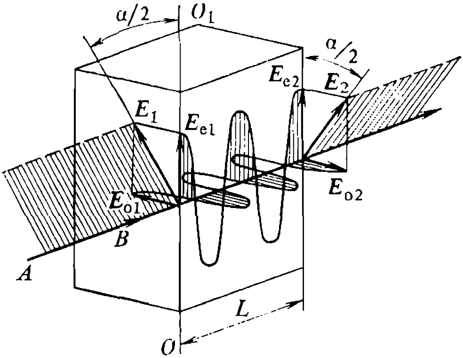


Fig. 8.14.

ization is to rotate the plane of polarization of the plane-polarized beam of light through some angle α.

Let a beam of light be normally incident on a crystalline plate whose optic axis is perpendicular to the beam. Figure 8.14 shows the cross section of the plate by the plane of the principal section; here OO_1 is the optic axis, and AB is the incident beam of light. The plate should be oriented so that the optic axis OO_1 is at an angle $\alpha/2$ to the electric vector $\mathbf{E}_1$ of the incident beam. The vector $\mathbf{E}_1$ has the components $\mathbf{E}_{o1}$ and $\mathbf{E}_{e1}$ corresponding to the ordinary and the extraordinary ray. In the case under discussion the two rays coincide but the velocities of the ordinary and the extraordinary waves are different, viz. v_o for the ordinary wave and v_e for the extraordinary one (the reader should look back at Fig. 8.6*b*). Hence, the wavelengths of the ordinary and the extraordinary rays will also be different. The wavelength λ is the distance covered by a wavefront during the wave period $T = 1/\nu$, where ν is the frequency of emission. Thus for the ordinary and the extraordinary waves inside the plate we can write

$$\lambda_o = \frac{v_o}{\nu}, \quad \lambda_e = \frac{v_e}{\nu}. \tag{8.4}$$

Assume further that the thickness L of the plate is selected so that the number of the ordinary wavelengths (L/λ_o) the plate accommodates is 1/2 more than the number of the extraordinary wavelengths (L/λ_e) it can accommodate, i.e.

$$\frac{L}{\lambda_o} - \frac{L}{\lambda_e} = \frac{1}{2}. \tag{8.5}$$

As can be seen from Fig. 8.14, the orientation of the vectors $\mathbf{E}_{o2}$ and $\mathbf{E}_{e2}$ with respect to each other when the waves leave the plate is such that the vector $\mathbf{E}_2 = \mathbf{E}_{o2} + \mathbf{E}_{e2}$ is shifted with respect to the optic axis by the same angle $\alpha/2$ but in the direction opposite to the vector $\mathbf{E}_1$. Thus at the end $\mathbf{E}_2$ is shifted with respect to $\mathbf{E}_1$ by the required angle α.

The plate of this kind is called *half-wave* plate because it shifts the ordinary and the extraordinary wave half a wavelength with respect to each other, the original and final waves are shifted in relation to each other in phase by the value π. By substituting (8.4) into (8.5), we get the following formula for the thickness L of a half-wave plate:

$$L = \left[2\nu \left(\frac{1}{v_o} - \frac{1}{v_e} \right) \right]^{-1}. \tag{8.6}$$

Let $\nu = 4.5 \times 10^{14}$ Hz (red light). At this frequency, we have for Iceland spar $v_o = 1.81 \times \times 10^8$ m/s and $v_e = 2.02 \times 10^8$ m/s. Substituting these numbers into (8.6) gives us $L = = 2$ μm.

This result corresponds to the minimum thickness of a half-wave plate made of Iceland spar. The thickness of a real half-wave plate can be $2N + 1$ times greater than L, where N is a whole number.

Consider a half-wave plate made of quartz. Unlike Iceland spar, quartz is a positive uniaxial crystal, and $v_e < v_o$. Hence (8.6) should be replaced by

$$L = \left[2\nu \left(\frac{1}{v_e} - \frac{1}{v_o} \right) \right]^{-1}. \tag{8.7}$$

For quartz, if $\nu = 4.5 \times 10^{14}$ Hz we have $v_o = = 1.945 \times 10^8$ m/s and $v_e = 1.934 \times 10^8$ m/s. Substituting these values into (8.7) gives us $L = 37$ μm.

Chapter Nine

What is fibre optics

This chapter is about one of the new branches of optics which took shape in the late 1950's. Its rapid development nowadays is closely connected with the progress achieved in optical communication, optical and electronic data processing systems, and so on.

