tewart, R. M.
Cal Tech

OPTICS

Robert A. Millikan

Duane Roller

Earnest C. Watson

Carl D. Anderson

Wolfgang K. Panofsky

California Institute of Technology

1943

CHAPTER ONE

BASIC FACTS AND CONCEPTS IN GEOMETRICAL OPTICS

The manifestation of light is one of the most obvious of physical phenomena, and optical effects such as the rainbow, halos, the mirage and even the sunny sky -- which is "as a molten looking glass"[1] -- must have provided early man with some of his most vivid experiences. Familiarity with many simple optical phenomena is implied in all the astronomical lore of the ancient world, and mirrors of polished metal and "burning glasses" are among the oldest of devices. Thus it is natural that optics, which treats of the properties and nature of light and vision, should be one of the oldest branches of physics. Growing out of practical lore, optics progressed more rapidly as the invention of instruments such as the microscope and telescope increased the demands for optical knowledge; and it developed into an accurate and significant science when controlled, quantitative investigation came to be a basic and integral part of its methods.

1. Rectilinear Propagation in a Homogeneous Medium. In the early ages when windows were without glass, and dust abounded, a straight shaft of brilliant sunlight piercing the dusty atmosphere of a habitation was one of the most common of sights. Thus doubtless rose the notion that light travels in straight lines so long as it remains in the same medium. It was tacitly assumed in the astronomy of the ancients, and acquired the

---

[1] Job, Chap. 37, verse 8.

status of a principle in the Optics of Euclid.[1] (c. 300 B.C.) This principle of the rectilinear propagation of light in a homogeneous medium has innumerable modern applications, for example, in all measurements of angles made with astronomical and surveying instruments, and in the explanation of the phenomenon of parallax, or apparent displacement of an object due to an actual displacement of the observer (Fig. 1). Another simple illustration is furnished by the pinhole camera,[2] in which a small hole takes the place of the usual lens. As indicated in Fig. 2, a narrow cone of light from

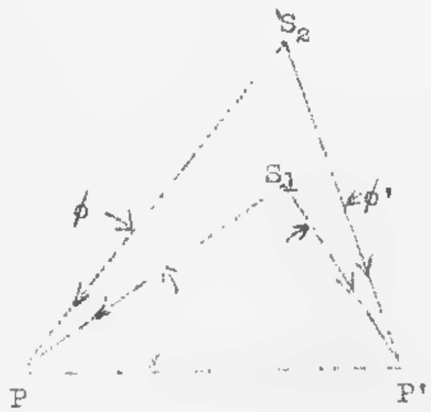


Fig. 1. If the observer moves from P to P', the nearer object S_1 will appear to be displaced an angular distance of $\phi + \phi'$ to the left with reference to the farther object S_2.

---

1 Euclidis Optica (Teubner, Leipzig, 1895); this is Vol. 7 of the authoritative edition of Euclidis opera omnia, ed. by G. J. L. Heiberg and H. Menge (8 vols., 1883-1916). It is difficult to attribute a book containing so many inaccuracies to one whose geometry is characterized by accurate reasoning and lucidity; however, the logical structure of all the ancient works on physics is very loose, and not to be compared with that of the mathematical writings.

A brief summary of the Optics will be found in *T. L. Heath, A Manual of Greek Mathematics (Oxford, 1931), pp. 267-268, and also in an article on "Optics" in the Encyclopaedia Metropolitana (London, 1845), Vol. III, p. 394. It, together with the Catoptrica (theory of mirrors), which is also usually attributed to Euclid but is really a compilation made much later from ancient works on the subject, gives a good idea of the views regarding light that prevailed among the Greeks.

2 Although the photographic plate was not invented until the nineteenth century, the principle here discussed -- that of the pinhole camera or "camera obscura" -- was discovered early in the sixteenth century and is described in detail by G. B. Porta (1536-1615) in his Magia naturalis (1558).

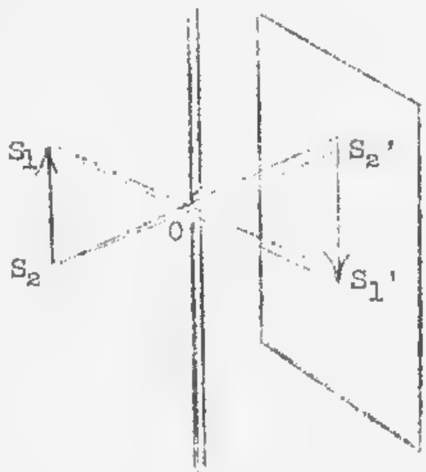


Fig. 2. Formation of an image by a pinhole camera.

a point S_1 of the object S_1S_2 passes through the pinhole O and illuminates a small spot S_1' on the photographic plate. Cones of light from other points of the object illuminate other corresponding spots on the plate. The result is an inverted image $S_1'S_2'$. This image is said to be more clearly defined the more nearly the points of the image and of the object approach a one-to-one correspondence. The definition of the image accordingly will be lessened when the pinhole is made larger, or when the object or plate is brought closer to the pinhole, for then points on the object will register on the plate as overlapping spots. But the definition will also be lessened if the pinhole is made too small, for then spots will again overlap, but now for an entirely different reason; namely, because the cones of light after passing through the pinhole spread out laterally through angles, which, while small, become appreciable as the size of the pinhole is decreased. In other words, light exhibits the same phenonomen of diffraction as do sound and water waves,[1] although the effect in the case of light is relatively small (Chap. 4). Lateral spreading of a beam of light always occurs, even in a homogeneous medium, but it is small enough to be ignored in treating certain important classes of problems.

---

[1] See, for example, Millikan, Roller and Watson, Mechanics, Molecular Physics, Heat and Sound (Ginn, 1937), pp. 387, 389. This book will be designated hereafter by the abbreviation MRW.

With effects due to diffraction ignored, and the principle of rectilinear propagation thus accepted as generally valid, it is possible to develop a special division of optics in which descriptions in terms of geometrical relations are the most natural and simple. This division, called geometrical optics or, sometimes, the theory of optical instruments, enables one to trace the passage of light through optical instruments in detail, and thus to determine the principles of their construction. Despite the simplifying assumptions which form its basis, geometrical optics has proved to be of enormous practical value, and its concepts and methods have so permeated the whole science of optics that it is essential to have an understanding of them. As we shall see later, geometrical optics is able to deal with phenomena that can be treated successfully without taking into account any definite hypothesis concerning the nature of light or of its interactions with matter. It is thus to be contrasted with physical optics, the division that deals with theories of the nature of light and of its interactions with matter, and with the experimental bases and verifications of these theories. A third main division of the science -- physiological optics -- is concerned with the physiology and physics of vision.[1]

In geometrical optics, the basic concept is the ray, which we may best define as the purely fictitious axis of a narrow cone of light. It is also sometimes useful to think of a ray of light as a path of energy-

---

[1] Since we shall be concerned mainly with geometrical and physical optics in this book, the student who is interested in problems of physiological optics should consult the textbooks and treatises devoted primarily to that field. A good elementary treatment will be found in J. P. C. Southall's Introduction to Physiological Optics (Oxford University Press, 1937). By far the most important treatise in the field is Helmholtz's Physiological Optics (1856-1866); an English translation of this great work was published in 1924-1925 by the Optical Society of America.

transfer. This of course involves the modern conception of light as a form of energy, a conception that was not clearly enough formulated so as to be very useful until the middle of the nineteenth century, when energy became a clear-cut concept of fundamental importance in mechanics and heat.[1] Today we have no difficulty in conceiving of light as a form of energy, because of our familiarity with the energy concept, and with such phenomena as the increases of temperature or the chemical changes observed to occur in various bodies when they are exposed to a source of light.

2. The Principle of Superposition. An important notion implicit in early optical lore was that light from various sources traverses the common region between the sources without getting mixed up, as it were. An obvious example is the fact that two people can see each other simultaneously without distortion. This notion, stated more precisely, becomes the important, basic principle that, if light rays from two or more sources intersect, each will thereafter be the same -- that is, will be able to produce the same effects -- as if it had traversed the region alone; the rays while intersecting acquire no properties by virtue of their number that they do not already possess individually. Any phenomena for which this latter is true are said to be superposable, and hence the foregoing principle is referred to as the principle of superposition for light.[2]

---

1 MRW (1937), p. 76.

2 An important example of superposable phenomena in mechanics is evidently that of forces acting simultaneously on a particle, for each force produces its own effect independently of the action of the other forces; in other words, the principle of the independence of forces [MRW (1937), p.53] is merely a special case of the general principle of superposition. Another special case arising in mechanics is the Fourier theorem [MRW (1937), p.332]. In general, any effect is superposable if it can be described by means of a linear differential equation; that is, a differential equation in which neither the dependent variable nor its derivatives enter in any power higher than the first power. The principle of superposition is a clearly defined

The concept of the ray, the principles of rectilinear propagation and of superposition, and the two laws of reflection and of refraction treated in the remainder of this chapter provide the entire basis of geometrical optics. When effects due to diffraction must be taken into consideration, they must be treated separately, by means of the theory of diffraction (Chap. 4).