We are used to the idea that light travels along straight lines. True, we know that in a medium with a gradually changing refractive index, for example, in the Earth's atmosphere, a light ray curves (see Chapter Two); nevertheless, we invariably associate the path of a light ray with a straight line or, in a more general sense, with a broken line made up of several linear sections. But can a light ray be wound round one's arm like, for instance, a length of rope or cord? The question may seem weird. However, this too is feasible if the light is "locked" inside a transparent flexible *optical fibre.*

A Luminous Jet of Water. The propagation of light along a curved fibre can be illustrated by the following experiment. Figure 9.1 shows a diagram of it. A vessel is filled with water; in the lower part of it there is a hole through which water can flow out in a steady jet. There is a light source behind the outlet, the light being focused to the outlet. Light gets inside the water jet and seems to run along it. One gets the impression that the water carries the light along, while the jet appears to be luminous from the inside.

Many of us have admired the beautiful sight of fountains lit from below in the evening. With

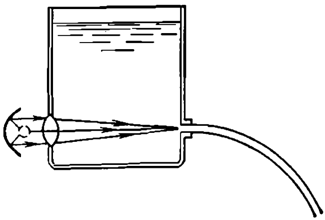


Fig. 9.1.

proper lighting the jets of fountains seem to be luminous, like the jet in the experiment.

A luminous jet of water was the precursor of an optical fibre. In both cases we are dealing with the phenomenon of *total internal reflection*. This is what makes the light run along the fibre obediently following all its curves.

Light Rays in Straight and Curved Cylindrical Fibres. Consider the following few problems which will give you an idea how light travels along optical fibre. For the sake of simplicity, consider only those rays which lie in a plane passing through the axis of the fibre; this kind of rays are termed meridional. The path of a ray of light in the fibre depends on the incidence of the ray on the end face of the fibre. Rays incident on the end face at large angles will not be caught by the fibre, they will emerge through the lateral surface.

Solve the following problem. *Find the tolerance angle of incidence α of the ray on the end face of a straight fibre with a refractive index n.*

Figure 9.2 shows the end face of the fibre and the path of a ray inside it. For the light ray not to slip out, the angle θ must not exceed the critical angle for total internal reflection. Considering the limiting case, let the angle θ equal the critical one, i.e. assume (see Chapter One) that

$$\sin\theta = \frac{1}{n}. \tag{9.1}$$

The law of refraction at point A gives $\sin\alpha/\sin\beta = n$, or, since $\beta = 90° - \theta$,

$$\frac{\sin\alpha}{\cos\theta} = n. \tag{9.2}$$

Using (9.1), we rewrite (9.2) as

$$\frac{\sin\alpha}{\sqrt{1-(1/n^2)}} = n.$$

It follows from here that

$$\alpha = \arcsin\sqrt{n^2-1}. \tag{9.3}$$

For example, if $n = 1.3$, then $\alpha = 56°$.

Evidently, straight fibres are of no great interest. *Curved* fibres are much more interesting. Naturally, we

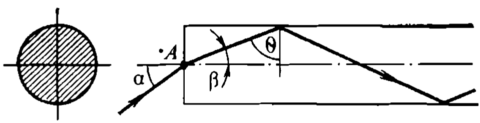


Fig. 9.2

wish to know to what extent the light guide is tolerant of curves. So consider the following problem. *We know* *the diameter D and the refractive index n of a fibre that is curved in the shape of a segment of a circle. A ray of light is incident on the end face of the fibre at an angle α. Find*

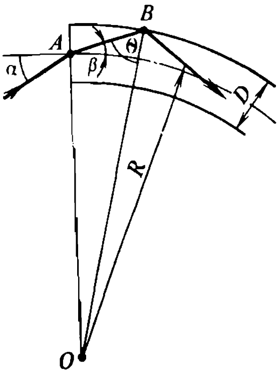


Fig. 9.3.

the least tolerant (from the point of view of total internal reflection) radius R of the curvature of the fibre.

We apply the law of sines to the triangle OAB (Fig. 9.3):

$$\frac{AO}{\sin\theta}=\frac{OB}{\sin(90^\circ+\beta)}. \tag{9.4}$$

Since $AO = R$ and $OB = R + D/2$, equation (9.4) becomes

$$\frac{R}{\sin\theta}=\frac{R+D/2}{\cos\beta}. \tag{9.5}$$

Recall now that $\cos\beta = \sqrt{1-\sin^2\beta} = \sqrt{1-\sin^2\alpha/n^2}$ and that θ must equal the critical angle for total internal reflection: $\sin\theta = 1/n$. Therefore, it follows from (9.5) that

$$Rn=\frac{R+D/2}{\sqrt{1-\sin^2\alpha/n^2}}.$$

Thus,

$$R=\frac{D}{2(\sqrt{n^2-\sin^2\alpha}-1)}. \tag{9.6}$$

Assume that $\alpha = 0$ (the direction of the ray of light coincides with the axis of the fibre). In this case (9.6) takes the form

$$R=\frac{D}{2(n-1)}. \tag{9.7}$$

If $n = 1.5$, then $R = D$.