Regular Reflection

Light arriving at a surface separating two different mediums is, in general, partly reflected into the medium into which it originally was travelling and partly transmitted into the new medium. The percentage of light reflected increases as the rays strike the surface at more nearly grazing incidence; for example, when light is incident perpendicularly on a smooth surface of water, only about 2 percent of it is reflected, whereas, for nearly grazing incidence, about 72 percent is reflected. Another factor which may be easily observed to affect the percentage reflected is the character of the two mediums at whose interface the reflection occurs; for example, a piece of glass immersed in water reflects much less light than it would in air under the same circumstances.

Light reflected from a smooth surface does not appear to the eye to come from the surface but from an image located behind or in front of the surface; the rays are reflected in definite directions and hence are said

---

property of such an equation: for we know that if y is a function of x that satisfies any given linear differential equation, and if z is another function of x that satisfies the same equation, then $y + z$ is a function of x that satisfies it; or, more generally, the sum of any number of individual solutions of a given linear differential is also a solution. Superposable phenomena are relatively easy to investigate and thus have usually been the first to be studied in physical science. In dealing with phenomena that cannot be treated as superposable, a nonlinear differential equation, which involves so-called combination terms, or some other mode of treatment must be employed.

to be regularly reflected. On the other hand, rays reflected from a very rough surface pass out from it in all directions, as if the surface itself were the original source of the light, and hence objects are not seen reflected in it; the light from a rough surface is diffusely reflected. Even the smoothest obtainable mirror reflects some of the light diffusely, because of slight irregularities due to the molecular structure of the surface. Contrariwise, most mat surfaces, such as rough paper, reflect an appreciable part of the light regularly, the percentage reflected increasing as the rays strike the surface at more nearly grazing incidence.

3. The Law for Regular Reflection. Experiment shows that the law for regular reflection is the same as that for sound[1]; namely, (a) the incident and reflected rays make equal angles θ and θ' with the normal drawn to the surface at the point of incidence, and (b) the two rays and the normal lie in one plane. This specification of the plane in which the reflected ray lies appears to have been first emphasized by Alhazen (c. 965-1039), in his Treasury of Optics.[2] But that the angle of incidence θ and the angle of reflection θ' are equal certainly was known to the Greeks,[3] and Heron of Alexandria[4] even deduced this equality from a more general

---

1 MRW (1937), pp. 386, 390.

2 This treatise was translated from Arabian into Latin in 1270 and printed at Bâle in 1572 under the title, Opticae thesaurus Alhazeni libri VII, cum ejusdem libro de crepusculis et nubium ascensionibus. It remained a standard authority on optics down to the seventeenth century. Alhazen was the greatest Muslim physicist and one of the greatest students of optics of all time.

3 See Euclid's Optics, Prop. 19, and the Catoptrica.

4 Heron lived sometime in the period between 150 B.C. and 250 A.D. The Capotrica, a treatise on reflection ascribed to him, appears in Latin and German translations in the authoritative edition of his works, Heronis Alexandrini opera quae supersunt omnia, ed. by W. Schmidt (Leipzig, 1901), Vol. II. A brief summary of its contents will be found in *T. L. Heath, A Manual of Greek Mathematics (Oxford, 1931), p. 433.

assumption; namely, that light in traveling from one point to another follows the shortest path. Although the rectilinear propagation of light obviously is also deducible from it, we shall see in Sec. 4 that the assumption is by no means generally valid. It holds only if the medium is homogeneous; moreover, in certain cases of reflection from a concave mirror, the path is a maximum instead of a minimum. The historical significance of Heron's principle of the shortest path is that it represents an early attempt to describe a physical situation in terms of some minimum value. Minimal principles of various forms today provide methods of great power and elegance for attacking a variety of involved problems.

Refraction

The phenomenon of refraction, or change in direction of a beam of light when it is transmitted from one medium into another, was familiar to the Alexandrian astronomers,[1] who realized that a correction for atmospheric refraction enters into the important practical problem of computing times of rising and setting of heavenly bodies from observations of their earlier positions in the sky. In an attempt to determine how much change in direction occurs in refraction, Ptolemy[2] (c. 70-147 A.D.) made experiments on the passage of light from air into water and other substances, and compiled tables showing corresponding observed values for the angle of incidence θ_1

---

[1] The Catoptrica attributed to Euclid notes that a coin in a cup can be lifted into sight by pouring in water.

[2] Claudius Ptolemaeus was an Alexandrian astronomer, mathematician and geographer of extremely great ability, his influence during the first sixteen centuries, A.D., being second only to that of Aristotle's. Ptolemy's work on refraction is described in the fifth, and last, book of his Optics, the text of which is known only through a twelfth century Latin translation from the Arabic; the modern reprint of this translation is L'ottica di Claudio Tolomeo (G. Govi, Torino, 1885).

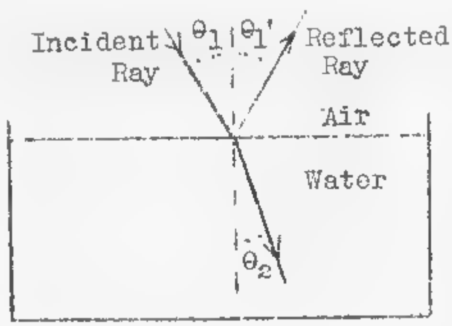


Fig. 3. Ray of light passing from air to water.

and the angle of refraction θ_2 (Fig. 3). These data from one of the few recorded experimental investigations of antiquity enabled Ptolemy to make empirical corrections for effects involving refraction. He also concluded that the ratio θ_1/θ_2 is always the same value for a particular pair of mediums. Ptolemy's own data (Table I) fail to substantiate this rule; yet it was not until some nine hundred years later that Alhazen, performing similar experiments on refraction, recognized that the rule holds only for small angles. Although the general relation between the angles escaped him, Alhazen succeeded in formulating one part of the law of refraction as we now know it; namely, that the refracted and incident rays lie in the plane containing the normal to the refracting surface.[2]

Table I. Ptolemy's values of angles of incidence and refraction for white light passing from air to water.[1]

θ_1	θ_2	θ_1/θ_2
0°	0°	-
10°	8°	1.3
20°	15 1/2°	1.3
30°	22 1/2°	1.3
40°	29°	1.4
50°	35°	1.4
60°	40 1/2°	1.5
70°	45 1/2°	1.5
80°	50°	1.6

[1] L'ottica di Claudio Tolomeo G. Govi, Torino, 1885 , Bk. V, p. 142.

All of Alhazen's work had great influence on European thought and was known to Kepler. When the telescope was invented, in 1609, Kepler became interested in finding a geometrical explanation for this instrument. To experimental data on

[2] Opticae Thesaurus, Bk. 7.

refraction he applied the inductive method that he had employed so successfully in arriving at the laws of planetary motion,[1] but he was unable to arrive at the general law connecting angles of incidence and refraction. Nevertheless, by using the rule that, for angles less than about 30°, θ_1/θ_2 is practically constant for a given pair of mediums, he was able to predict the course of light rays through various types of lenses and lens combinations, including the astronomical telescope, and to obtain an approximate geometrical theory of their action.[2] So notable are these advancements over the achievements of his predecessors that Kepler may be regarded as the founder of modern geometrical optics.

4. The Law of Refraction. It remained for Willebrord Snel van Royen (1580-1626), a Dutch physicist, geodisist and mathematician, to formulate the complete quantitative law of refraction.[3] In the course of experiments on refraction, Snel observed that the length of path OS_2 of a ray S_1OS_2 passing from air to water (Fig. 4) and falling on the vertical side of the containing vessel, bears a constant proportion to the path OS_2'' which the ray would have

---

1 MRW (1937), p. 41.

2 Kepler's two works on optics were Ad Vitellionem Paralipomena, quibus Astronomiae Pars Optica traditur (Frankfurt, 1604), containing important discoveries in the theory of vision and a statement of the approximate rule for diffraction, and the more important Dioptrice (Augsburg, 1611), on the theory of lenses. Both books appear in Joannis Kepler opera omnia, ed. by C. Frisch (8 vols., Frankfurt, 1858-1871), and in the more recent Johannes Kepler Gesamelte Werke, ed. by M. Caspar (Munich, 1938-). English summaries of their contents will be found in *A. Wolf, A History of Science, Technology, and Philosophy in the 16th and 17th Centuries (Allen & Unwin, 1935), pp. 245-250, and in *E. Mach, The Principles of Physical Optics, tr. by J. S. Anderson and A. F. A. Young (Dutton, 1925), pp. 13, 29-32, 43-47, 53-54.

3 His discovery is described in an unpublished manuscript written in 1621. Huygens, who saw this manuscript, credits Snel with the discovery in his Dioptica (1653) Opera posthuma (Leyden, 1703) , p. 2. A French translation, in parallel with the original Latin, will be found in the Oeuvres Complètes de Christiaan Huygens (Nijhoff, The Hague, 1888-1927), Vol. 13, pp. 6-8.