So, the least tolerant radius of the curvature of the fibre may be as small as the diameter of the fibre itself! Certainly, curves and bends as acute as this are not practically important, because, among other things, the flexibility and strength of the fibre must also be taken into account. It appears, therefore, that an optical fibre can withstand quite acute bends as a light guide, which is certainly very important.

Rays in Conic Fibres. Up to now, we have dealt with the fibre whose diameter was constant all along its length. What will happen if, instead of a cylindrical fibre, we take a conic one, i.e. one with a gradually decreasing diameter? Can this kind of fibre be used to increase the density of light energy, i.e. to concentrate the light?

Fibres of this kind do exist. They can really be used to collect light; however, the concentration of light which can be achieved is limited. The trouble is that as the light travels along the fibre the angle θ at which a ray is incident on the lateral surface of the fibre increases with every reflection, until the total internal reflection condition is violated, thereupon the ray will leave the fibre (Fig. 9.4*a*).

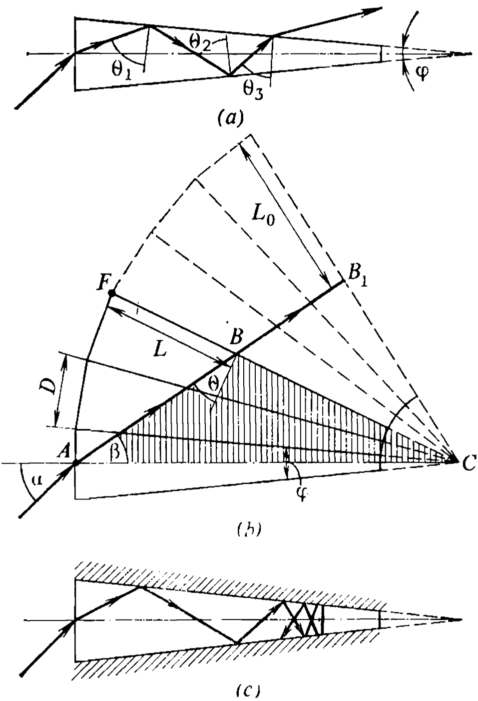


Fig. 9.4.

Consider the following problem. *Light falls at an angle α on the end face of a tapering fibre with a cone angle φ; the refractive index of the fibre is n, and the diameter of the face is D. Find the length L of the segment of the fibre which will contain the ray.*

At first sight the problem appears quite complicated as the angle of incidence of the ray changes with every reflection. However, there is a clever trick which makes the task a lot simpler. It is based on the fact that the angle of incidence equals the angle of reflection, and becomes quite clear if you look at Fig. 9.4*b*. The real path of the ray is a broken line which can be represented as a straight one (the line AB in the figure) that traverses a 'fan' of conic fibres. The distance $L = FB$ can be determined because the perpendicular at B to FC must be at an angle θ to the line AB, and θ is the angle of total internal reflection ($\sin\theta = 1/n$). Since φ is small, we can assume that $AC = FC = D/\varphi$. The law of the sines for the triangle ABC gives

$$\frac{AC}{\sin(90°+\theta)} = \frac{BC}{\sin\beta}$$

or, in a different way,

$$\frac{D/\varphi}{\cos\theta} = \frac{D/\varphi - L}{\sin\alpha/n}. \tag{9.8}$$

Hence, since $\sin\theta = 1/n$,

$$\frac{D/\varphi}{\sqrt{n^2-1}} = \frac{D/\varphi - L}{\sin\alpha},$$

and finally

$$L = \frac{D}{\varphi}\,\frac{\sqrt{n^2-1}-\sin\alpha}{\sqrt{n^2-1}}. \tag{9.9}$$

This shows that light can't be locked inside tapering conic fibre. It can't be done even if the lateral surface of the fibre is coated with a mirror. Figure 9.4*c* shows the path of a ray in such a fibre. The ray penetrates the fibre to the depth L_0 and then returns. The distance L_0 can easily be found with the help of Fig. 9.4*b*. Since AB_1C is a right-angled triangle, $B_1C = (D/\varphi)\sin\beta$ and therefore

$$L_0 = \frac{D}{\varphi}(1-\sin\beta) = \frac{D}{\varphi}\left(1 - \frac{\sin\alpha}{n}\right). \tag{9.10}$$

Note that (9.10) can be deduced from (9.9) if we assume that $n \gg 1$.

Suppose α_1 is the angle of incidence of a ray on the first end face of a conic light guide, and α_2 is the angle at which the ray emerges from the light guide at the second end face. It is clear enough that in a tapering light guide $\alpha_2 > \alpha_1$, while in a diverging one, on the contrary, $\alpha_2 < < \alpha_1$.