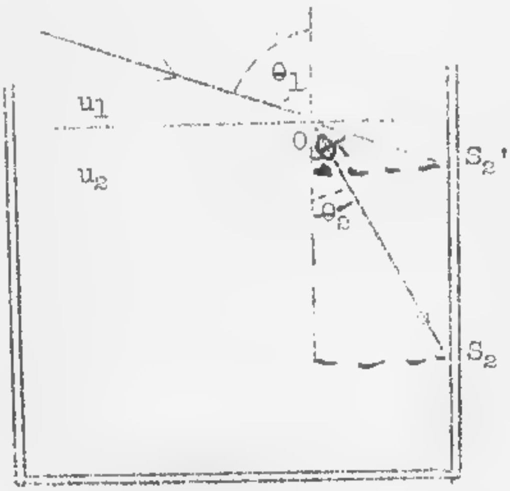


Fig. 4. Snel's law of refraction.

traversed had the water not been present. He correctly concluded that, for any particular pair of mediums, the ratio OS_2/OS_2' is constant for all angles θ_1 and θ_2. This being the case, a number μ_1 may be assigned to some one medium, taken as the standard medium, and a corresponding number μ_2 may be defined for any other medium by means of the equation

$$OS_2/OS_2' = \mu_2/\mu_1. \qquad (1)$$

The number μ is called the refractive index of the medium. Selecting a vacuum as the standard "medium", we arbitrarily make its refractive index unity. The refractive index of any substance referred to a vacuum as the standard medium is sometimes called the absolute refractive index of the substance.

As is evident from Fig. 4, $OS_2/OS_2' = \sin\theta_1/\sin\theta_2$, and hence Snel's conclusion is equivalent to the statement that the ratio of the sines of corresponding angles of incidence and refraction is constant for any two given mediums. This more useful and elegant formulation is due to Descartes.[1] We may, therefore, summarize the complete law of refraction in the following statement:

---

[1] Whether Descartes arrived at the law independently, or had seen Snel's manuscript, is not known. Nevertheless, the first published statement of the general law of refraction appears in La Dioptrique (1637), "Discours II", an essay prepared by Descartes to illustrate the system of methodology expounded in his great work, Discours de la Methode, and published as a supplement to that book. An extract from the essay appears in *A Source Book in Physics (1935) pp. 265-273. Descartes' complete works are collected in Oeuvres de Descartes, ed. by C. Adam and P. Tannery (13 vols., Paris, 1897-1911).

When a ray of light passes from a medium 1 into a medium 2, (a) the angles θ_1 and θ_2 which the rays make with the normal to the surface separating them are related by the equation

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2, \qquad (2)$$

μ_1 and μ_2 being the respective refractive indexes of the mediums, and (b) if the mediums are isotropic, the rays in the two mediums lie in the same plane with the normal, and on opposite sides of the normal.

If a substance is in the solid or liquid state, μ obviously may be determined directly, by placing a sample of the substance in a vacuum, for which μ is 1, measuring the angles θ_1 and θ_2, and applying Eq. (2). Various indirect but more accurate methods will become evident as we proceed.

The refractive index of a gas or liquid varies appreciably with the temperature. For all substances, as was first shown by Newton (Sec. 9), the refractive index also depends to a small, though important, extent on the color of the light used to determine it (Table II). This phenomenon is known as the dispersion of light.

Table II. Refractive indexes relative to air.

Substance	Refractive indexes relative to air			
	μ'_C(red)	μ'_D(yellow)	μ'_F(blue)	μ'_G(violet)
Water (20° C)	1.3312	1.3330	1.3372	1.3404
Ethyl alcohol (20° C)	1.3605	1.3618	1.3666	1.3700
Carbon disulfide (20° C)	1.6182	1.6276	1.6523	1.6748
Crown glass, No. 123*	1.51458	1.51714	1.52326	1.52859
Flint glass, light, No. 188*	1.57638	1.58038	1.59029	1.59931
Flint glass, dense. No. 76*	1.65007	1.65548	1.66911	1.68181

*National Bureau of Standards melting number.

In the case of gases, the refractive indexes listed in tables are always absolute indexes. In the case of solids and liquids, on the other hand, much of the practical work is carried out in air, and then it is usually more convenient to select air at $0^{o}C$ and 1 A_s, rather than a vacuum, as the standard medium to which the value unity for the refractive index is assigned. For this reason the value for a solid or liquid listed in tables such as Table II is usually the <u>refractive index relative to air</u>, μ'. Since the refractive index of air itself is approximately 1.0003, at $0^{o}C$ and 1 A_s, the absolute refractive index μ of any substance evidently is 1.0003 μ'. For most purposes it is sufficiently accurate to consider μ and μ' for any substance as equal.

<u>Example 1</u>. If a ray of light is passed from air through a series of glass plates, or other transparent mediums with parallel faces (Fig. 5), the ray is observed to emerge into the air parallel to its original direction, although it undergoes a lateral displacement. Show that this fact is predicted by the law of refraction.

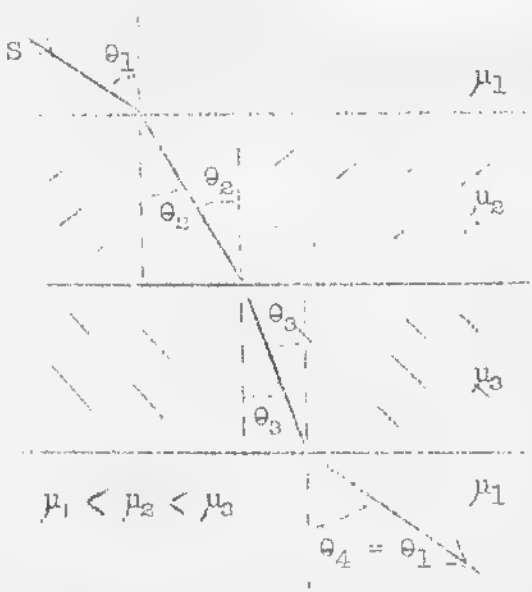


Fig. 5. When a ray of light passes through any number of mediums bounded by plane parallel refracting surfaces, the angle between the ray and the normal in any one medium is independent of all the intermediate mediums passed through.

<u>Solution</u>. Application of Eq. (2) successfully to the conditions existing at each of the refracting surfaces yields

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2$$
$$= \mu_3 \sin \theta_3 = \cdots = \mu_1 \sin \theta_n.$$

Therefore, $\theta_n = \theta_1$; that is, the total <u>deviation</u>, or total change in the angular direction of the ray, is zero. In fact, the angle between the ray and the normal in any medium is seen to be independent of all the intermediate mediums passed through.

In the case of the foregoing example, the direction in which the light is propagated clearly may be reversed without changing the <u>path</u> of the ray. That this is generally true both for reflection and refraction was first emphasized by Alhazen and is referred to as the principle of the reversibility of light rays.

5. The Principle of Extreme Time. It was mentioned in Sec. 3 that both the principle of rectilinear propagation and the law of reflection are deducible from the assumption that, in a homogeneous medium, light follows either the shortest or the longest path between two points. But any attempt to deduce the law of refraction from this assumption fails. The reason for the failure became clear when Fermat demonstrated, in 1662, that it is the time required for light to pass between two points, rather than the length of the geometrical path, that is always either a maximum or a minimum. Applied to the case of refraction illustrated in Fig. 6, for example, the Fermat principle asserts that the path light actually will take in traveling between points S_1 and S_2 is such that the time is smaller or, in some cases, larger than it would be for any other path connecting the two points. If the refracting surface is plane, as in Fig. 6, it turns out that the time is always a minimum. If the refracting surface is curved, the time is in some cases a minimum and in others a maximum. Similar remarks apply to reflection from plane and curved mirrors.

Fig. 6. The Fermat principle applied to refraction at a plane surface.

Let us see what law of refraction can be deduced from the Fermat principle. Denote by t_1 and t_2 the times required for the light to traverse the paths S_1O and OS_2, respectively. Since the refracting surface XX' in

Fig. 6 is plane, $t_1 + t_2$ is a minimum; or, if y_1 and y_2 be the respective constant, but different, speeds in the two mediums,

$$\frac{S_1O}{v_1} + \frac{OS_2}{v_2} = \text{minimum.} \tag{3}$$

Since $S_1O = \sqrt{x^2 + y_1^2}$ and $OS_2 = \sqrt{(l - x)^2 + y_2^2}$, we have

$$\frac{\sqrt{x^2 + y_1^2}}{v_1} + \frac{\sqrt{(l - x)^2 + y_2^2}}{v_2} = \text{minimum.}$$

The left-hand member of this equation being a minimum, its partial derivative with respect to x must necessarily be zero. Performance of the differentiation yields

$$\frac{x}{v_1\sqrt{x^2 + y_1^2}} - \frac{(l - x)}{v_2\sqrt{(l - x)^2 + y_2^2}} = 0;$$

or, finally,

$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2} \tag{4}$$

If this equation, which expresses the law of refraction derived from the Fermat principle, be compared with Eq. (2), the law of Snel and Descartes, we find that[1]

$$\frac{\mu_2}{\mu_1} = \frac{v_1}{v_2} \tag{5}$$

This extremely important relation, according to which the speed of light in different mediums varies inversely as the refractive index, was finally confirmed in 1862, when Jean Bernard Léon Foucault[2] (1819-1868), in a crucial

---

[1] Fermat's original paper on the derivation of the law of refraction appears in his collected works, ed. by P. Tannery and C. Henry (Paris, 1891-1894); it is reproduced in *A Source Book in Physics (1935), pp. 278-280. See also Huygens, Traité de la Lumière (1690), Chap. III, tr. by S. P. Thompson (Macmillan, 1912).

[2] Comptes Rendus 55, 501 (1862); reproduced in *A Source Book in Physics (1935), pp. 343-344.

experiment, measured the speed of light in water and showed it to be less than that in air.

Since the refractive index varies with the color of the light, Eq.(5) also predicts that the speed of light in any particular medium is different for different colors. In any one of the substances listed in Table II, for instance, the speed of red light should be larger than that of blue light, the exact ratio of these speeds being the ratio of the refractive indexes for the two colors. This prediction was first confirmed by Albert Abraham Michelson[1] (1852-1931), who measured the speeds of red and blue light in carbon disulfide, and also in water and in air. As reference to tables of refractive indexes will show, the differences in the speeds for different colors may be considerable in the case of a solid or a liquid, but are inappreciable in the case of a gas. In a vacuum the speed of light is[2] $(2.99776 \pm 0.00004) \times 10^8 \ m\cdot sec^{-1}$, regardless of the color. That the speed in a vacuum is independent of color is confirmed by observations of various types, both terrestrial and astronomical; for instance, by observations on variable stars, and by the fact that a star does not change in color at the moment before or after its eclipse by the moon.

If the light remains in a homogeneous medium, as it does in the case of reflection, the speed v remains constant and Eq. (3) reduces to Heron's principle. All the principles and laws of geometrical optics are thus

---

1 Report of the British Association for the Advancement of Science (1884), p. 654; U. S. Nautical Almanac Office, Astronomy Papers 2 (1891), Part IV, pp. 231-258 (1885).

2 This value is based on a careful review made by R. T. Birge [Nature 134, 771 (1934)] of the data from various experimental determinations. For general accounts of the various methods for determining speeds of light, from the first astronomical method of Ole Roemer (1676) to the Mt. Wilson determinations of Michelson (1924-1926) see: *A. A. Michelson, article "Velocity of Light", Encyclopaedia Britannica (ed. 14); *A. A. Michelson, Studies in Optics (Univ. of Chicago Press, 1927); *T. Preston, The Theory of Light (ed. 5, Macmillan, 1928), Chap. XIX. The report on the "One-Mile Evacuated Pipe Experiment", which was begun in 1929 by Michelson, Pease and Pearson but not completed until after Michelson's death, appears in Astrophysical Journal 82, 935 (1935).

deducible from the Fermat principle of extreme time. If the light passes through more than two mediums in traveling between two points, then it is evidently the quantity $t_1 + t_2 + \cdots t_n$, or $\sum_{i=1}^{i=n} s_i/v_i$, that is extreme. If the medium is continuously nonhomogeneous, so that the refractive index and, hence, the speed of light changes continuously from one point to the next, then $\int_{S_1}^{S} ds/v$ is extreme, ds being the element of path, v the speed in that element, and S_1 and S the initial and terminal points of the path. This is the Fermat principle in its most general form.

Combination of Eqs. (3) and (5) yields the expression

$$\mu_1 \cdot S_1O + \mu_2 \cdot OS_2 = \text{minimum or maximum} \qquad (6)$$

for light traveling between the points S_1 and S_2 in Fig. 6. Since the quantity $\mu_1 \cdot S_1O$ or, in general, μs, is the length of the geometrical path multiplied by the refractive index of the medium, it is appropriately called the optical length of the geometrical path. Hence, Eq. (6) leads to a very useful alternative statement of the Fermat principle; namely, that the actual path taken by light in passing between two points is that one whose optical length, $\sum_{i=1}^{i=n} \mu_i s_i$, will be smallest or largest; or, more generally $\int_{S_1}^{S} \mu\, ds$ = extreme.

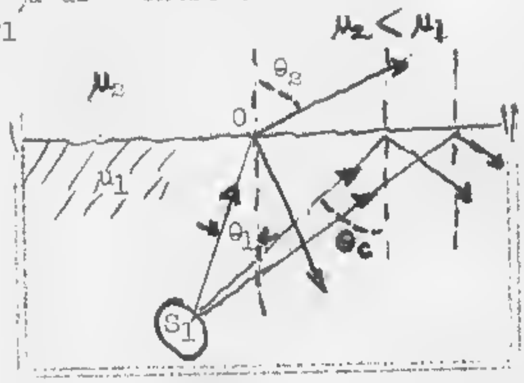


Fig. 7. The phenomenon of total reflection.

6. Total Reflection. When a ray of light passes obliquely from a medium of refractive index μ_1 to a medium of smaller refractive index μ_2, say from water to air, the ray is observed to bend always away from the normal. Thus, if S_1 in Fig. 7 is a source of light under water, a ray from it such as S_1O strikes the

water-air interface at some angle θ_1 and emerges into the air at some larger angle θ_2. This observation is in accordance with Eq. (2), which may now be conveniently rewritten in the form

$$\theta_1 = \sin^{-1} \left(\frac{\mu_2}{\mu_1} \sin \theta_2\right) \qquad (7)$$

But θ_2 cannot exceed 90°, and hence θ_1 has a maximum limit θ_c such that

$$\theta_c = \sin^{-1} (\mu_2/\mu_1) \qquad (8)$$

No ray incident on the interface at an angle larger than θ_c, which is known as the <u>critical angle of refraction</u>, can emerge into the air. It is observed, instead, to be <u>totally reflected</u>, the usual law of reflection being applicable.[1] Reflection of course occurs for all values of the angle of incidence θ_1, the percentage of reflected light increasing at the expense of the refracted light as θ_1 increases from 0° to θ_c; but for angles of incidence larger than the critical angle θ_c, all the light is reflected and there is no refracted beam. Since θ_c is determined by the refractive indexes, it differs for different pairs of mediums, and also for different colors. Totally reflected rays of, say, white light do not separate into colors, however, since angles of reflection are independent of the color. For water in a vacuum or in contact with air, μ_2/μ_1 is approximately equal to 1.00/1.33 for any color; hence, by Eq. (8), θ_c for a water-vacuum or water-air interface is approximately 48.8° for any color.[2] Precious stones are characterized by very small critical angles, and the large amount of light accordingly totally reflected by them accounts for their brilliancy.

---

[1] Kepler (Dioptrice, XIII) was the first to suspect the phenomenon of total reflection and also the first to demonstrate it experimentally. His argument is reproduced and discussed in *E. Mach, <u>The Principles of Physical Optics</u>, pp. 30-32.

[2] An interesting discussion of how the external world appears when viewed from under water, illustrated with "fish-eye views" made with a pinhole camera, will be found in *R. W. Wood, <u>Physical Optics</u> (Macmillan, 1934) pp.37-

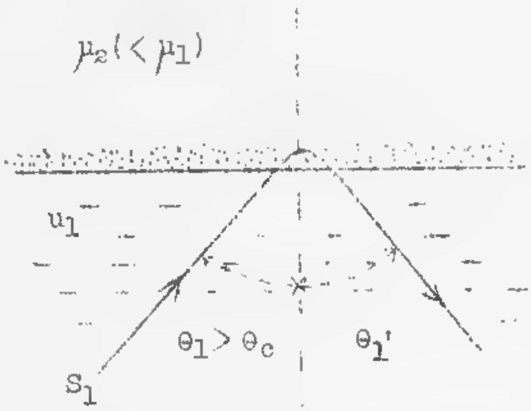


Fig. 8. Total reflection interpreted as due to refraction.

Fig. 9. One type of mirage.

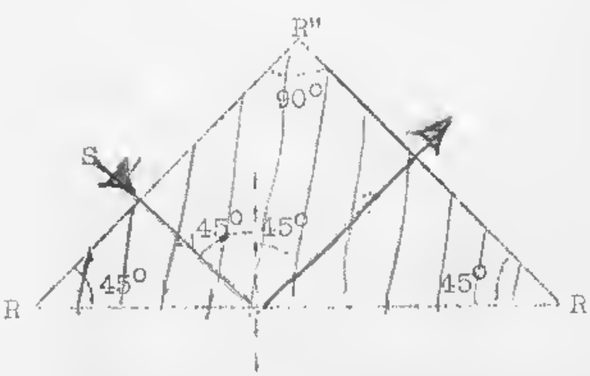


Fig. 10. A total-reflection prism.

The interface of a pair of mediums may sometimes consist of a transition layer of finite thickness, which is the result of the interpretation of the two mediums, or of occlusion at the surface. In this case it is reasonable to assume that the refractive index changes gradually, rather than abruptly, from μ_1 to μ_2. Therefore certain rays passing from the medium of the greater index of refraction into the transition layer may be deviated sufficiently so as to return into the more refractive medium, in a manner similar to that indicated in Fig. 8. A large scale phenomenon that is similarly explained is the type of mirage seen on a hot day in the desert or on a paved highway. Since the heated air next to the ground has a smaller refractive index than the cooler air immediately above it, rays from above that ordinarily would strike the ground are turned upwards and reach the eye. Thus a distant object may be seen directly and also due to the refracted rays an image of it may appear below the level of the ground.