The Influence of Fibre Bending. In the problems we have just considered the path of a ray of light

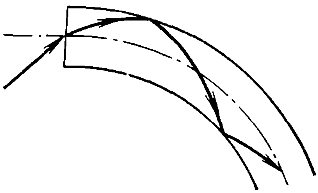


Fig. 9.5.

inside the fibre was assumed to be made up of straight lines. This assumption should only be taken as a first approximation. In reality it often turns out that this is inappropriate.

First of all, we have to take into account that when a fibre is bent its inner surface is compressed, while its outside surface is stretched. As a result, the refractive index of the inner part becomes greater than that of the outside. This leads, in turn, to the *curvature* of the ray in the fibre. The path of the ray becomes convex towards the lower refractive index. In other words, the curvature of the ray follows the curvature of the fibre. Thus, the path of a light ray in a bent fibre consists of curved segments, and not of straight lines as shown in Fig. 9.5.

Gradient Optical Fibres. In order to contain light inside a fibre more effectively, the fibre is often made so that its refractive index is highest along the axis of the fibre and gradually decreases towards the outside. Figure 9.6 shows

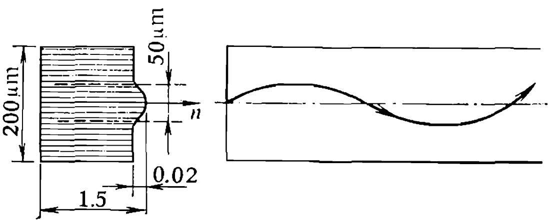


Fig. 9.6.

the change of the refractive index in the cross section of the fibre, and next to it is the path of a light ray in such a fibre. Besides, the figure gives the dimensions of this kind of fibre. If the diameter of the fibre is 200 μm, its central (light-guiding) part has a diameter of about 50 μm; the relative decrease of the refractive index in the direction from the axis towards the outside is a little over 1 % in this case. Recently, a sophisticated technology for producing optical fibres with a varying refractive index has been developed. They are called *gradient fibres*.

Thin Fibres. All of the above-described holds true for fibres whose diameter is much greater than length of a light wave, i.e. for the so-called *thick fibres*. There are *thin fibres*, too, their diameter being comparable to the length of a light wave or is even smaller. The diameter of a thin fibre is sometimes 0.1-1 μm. Obviously, we can hardly speak about the path of a ray inside such a fibre; only *wave* notions are appropriate. The field of the light wave travelling along the thin fibre fills the whole volume of the fibre and, furthermore, occupies some space around the fibre.

By now the reader hopefully has some idea of the complexity of the problems which have to be solved when designing, manufacturing,

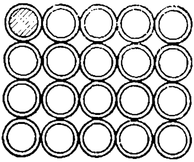


Fig. 9.7.

and applying fibre light guides. Nowadays this kind of guide is used to transmit signals over distances from several metres up to several kilometres. It goes without saying that long fibres must have very weak light absorptions, as well as quite insignificant losses of light through the walls. Modern uses for long light guides require that the losses of light energy must not exceed several decibels per kilometre of the light guide. Naturally, the requirements for light guides that will be used to transmit signals over short distances (up to a few dozens of metres) are not so rigid.

Transmission of Optical Images over a Fibre Braid. Optical fibres are widely used to transmit not only light signals but also *optical images* (two-dimensional pictures). The fibres are *braided* for this purpose. Figure 9.7 shows a cross section of a braid consisting of a relatively small number of fibres. The number of fibres in practically applied braids can sometimes reach a million. Each fibre in the braid has a special casting to prevent from seepage of light energy from one fibre to another.

In principle the transmission of images over a fibre braid is simple enough. The rays of light

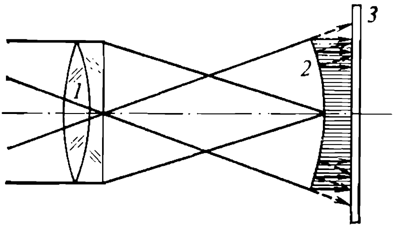


Fig. 9.8.

reflected (or emitted) by one element of the image to be transmitted are caught by one fibre in the braid which goes all along the braid, and at the other end they reproduce this element of the image. In other words, we "take possession" of each ray going from the object or from the original image, and the ray is sent along the fibre to where it is required. By keeping the relative positions of fibres in the braid the same both at the inlet and at the outlet, we can reproduce the image at the outlet which was fed in at the inlet, i.e. the image will be transmitted over a distance.

One special example, the method makes it possible to receive images of objects situated in cavities that are not easily accessible, or in places which cannot be reached using conventional optical equipment. Thus, fibre optics has opened up new ways of looking inside a human body.