Total reflection has many applications in optical devices, a refracting medium of prismatic form being most suitable for such purposes. For most kinds of glass in contact with air the ratio μ_2/μ_1 in Eq. (8) exceeds 1.0/1.5, and hence the critical angle of refraction exceeds 42°. Thus, if a beam of light enters a 45° - 90° glass prism at normal incidence (Fig. 10) it will strike the glass-air interface RR' at 45°, will be totally reflected and will pass out normally through the face R'R" without refraction or separation into colors. The intensity of the beam is not sensibly diminished, as is often the case with a metallic mirror, whose reflection factor, or ratio of the light reflected by a surface to that incident upon it, is usually far from unity, varies with the color and decreases as the mirror corrodes.[1] Total reflection prisms are therefore often used in place of mirrors in field glasses, submarine periscopes and many optical instruments of precision.

7. Determination of Refractive Indexes by Means of Total Reflection. The phenomenon of total reflection provides a means for determining the refractive index μ_s of any solid that can be put into the form of a prism with three polished faces. One face of the prism is illuminated with a broad, nonparallel beam of light of some particular color -- say, the yellow light from a sodium burner -- so that the light will be reflected and refracted to a telescope T in the manner indicated in Fig. 11. Those rays from the source that strike the second face RR' of the prism at angles larger than the critical angle θ_c will be totally reflected, while those that strike RR' at angles smaller than θ_c will be partially reflected and partially transmitted. If SOO' is the ray incident on RR' at exactly the critical angle of reflection, all the light that comes to the telescope T from the portion RO' of the face RR' will have undergone total reflection; and all the light that comes to T from the

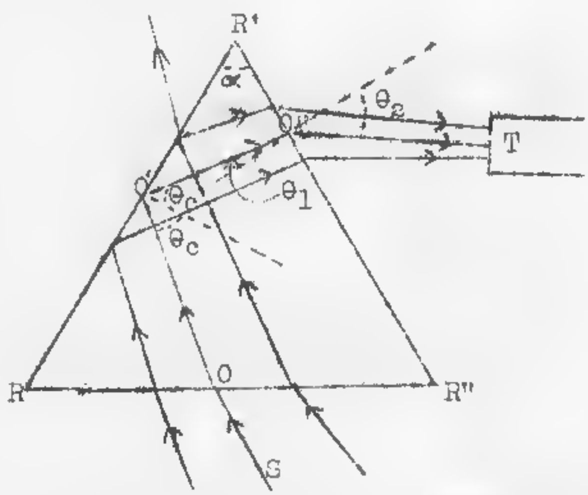


Fig. 11. Method of determining the refractive index by total reflection.

---

[1] For a discussion of various types of total-reflection prisms, see I. C. Gardner, Bureau of Standards Scientific Paper No. 550 (1927).

remaining portion O'R' of RR' will have undergone only partial reflection, the other part having been transmitted. Hence the surface RR' as seen by an observer at T should appear to consist of two parts of unequal brightness. Since the line of demarcation of the two portions is quite sharp, the angle θ_2 (Fig. 11) may be accurately measured by setting a cross hair of the telescope T on this dividing line and then rotating the telescope until it is at right angles to the prism face R'R".

To obtain the equation for computing μ_s, the refractive index of the prism material for light of the color employed, we note that the equations that describe the path of the ray SOO'O"T are

$$\sin \theta_c = \mu_f/\mu_s \tag{9}$$

and

$$\mu_s \sin \theta_1 = \mu_f \sin \theta_2, \tag{10}$$

where μ_f is the refractive index of the fluid surrounding the prism. Moreover, as can be easily proved from the geometry of Fig. 11,

$$\alpha = \theta_1 + \theta_c, \tag{11}$$

where α is the angle RR'R" of the prism, which can be easily measured. From these three equations we have only to eliminate θ_c and θ_1 in order to obtain an expression for μ_s in terms of measurable quantities; namely,

$$\mu_s = \mu_f \sqrt{1 + \left(\frac{\sin \theta_2 + \cos \alpha}{\sin \alpha}\right)^2} \tag{12}$$

If the prism is in air, μ_f is unity to the degree of accuracy with which the other quantities involved usually can be measured; that is, $\mu_s = \mu'_s$. After the refractive index of the solid material forming the prism has been determined, the refractive index μ_f of any fluid can be found by performing the foregoing experiment while the prism is immersed in the fluid and then applying Eq. (12).

8. Refractive Index Determined by the Deviation Produced by a Prism. Suppose that a beam of parallel rays of a single color is passed through a prism in the manner indicated in Fig. 12. At the first face of the prism the deviation, or change in angular direction, of any ray of the beam is $\theta_1 - \theta_2$; at the second face the deviation is $\theta_4 - \theta_3$. Hence the total deviation δ produced by both faces is $\theta_1 - \theta_2 + \theta_4 - \theta_3$. From geometry of Fig. 12 it is easily proved

$$\theta_2 + \theta_3 = \alpha \tag{13}$$

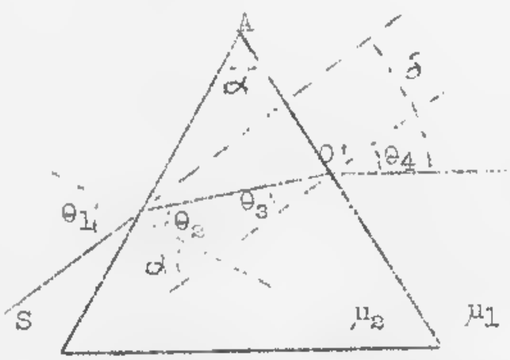


Fig. 12. Deviation produced by a prism.

where α is the angle between the two refracting faces, called the refracting angle of the prism. Therefore, the expression for the total deviation becomes

$$\delta = \theta_1 + \theta_4 - \alpha \tag{14}$$

If the direction of the beam were reversed it would travel over exactly the same path and would undergo the same total deviation δ. Hence there are two values for the angle of incidence of the ray SO on the dirst face, namely θ_1 and θ_4, such that δ is the same. Suppose then that the incident beam SO is kept fixed in direction, and the prism is rotated so as to cause the angle of incidence to vary from a value θ_1 to a value θ_4. It follows, since the deviation δ is observed to change continuously between these two values, that δ must pass through either a maximum or a minimum. Both experiment and theory[1] show that it passes through a minimum. Furthermore, since this minimum deviation δ_{min} must occur for an angle of incidence whose value lies between θ_1 and θ_4, no matter

---

[1] See Problem 16, p. 33.

how slightly these two angles differ, it must actually occur when θ_1 equals θ_4 (Fig. 13). Thus, for minimum deviation, $\theta_1 = \theta_4$, $\theta_2 = \theta_3$ and Eqs. (13) and (14) reduce to $\theta_2 = \alpha/2$ and $\delta_{min} = 2\theta_1 - \alpha$. Substituting these values for θ_1 and θ_2 in the equation $\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2$, we obtain, after rearrangement,

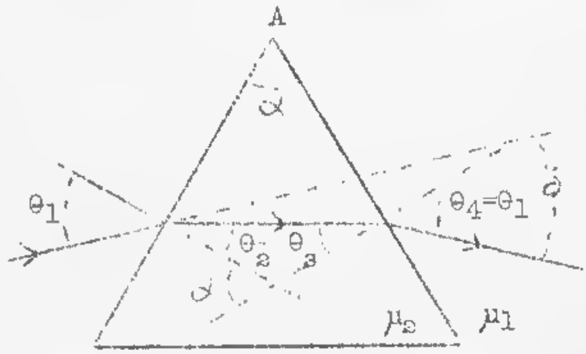


Fig. 13. The conditions for minimum deviation.

$$\mu_2 = \mu_1 \frac{\sin \frac{1}{2} (\delta_{min} + \alpha)}{\sin \frac{1}{2} \alpha} \quad (15)$$

This equation forms the basis of a convenient method for determining the refractive index μ_2 of a solid substance for light of a particular color. A beam of the desired color is sent through a prism of the substance to be tested, and the prism is rotated until the emergent beam is the least deviated. The angles δ_{min} and α can then be measured with great accuracy by means of a spectrometer,[1] and, since μ_1, the refractive index of the surrounding air, is known, μ_2 or μ_2' $(= \mu_2/\mu_1)$ can be computed by means of Eq. (15). A liquid can also be tested by placing it in a hollow glass prism, for the walls, provided they are plane-parallel, do not affect the total deviation (Sec. 4).

9. Prisms and the Dispersion of Light. Since the refractive index of any particular substance varies with the color (Sec. 4), the deviation produced by a given prism will not be the same for all colors. This gives rise to striking color effects that were doubtless familiar to the ancients who possessed glass ornaments of prismatic form. That the colors thus

---

[1] See spectrometer experiment.

formed are similar to the colors of the rainbow appears to have been noticed by the Roman philosopher Seneca; and that they are caused by refraction was probably recognized as early as the fourteenth century. But color phenomena did not receive much serious attention until the time of Descartes and Newton, when optical instruments of high magnification were coming into use and the elimination of the color fringes which blurred the images produced by these instruments became an important problem. Newton's first experiments on light date back to his student days, in 1664, and his classical researches on the analysis and synthesis of white light were begun two years later when he procured "a Triangular glass-Prisme, to try therewith the celebrated Phaenomena of Colours."[1] The experiments which he made formed the basis of all explanations of the phenomena of colors for two centuries to come.