Fibre Equalizer of the Light Field. Fibre optics makes much more than just the transmis-

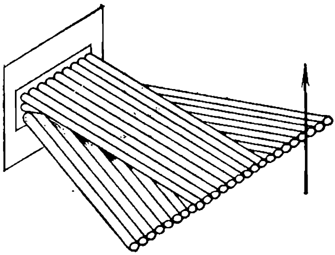


Fig. 9.9.

sion of images possible, because each ray travelling from the object is, as it were, in our hands. We can therefore increase the luminance of any ray at will, and thereby increase the brightness of the image or correct the paths of certain rays to diminish aberrations in the image obtained. Figure 9.8 shows a *fibre equalizer of the light field* which can eliminate the aberrations caused by a lens system. The figure contains: *1*—a lens system, *2*—a fibre equalizer, and *3*—a photographic plate. The dotted lines show the paths of the light rays in the absence of the equalizer, while the solid lines represent their actual paths. The correction of the paths is to eliminate the aberrations of the lens system.

Fibre Image Dissector in High-Speed Photography. Fibre optics is also used in high-speed photography. Supposing one has to photograph a fast process and so the film in the camera has to be moved very fast. This is not easy to accomplish technically. Besides, one has to take into account the time necessary for the exposure of the film. Instead a fibre device called an *image dissector* can be used. It can transform a two-dimensional picture into a line (Fig. 9.9). The two-

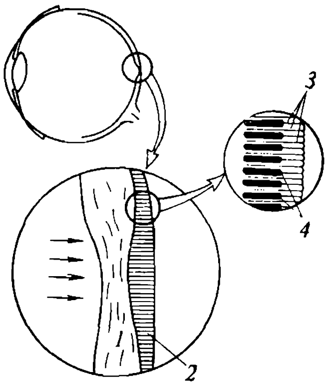


Fig. 9.10.

dimensional picture is, so to speak, cut into strips and the strips connected into one. The arrow in the figure shows the direction of the movement of the film in relation to the image. Obviously, far more exposures can be made on the same length of film if the 'pictures' are linear instead of two-dimensional. After that the fibre image dissector can turn the linear image back into a two-dimensional one, which is then rephotographed onto another film. The first film must move more slowly, and if the speed of the second film is normal, the original process will be shown in slow motion.

The Retina as a Fibre-Optics Device. To conclude the chapter, let's see an example of a *natural* fibre-optics system—the retina of the human eye. Figure 9.10 shows a diagram of the retina around the macula. Light falls on the retina and passes first through layer *1* consisting of nerve cells and fibres, and then reaches the light-sensitive elements in layer *2*. This layer is, in fact, a kind of a fibre-optics device. It contains two types of fibre (both are shown in the figure): the thinner fibres (*3*) and thicker ones (*4*) called rods and cones respectively. In the last few years scientists have come to the conclusion that there are far more natural fibre-optics components than was originally supposed.

Afterword

Hopefully, we have shown that the refraction of light covers a wide range of questions. Nevertheless, the subject has not been exhausted. Many interesting problems remained beyond the scope of this book. We will outline at least some of them.

Control of the Refractive Qualities of a Substance. Can the refractive qualities of a substance be *controlled*? In other words, can the refractive index be controlled? Modern science and technology give a positive answer to the question. There are now several techniques for controlling the refraction of light by using external actions of different kinds to alter the refractive index.

Thus, the refractive index of a semiconductor depends on the number of conduction electrons per unit volume. In theory the refractive index decreases when the number of conduction electrons increases. In turn, the number of conduction electrons depends on a number of factors. For example, it increases when the semiconductor is irradiated by the light of a frequency specific to the semiconductor. Imagine a thin semiconductor plate. By exposing certain parts of it to light thus decreasing the refractive index, we can obtain a kind of flat lens or prism for light travelling along the plate (it must be transparent enough). Figure A.1 illustrates the idea. The shaded areas represent the irradiated part of the plate, while the arrows show the direction of the rays of light moving along the plate. The figure shows four cases: (a) the deflection of a

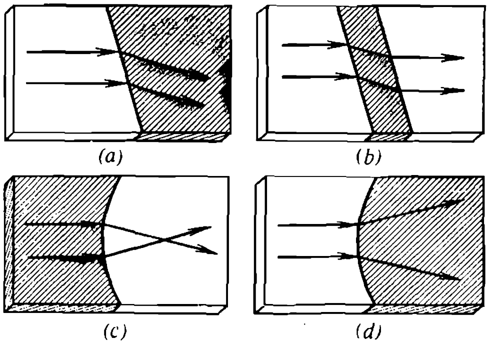


Fig. A.1.

beam of light, (b) parallel shift of the beam, (c) focusing of the beam, and (d) defocusing of the beam.