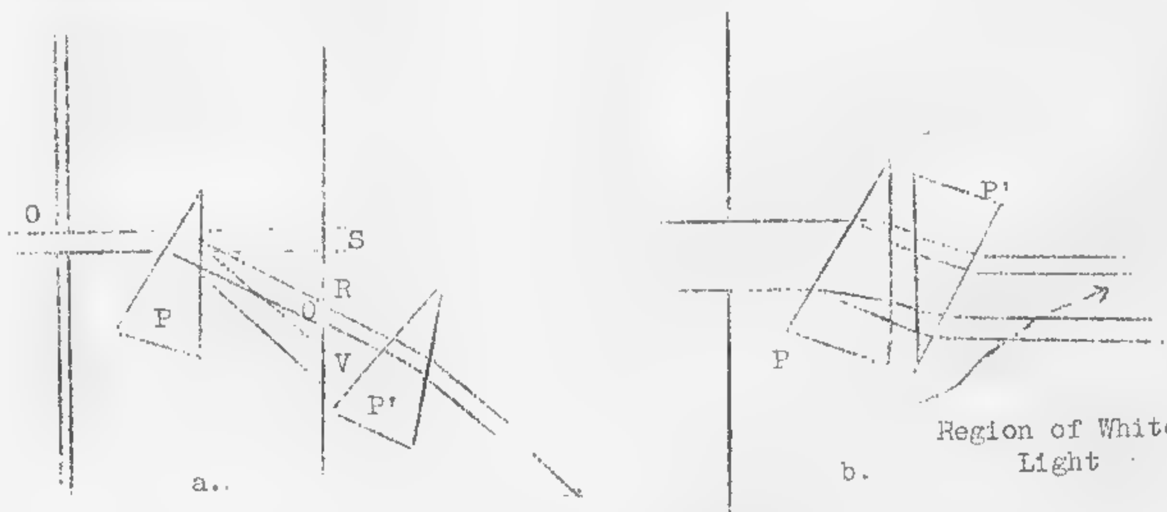


Fig. 14. The analysis and synthesis of light.

---

[1] These celebrated experiments were described by Newton in his first scientific paper, written when he was 25 years old and published in the Philosophical Transactions of the Royal Society, No. 80 (1672); the paper is reproduced in *M. Roberts & E. R. Thomas, Newton and the Origin of Colours (G. Bell, 1934), pp. 71-91 and in *A Source Book in Physics (1935), pp. 298-305. An excellent account of the experiments was later given by Newton in his classical Opticks: or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light (1704), Book One, Part I. The fourth edition, corrected (1730), has recently been reprinted (McGraw-Hill, 1931).

The experiments consisted first of all in admitting sunlight through a small circular aperture O (Fig. 14a) into a darkened room and observing that the round image S of the sun which appeared on the screen before the prism P was placed in the path of the beam became replaced upon the interposition of the prism by the band of colors RV. This band, which Newton called a spectrum, was red at the end which corresponds to the smallest amount of deviation, and changed through an infinite variety of yellows, greens and blues into violet at the other end. Newton further placed a second upright prism behind a second aperture O' in the screen and observed that it was impossible to decompose any one of the spectral colors into more elementary parts. But when he inverted this second prism (P', Fig. 14b) he found that the colors were recombined on the wall into white light.

In view of these experiments it has been customary, since the time of Newton, to regard white light as composed of a mixture of light of all conceivable colors between that of the extreme red and that of the extreme violet. As a matter of fact, Newton's experiments show, not that white light actually consists of all those colored lights, but merely that white light is decomposed by a prism into beams of these colors, and that by recombining the beams we do actually reproduce upon the retina of the eye the effect of white light. However, we are not led to any conclusions that are at variance with experiment if we adopt Newton's point of view as to the nature of white light, and we shall therefore make it the basis of much of our reasoning. We shall return to a more critical analysis of this subject in Chap. 4.

10. Production of Purer Spectra; the Fraunhofer Lines. Newton's arrangement in Fig. 14 was essentially a pinhole camera (Sec. 1) combined with a prism, and thus his spectrum consisted merely of a row of overlapping circular images of the sun in different colors; the color at any point thus

was impure, being a combination of two or more colors. Later he replaced the circular aperture by a narrow slit placed parallel to the refracting edge A of the prism and thus was able to obtain a somewhat purer spectrum.[1] It was in the purer spectrum of sunlight obtained with a slit that William Hyde Wollaston[2] (1766-1828) first noticed the existence of a number of dark lines which were parallel to the slit. He did not follow up the matter and it was left for Joseph Fraunhofer[3] to rediscover these dark lines, to make careful measurements of their positions with greatly improved apparatus, and thus to uncover a fact of the greatest importance for optical progress; namely, that they correspond to definite colors in which sunlight is always deficient, and hence can be used, instead of vaguely defined colors, as standard reference lines for measurements of refractive indexes and spectra. It is from his time that accurate knowledge of refractive indexes dates.

Fraunhofer was a skilful manufacturer of fine optical glass and instruments, and he was able to produce prisms giving much purer spectra than Newton and Wollaston were able to obtain. Furthermore, he used a telescope to view the spectrum. Addition of a lens to render the light rays from the narrow slit parallel before they enter the prism is all that is needed to make this arrangement the same as we use today in prism spectrographs (Fig. 15). The spectrum thus produced is not a row of overlapping colored images of the sun but a row of adjacent line images, in different colors, of the slit. Hence it is a very pure spectrum, although never perfectly pure, of

---

1 Newton's Opticks (ed. 4 reprinted, 1931), p. 70.

2 Philosophical Transactions (1802), p. 378.

3 "Bestimmung des Brechungs - und Farbenzerstreuungs - Vermoegens verschiedener Glasarton", Denkschriften der Koniglichen Akademie der Wissenschaften zu Muenchen fur die Jahre 1814 und 1815, Vol. 5 (1817).

course, since no slit can be made infinitesimally narrow. It is merely because a straight slit is most generally used that we speak of spectrum <u>lines</u>.

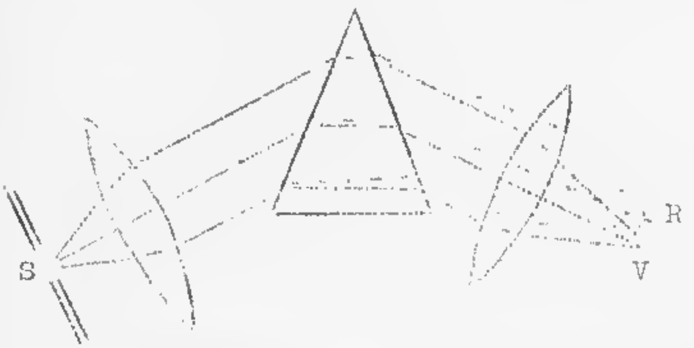


Fig. 15. Essential elements of a spectrograph.

Fraunhofer mapped the positions of some 700 dark lines in the solar spectrum, labelled the most prominent ones with letters of the alphabet, and measured the refractive indexes of various substances for the definite and reproducible colors which are absent in sunlight and which correspond to these lines (Table II). In Chap. 2 we shall see how Gustav Kirchhoff (1824-1887) was able to show that most of the Fraunhofer lines are due to the absorption of the light of the corresponding colors by vapors in the sun's atmosphere. These discoveries by Newton, Wollaston, Fraunhofer and Kirchhoff laid the foundations for the sciences of spectrum analysis and of astrophysics.

<u>Example 2</u>. Rays of white light, rendered parallel by means of a lens, are incident on a prism of refracting angle $60^\circ\ 0.0'$ which has been rotated into the position of minimum deviation for D-light. The prism is made of Bureau of Standards No. 188 light flint glass (Table II). Calculate (a) the angle between the emerging D- and F-rays and (b) the deviation of the D-rays.

<u>Solution</u>: We will employ the notation indicated in Fig. 12. For the D-light, since the deviation is a minimum, $\theta_2 = \theta_3 = 30^\circ\ 0.0'$ and $\theta_4 = \theta_1 = \sin^{-1}(\mu_D \sin\theta_2) = \sin^{-1}(1.58038 \sin 30^\circ\ 0.0') = 52^\circ\ 12.2'$; for the F-light, since the rays of the incident beam are parallel, $\theta_1 = 52^\circ\ 12.2'$. But, since the prism is not in the position of minimum deviation for F-light, θ_4 must be computed by dealing with the refraction at each prism face separately: at the first face, $\theta_2 = \sin^{-1}(\frac{1}{\mu_F}\sin\theta_1) = \sin^{-1}(\frac{1}{1.59029}\sin 52^\circ\ 12.2') = 29^\circ\ 47.6'$; at the second face, $\theta_4 = \sin^{-1}(\mu_F \sin\theta_3) = \sin^{-1} 1.59029 \sin(60^\circ\ 0.0' - 29^\circ\ 47.6') = 53^\circ\ 8.4'$. Therefore, the required angle between the emergent D- and F-rays is $53^\circ\ 8.4' - 52^\circ\ 12.2' = 0^\circ\ 56.4'$. (b) The deviation of the D-rays, since it is minimum, is $2\theta_1 - \alpha = 44^\circ\ 24.4'$.