In addition, the refraction of light can be controlled by placing the refracting substance in an external electric field of varying intensity. For instance, if a liquid is placed into an electric field it acquires the properties of a uniaxial (positive or negative) crystal, with the direction of the external field being its optic axis. Two plane-polarized waves of light—the ordinary and the extraordinary ones—travel in such an anisotropic liquid in the same way they do in a uniaxial crystal.

The difference $v_0 - v_e$ is proportional to the square of the intensity of the field: $| v_0 - v_e | \sim \sim E^2$. This is called the *Kerr electrooptic effect.*

We pointed out (see Chapter Nine) that the refractive index of a curved fibre varies from point to point as a result of nonuniform stress. This holds true for any substance in which the stress arises from natural causes or is generated on purpose, for example, by sending an ultrasonic wave through the substance. Thus, the refractive index of a substance can be controlled with the help of ultrasonic waves. This idea has given rise to a new branch of modern optics which is called *acousto-optics.*

The index of refraction depends on the temperature of the substance, too. In some substances it increases as the temperature rises, while in others it decreases. A change in temperature, by accident or on purpose, at the interface between two regions in a substance can create a *thermal lens.* The focal power of such a lens depends on the thermal condition, i.e. on the degree of heating or the nature of the heat abstraction. A thermal lens appears, for example, in the active element of a laser excited by intense optical radiation which heats the active element.

If a very intense beam of light, for example, the beam of a powerful laser, is propagating through a substance, a peculiar phenomenon can occur which can be regarded as a *self-induced* effect of the beam; the beam changes the refractive index of the substance thus affecting the propagation of the light in the substance. The refractive index is usually higher along the axis of the beam (where the intensity of the light field is higher) and decreases towards the beam's boundary. The reader will know about the curvature of light in optically nonhomogeneous media (see Chapter Two), so he will find it easy to understand why the beam will focus itself. Upon entering the substance it will not be dispersed but collected

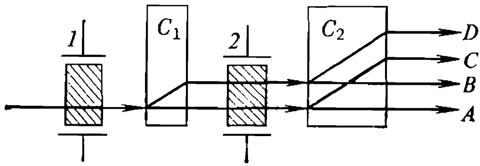


Fig. A.2.

into a pencil of light. The *self-focusing* of powerful lasers in different media is the subject of careful research nowadays.

Electrooptic Deflector. Thus, there are various means of influencing the refractive qualities of a substance and the controlled refraction of light is widely used in modern optical instruments. By way of an example, consider an electrooptic device which can shift a beam of light without changing its direction. Note that the operation is carried out with extraordinary speed and takes less than 10^{-6} s. The device is called an *electrooptic deflector* and Fig. A.2 shows the diagram of its simplified variant. Here C_1 and C_2 are two birefringent crystals oriented in the same way, and *1* and *2* are Kerr cells, i.e. cells filled with liquid and placed in an electric field, the direction of the field being perpendicular to that of the original beam of light. In the figure the beams of light are represented by arrows.

Assume that the original beam is plane-polarized and that when a voltage is applied to a Kerr cell, it acts as a half-wave plate (see Chapter Eight) turning the plane of the polarization of the beam by 90°. To obtain this angle, the field in both cells must be oriented at 45° to the direction of the polarization of the original beam of light (note that the direction of the field in both cells is the same). Let the crystals C_1 and C_2 be oriented in such a way that the original beam is the ordinary beam of light for them (it is polarized perpendicular to the plane of the principal section of the crystals).

Suppose that a voltage has not been applied to either of the cells (they are both switched off). In this case the beam of light passes through the crystals without being deflected, i.e. it acts as the ordinary beam, and the beam emerges from the deflector in position *A* (see the figure). When both cells are switched on, the plane of the polarization of the beam is turned 90° after passing through cell *1*, and becomes the extraordinary beam for the crystal C_1 and is therefore deflected by it. After the plane of polarization is turned by 90° in cell *2*, the beam reaches the C_2 crystal as ordinary and for this reason is not deflected by it. As a result, the beam emerges from the deflector in position *B*. It is evident that when cell *1* is switched off and cell *2* is on, the beam will emerge from the deflector in position *C*, and when cell *1* is on and cell *2* is off in position *D*.

For the sake of simplicity, we have considered a two-stage diagram which has only four positions at the output. If the number of stages is n, the number of positions at the output will be 2^n.

Modern deflectors have ten stages and can 'hit' 1024 positions without difficulty.

Cosmic Lenses. Finally, let us touch upon another question connected with the refraction of light—so-called *cosmic* (or gravitational) *lenses.* When using this term astronomers usually mean one of two optical phenomena. The first is connected with the refraction of light in the atmosphere and was described in Chapter Two. The atmosphere of the Earth or of any other planet can act as a sort of lens, thus, radiation propagating from distant cosmic sources to the Earth can be focused by the Earth's atmosphere, which is then working as a gigantic lens with a diameter of over 10,000 km. The second phenomenon is connected with the curvature of star's rays in a gravitational field. We shall discuss it in greater detail.

It follows from the theory of relativity of the great 20th-century physicist Albert Einstein (1879-1955) that light rays passing close to a massive body, the Sun, for example, are made to curve. This curvature allows us to observe stars during a total eclipse of the Sun, which, according to precise calculations, should then be behind the solar disc (Fig. A.3). We have here therefore a special kind of cosmic lens which can be termed a *gravitational lens.*

A spectacular example of a gravitational lens was discovered in the middle of 1979, when two very similar quasars situated close to each other were discovered. The first *quasars* were discovered about twenty years ago and the name is applied to very distant cosmic objects which are sources of powerful radiation. Quasars are supposed to be the nuclei of distant galaxies where mighty processes take place which increase the luminosity of the galaxies to make them thou-

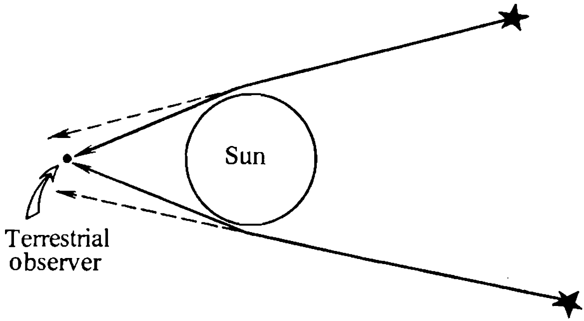


Fig. A.3.

sands of times brighter than ordinary (quiet) galaxies. The angular distance between the two quasars discovered in 1979 was 6 seconds of arc,

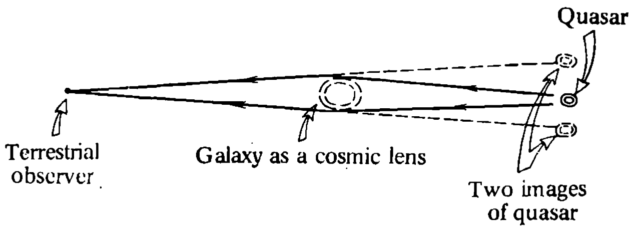


Fig. A.4.

which corresponds to 0.003 of the visible diameter of the full moon. This was quite surprising as all the quasars known at that time were more or less regularly scattered about the sky and a few degrees apart on average. Still more surprising was the fact that the spectra of the two quasars coincided completely. Such a coincidence had never occurred before. It has now been proved that the two quasars are the images of *the same quasar* caused by a powerful gravitational lens existing in space (Fig. A.4). Further research showed that the gravitational lens is formed by a galaxy situated between the Earth and the mysterious quasar.

* * *

This is the end of our journey around the world of refracted light, or, in other words, around the world of *geometrical optics*. The presentation of radiation as light rays has allowed us to investigate a lot of interesting phenomena. It should be remembered, however, that this is only a limiting case of the more general approach to light, which uses the concept of a light wave. We did mention the idea, albeit infrequently, but confined ourselves to the phenomena which could be analysed without consistently using it. Had we done otherwise, our journey would have been too long and we would have left the world of geometrical optics to travel in the more extensive world of wave optics.

Advice to a Reader Who Has Reached the End. Dear reader, You have probably noticed that this book was sometimes easy and sometimes difficult to read. It was easy when the historical backgrounds, the experiments and the physical aspects of the phenomena were described, but it became more complicated when concrete problems were considered (they are given in brevier), or when the path of the light through a device was traced. Perhaps you omitted the difficult places in the book or, at least, scanned through them. This is nothing to be ashamed of, but now you should take a pen and a notebook and go through the book again carefully analysing all the problems and the diagrams.

Index

Aberration,
chromatic, 141
spherical, 141
Absorption of light, 100
Accommodation, 174
Acousto-optics, 230
ALHAZEN (IBN AL-HAYTHAM), 45
Angle,
Brewster, 204
critical, 34
of incidence, 17
of refraction, 42
of view, 169
Angular dispersion, 104
Angular magnification, 151 154
ARISTOTLE, 84
Atmosphere, 42
boundary of, 48

Barometric height formula, 52
BARTHOLIN, ERASMUS 186
Beam,
reflected, 33
refracted, 33
Biprism, 81
Birefringence, 186
BOLTZMANN, LUDWIG, 53
Boltzmann constant, 53
BREWSTER, SIR DAVID, 203

Camera obscura, 134
CAVALIERI, FRANCESCO BONAVENTURA, 157
Cavalieri's formula, 157
CHERENKOV, PAVEL ALEKSANDROVICH 38
Cherenkov,
effect, 38
radiation, 38
CLEOMEDES, 12
COPERNICUS, NICOLAS, 155
Corpuscular theory, 203