11. Achromatic and Direct-Vision Prism Combinations. A prism that produces large deviations in the rays passed through it does not necessarily produce a correspondingly large spreading of the colors. In other words, a substance of large refractive index does not always give a spectrum of large angular width; although Newton supposed, from his early investigation of the subject, that such was invariably the case. Moreover, the reds and the yellows generally are relatively little separated, while the blues and the violets are spread considerably. Nor, indeed, are the spectra of prisms made of different materials found to agree with one another, the red C-light and blue F-light, for example, suffering a larger relative separation in one case than in another; the spectrum produced by the one prism is not simply a larger or a smaller copy of that formed by the other prism (Fig. 16).

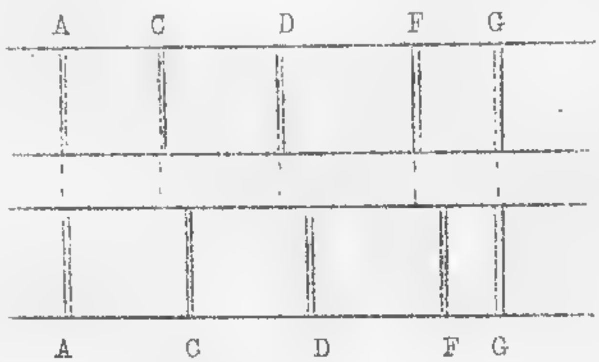


Fig. 16. Illustrating the irrationality of prismatic dispersion. If photographs of spectrums produced by two prisms of different materials are reduced to the same size, a given color does not occupy exactly the same position in both photographs, and the relative spacing of the colors is not the same.

This so-called irrationality of prismatic dispersion can best be made clear by quantitative examples. In general, the calculations must be carried out as in Example 2. But in the special, though practically important, case of a prism of small refracting angle α, a much simpler, approximate method is available; for, if α is small, then δ_{min} for rays of any given color will be small, and Eq. (15) reduces approximately to

$$\delta_{min} = \alpha(\mu' - 1), \tag{16}$$

where μ' is the refractive index of the prism material, relative to air and for the given color.

Example 3. Calculate the approximate angular separation of red C- and blue F-rays produced by a prism having a refracting angle of 10° and made of Bureau of Standards No. 76 heavy flint glass (Table II). (b) Make the same calculation for a 10°-prism made of No. 123 crown glass.

Solution: We will assume that the prism in each case is in the position for minimum deviation, so that Eq. (16) applies. (a) For the flint glass prism, $\delta_C = 10^\circ\,(1.6501 - 1) = 6.5^\circ$ and $\delta_F = 10^\circ\,(1.6691 - 1) = 6.7^\circ$; therefore, $\delta_F - \delta_C = 0.2^\circ$. (b) For the crown glass prism, $\delta_C = 10^\circ\,(1.5146 - 1) = 5.1^\circ$ and $\delta_F = 10^\circ\,(1.5233 - 1) = 5.2^\circ$; therefore $\delta_F - \delta_C = 0.1^\circ$. Thus we we see that the flint glass prism spreads out the C- and F-rays approximately twice as much as does the crown glass prism of the same refracting angle; but the deviations produced in these rays are only slightly larger in the case of the flint glass prism.

It is possible to combine two prisms of different kinds of glass, with their refracting angles turned in opposite directions, so that the combination will not produce any deviation in rays of some chosen color, and yet will spread out the colors into a spectrum (Fig. 17). Such a combination is called a direct-vision prism, and is often used in an optical instrument in which, for the sake of compactness, say, it is desired to keep the axis of the viewing telescope in approximately the same straight line with the source of light and the prism.

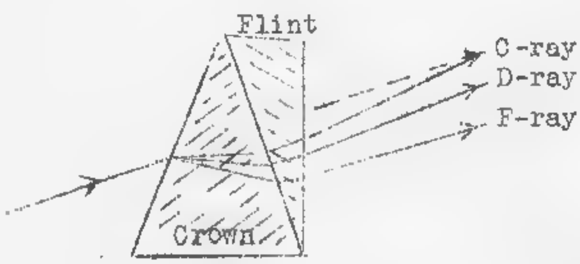


Fig. 17. A direct-vision prism.

Example 4. What refracting angle α must be chosen for a prism of No. 76 dense flint glass if it is to be combined with a 10°-prism of No. 123 crown glass (Table II) so as to form a combination that will not produce any deviation in D-light.

Solution: The requirement is that the two prisms produce equal but opposite deviations in the D-rays; or, if we assume that the prisms are kept in approximately the position for minimum deviations, that $10^\circ\,(1.5171 - 1) - \alpha(1.6555 - 1) = 0$. Therefore, $\alpha = 7.9^\circ$. This direct-vision combination will, of course, produce some deviation in rays of other colors; but, since D-light is near the middle of the spectrum, a beam of white light sent through the combination will not undergo much deviation as a whole, although it will be spread out into a spectrum.

The irrationality of prismatic dispersion also makes it possible to combine two prisms of different kinds of glass so that parallel rays of any two chosen colors will still be parallel after passing through the combination, but will undergo a net deviation (Fig. 18). Such a combination, which produces deviation without spreading the chosen colors into a spectrum, is said to be achromatic for the two chosen colors.

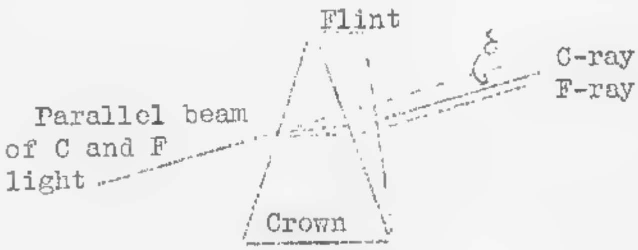


Fig. 18. A combination achromatized for C- and F-light.

Example 5. A prism of No. 76 dense flint glass is to be combined with a 10°-prism of No. 123 crown glass so that a parallel beam of C- and F-light incident on one face of the combination will emerge as a parallel beam from the other face. (a) What refracting angle α must be chosen for the flint glass prism? (b) Compute the net deviation produced in the beam.

Solution: If the deviations produced are close to minimum, Eq. (16) is applicable. (a) For the crown glass prism alone, the angle $\delta_F - \delta_C$ between the emerging F- and C-rays is $10^\circ\ (1.52326 - 1.51458) = 0.087^\circ$. If the inverted flint glass prism is to render these rays parallel, the angle $\delta_F - \delta_C$ which it produces must be 0.087°; that is, its refracting angle α must be such that $\alpha(1.66911 - 1.65007) = 0.087^\circ$. Therefore, $\alpha = 4.6^\circ$. (b) The net deviation of the beam is $10^\circ\ (1.52326 - 1) - 4.6^\circ\ (1.66911 - 1) = 2.15^\circ$.

Combinations intended for visual observation are usually achromatized for C- and F-light, since all rays of other colors will then emerge with only a slight amount of spreading. A combination of three different prisms can be chosen that will be achromatic for rays of three portions of the spectrum, and then the remaining colors will be spread still less than before. Although very little practical use is found for achromatic prisms, we shall find that the principle involved in them is of great importance in connection with achromatic lenses (Chap. 2).

Problems

1. From the law of reflection deduce the fact that a rotating mirror turns through one-half the angle through which the reflected ray is rotated.

2. A ray of light of fixed direction is incident on a plane mirror that is rotating about a vertical axis with an angular speed of ω radians per unit time. Show that the speed of the spot of light made on a vertical screen by the reflected ray is $2\omega s \sec^2 \phi$, where s is the length of the shortest line from the mirror to the screen, and ϕ is the angle between this line and the reflected ray.

3. (a) If the angle between two plane mirrors is 90°, is it possible to direct a parallel beam of light onto one of them so that the beam after two reflections will not be parallel to its original direction? (b) What must be the angle between two plane mirrors if a ray that is initially parallel to one mirror is, after two reflections, parallel to the other mirror?

4. Prove that any ray of light incident upon a system consisting of three plane mirrors joined together to form the inside corner of a cube will emerge from the system parallel to its original path.

5. Verify the truth of the following statements, the assumptions in each case being that the mirror is geometrically and optically perfect, and that the source of light employed is very small in size as compared with the size of the mirrors: (a) If a source is placed at one focus of an ellipsoidal mirror, the rays after reflection will all pass through the other focus of the ellipsoid; (b) if a source is placed at the focus of a paraboloidal mirror, the rays after reflection will be parallel to the axis of the paraboloid; (c) if a source is placed at one focus of a convex or concave hyperboloidal mirror, the rays after reflection will be so directed as to appear to have come directly from the other focus of the hyperboloid. Discuss the case of a source placed at one focus of an ellipsoid whose foci are a very large distance apart, and also the case where the foci of the ellipsoid coincide.

6. Prove that a ray passing from air through a transparent plate having plane, parallel faces undergoes a lateral displacement of amount $t \sin \theta \left[1 - \sqrt{\cos^2 \theta/(\mu^2 - \sin^2 \theta)}\right]$, where t is the thickness of the plate, μ is the refractive index of the material composing the plate and θ is the angle of incidence of the ray.

7. When a parallel beam of white light is passed obliquely through a thick glass plate having plane parallel faces, the edges of the emerging beam are observed to be colored. Explain with the help of a diagram. Do the widths of the colored edges vary with the thickness of the plate and the angle of incidence?