DEMOCRITOS, 170
DESCARTES, RENE, 17
Descartes error, 17
Dichroic plate, 209
Dichroism, 209
Diffraction of light, 84
Dispersion, 83
anomalous, 98

DOLLOND, PETER, 158
DE DOMINIS, MARCUS ANTONIUS, 117

EINSTEIN, ALBERT, 233
Electrooptic deflector, 231
EMPEDOCLES OF AGRIGENTUM, 170
EULER, LEONHARD, 96
Eye-piece, 151

Fibre(s),
  gradient, 222
  optics, 235
  thick, 222
  thin, 222
Field of vision, 177
FLANSTEED, JOHN, 50
FLEISCHER, JOHANNES, 117
FRANK, 41
FRESNEL, JEAN, 33
Fresnel's formula, 33
Fresnel zone plate, 163m

HALLEY, EDMUND, 157
Halo, 130
HARRIOT, THOMAS, 84
HERO OF ALEXANDRIA, 25
HUYGENS, CHRISTIAAN, 19

IBN AL-HAYTHAM, (ALHAZEN), 45
Illusion(s), 43
Image(s),
  inverted, 134
  real, 142
  virtual, 142
Image dissector, 225
Incidence,
  angle of, 17
  plane of, 27
Incident light, 39
Index,
  refractive, 23
Interface, 11

GALEN, CLAUDIUS, 170
Galilean telescope, 151
GALILEO, (GALILEI), 96

KEPLER, JOHANNES, 45
KERR, JOHN, 230
Kerr electrooptic effect, 230
KIKOIN, ISAAK KONSTANTINOVICH, 159
LRYLOV, ALEKSEI, 51

KEEUWENHOEK, ANTONY VAN, 157
Lens,
  collecting, 93
  concave-convex, 45
  convexo-concave, 145
  cosmic, 233
  Dollond's achromatic, 158
  double-concave, 145
  double-convex, 115
  gravitational, 233
  objective, 151
  power, 139
  thin, 137

LEIBNIZ, 28
LEONARDO DA VINCI, 172

MALUS, ETIENNE STEPHEN, 203
MANDELSHTAM, LEONID, 95
MARCI OF KRONLAND (JOHANNES MARKUS), 84
Mirage(s), 60
Mock suns, 133
Monochromator(s), 105

NEWTON, ISAAK, 50
NICOL, WILLIAM, 209
Nicol prism, 209

Ommatidium, 185
Optical centre, 169
Optical fibre, 214
Optical image, 134
Optics,
  geometrical, 235

Plane,
  of incidence, 27
  of polarization, 205
Plane-polarized light, 205
Polarized beam, 209
Prism,
  dispersive, 104
  Dove, 102
  Glan-Foucault, 210
  Nicol, 209
  polarization, 209
PYTHAGORAS, 170

Quasar(s), 233

Rainbow, 113
Ray(s),
  emergent, 122
  extraordinary, 186
  incident, 122
  monochromatic, 67
  ordinary, 186
  refracted, 67
Refractometer, 71
ROZHDESTVENSKY, DMITRY, 100

Scattering of light, 112
SNELL, WILLEBROD van ROIJEN, 17
Spectrograph(s), 105
Spectrometer(s), 105
Spectroscope(s), 105
Stellar scintillation, 59

VAVILOV, SERGEI, 96
Velocity,
  light wave, 192
  ray, 192

Wave-front(s),
  secondary, 21
  spherical, 20

YOUNG, THOMAS, 174

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book.

We would also be pleased to receive any other suggestions you may wish to make.

Our address is:

Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

Printed in the Union of Soviet Socialist Republics

SCIENCE FOR EVERYONE

The rainbow and the Galilean telescope, the spectre of the Brocken and illuminated fountains - what can all these have in common? The answer is that in all of them the refraction of light has an important role to play. That is why you will read about them in this book written by the well-known popular science authors Lev and Aldina Tarasov. Professor Tarasov is a prominent figure in the field of quantum electronics and optics; Aldina Tarasova is a teacher of physics and studies methods for teaching the basics of science.

The book treats refraction of light from two points of view. Firstly it describes the many, and sometimes quite surprising, forms that the refraction of light can assume. Secondly, the hook traces the long and difficult path mankind covered before it came to understand some of the mysteries of Nature. Recent inventions such as the laser and the transmission of images over optic fibres are also dealt with. The book was first published in Russian in 1982, and is now coming out in English. We are certain it will be well received by those who are interested in the history of science and the present development of theoretical optics and its related technology as well as by all lovers of Nature.

ИЗДАТЕЛЬСТВО · МИР ·