8. (a) In computing the total deviation produced in a ray of starlight during its passage through the atmosphere, why is it unnecessary to know the refractive index of the air for any point other than the one at

which the position of the star was observed? (b) What is the true elevation of a star above the horizon if its apparent elevation is 25°, as observed at a point where the refractive index of the air is 1.0003?

9. Prove that if the deviation δ produced in a ray by refraction at a single surface is small, then

$$\delta = (1 - \frac{\mu_1}{\mu_2}) \tan \theta_1,$$

where θ_1 is the angle of incidence, and μ_1 and μ_2 are the respective refractive indexes of the first and second mediums.

10. Is it possible to find a pair of mediums such that light incident on their interface at any angle whatever will be totally reflected?

11. A parallel beam of light is incident normally from air on the plane surface of a glass hemisphere of refractive index μ. At what distance from the center of the surface must a ray of the beam strike in order to be totally reflected?

12. (a) If S is the optical length of the actual light path S_1OS_2 in Fig. 6, Sec. 5, show that

$$\frac{dS}{dx} = \frac{\mu_1 x}{\sqrt{x^2 + y^2}} \quad \frac{\mu_2 (l - x)}{\sqrt{(l - x)^2 + y_2^2}}$$

With the help of Eq. (2), which has been verified experimentally, show that the right-hand member of the foregoing equation is zero, thus proving that the Fermat principle is valid for the case of refraction at a plane surface.

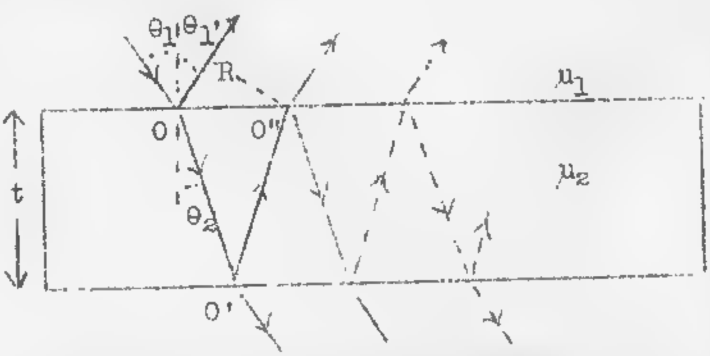


Fig. 19. Any ray SO incident on a transparent plate having plane, parallel faces gives rise to a series of reflected and refracted rays of rapidly decreasing intensity.

13. Prove (a) that the optical length of the path OO'O" in Fig. 19 is $2\mu_2 t \sec \theta_2$, where t is the thickness of the plate, (b) that the optical length of the path OR is $\mu_1 \cdot OR$, or $2\mu_2 t \sec \theta_2 \sin^2 \theta_2$, and, hence (c) that the difference in the optical lengths of these two paths is $2\mu_2 t \cos \theta_2$. (d) Express this difference in terms, not of θ_2, but of the angle of incidence θ_1.

14. (a) For any ray of light passing through a prism as in Fig. 12, Sec. 7, show that

$$\frac{d\delta}{d\theta_1} = 1 + \frac{d\theta_4}{d\theta_1} = 1 - \frac{\cos \theta_1 \ \cos \theta_3}{\cos \theta_2 \ \cos \theta_4}$$

(b) In order for δ to be a minimum or a maximum, it is necessary that $d\delta/d\theta_1 = 0$; show that this will be the case if $\theta_1 = \theta_4$. (c) Obtain the second derivative, $d^2\delta/d\theta_1{}^2$ and show that it is positive when $\theta_1 = \theta_4$, thus proving that this condition gives a minimum instead of a maximum value for δ.

15. Calculate the minimum deviation produced in a beam of light from a sodium burner when it passes through a light flint-glass prism that is immersed in water. The refracting angle of the prism is 60° 38.1', and the refractive indexes μ of the water and glass are 1.3335 and 1.6085, respectively.

16. Given a prism having a small refracting angle α, show that all rays incident on the first face at a small angle will undergo approximately the same total deviation, $\alpha(\mu_2 - \mu_1)/\mu_1$.

17. Prove that the refracting angle α of a prism must be less than $\sin^{-1}(\frac{\mu_1}{\mu_2}\sin\theta_1) + \sin^{-1}(\frac{\mu_1}{\mu_2})$ if a ray of light incident on the first face at an angle θ_1 is to be transmitted through the second face.

18. A ray of light SO (Fig. 20), incident from air on the surface of a spherical raindrop in a plane through the center of the drop is refracted at O, reflected at O' and refracted into the air at O". (a) Show that the deviation δ produced in the ray is $\pi - \phi$, or $\pi - (4\theta_2 - 2\theta_1)$. (b) Show that $d\delta/d\theta_1 = -4(d\theta_2/d\theta_1) + 2 = (-4\cos\theta_1/\mu\cos\theta_2) + 2$, where μ is the refractive index of water. (c) Show that minimum deviation occurs when

$$\theta_1 = \cos^{-1}(\tfrac{1}{2}\mu\cos\theta_2) = \cos^{-1}\sqrt{(\mu^2 - 1)/3}\,.$$

(d) Show that the minimum deviation for red light ($\mu = 1.329$) is 137.2° and for violet light ($\mu = 1.343$), 139.2°. The foregoing results provide the basis for the explanation of the primary rainbow, given first by Descartes and Newton.

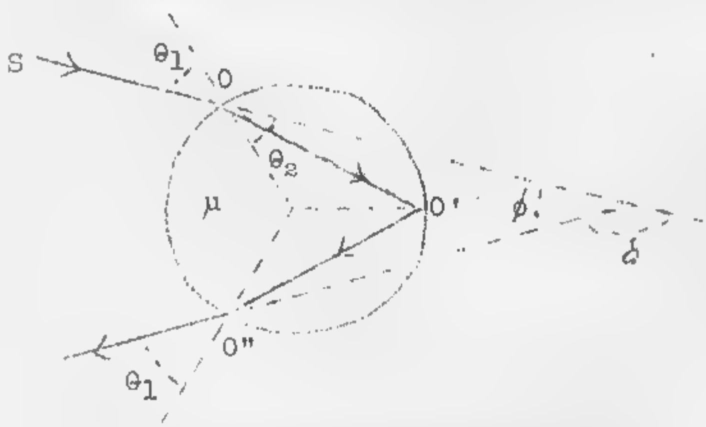


Fig. 20. Refraction and reflection of light by a spherical raindrop.

19. A parallel beam of white light is incident from air at an angle of 45° on one face of a prism of refracting angle 60° 0'. If the refractive indexes μ' of the prism material are 1.622 and 1.635 for D- and F-light, respectively, at what angles do the D- and the F-rays emerge from the prism?

20. When a certain 60°-prism is set in the position of minimum deviation for C-light, the angular separation of emerging C- and F-rays is observed to be 2°. The refractive index u' of the prism material is 1.58 for C-light. Calculate (a) the minimum deviation for C-light, (b) the refractive index u' for F-light and (c) the minimum deviation for F-light.

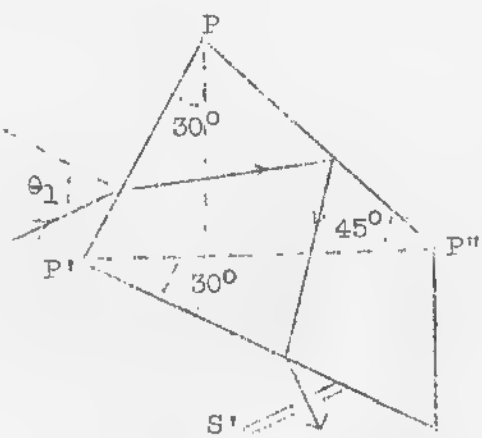


Fig. 21 A constant-deviation prism, as used in a monochromator, which is a device for illuminating an aperture with the different colors of the spectrum in succession.

21. A parallel beam of white light is incident on the prism shown in Fig. 21, which is a single piece of glass but which may be thought of as consisting of two prisms of refracting angle 30°, together with a total reflecting prism PP'P". (a) Show that the ray which undergoes a total deviation of 90°, and which thus passes through the center of the aperture S', consists of that single color for which the angle of incidence θ_1 is the correct angle for minimum deviation for a 60° prism of the same material. (b) Show that if the prism be rotated about an axis perpendicular to the paper while the aperture S' and the incident beam are kept fixed in position, the aperture will be illuminated in succession by the different colors of the spectrum. (c) What value must be given to θ_1 in order that the aperture be illuminated with D-light, the refractive index of the glass, relative to the air, being 1.58038 for this color?

22. If the direct-vision combination described in Example 4, Sec. 11, is set approximately in the position for minimum deviation and white light is passed through it, what will be the angular width of the spectrum between the Fraunhofer C- and G-lines?

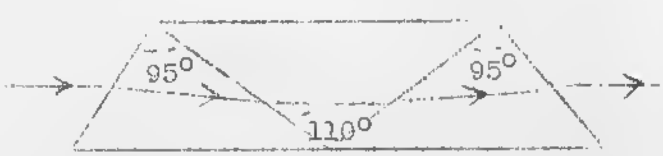


Fig. 22. A widely used type of direct-vision prism.

23. In the combination shown in Fig. 22, the two outside prisms consist of crown glass for which μ'_D is 1.520. (a) What must be the refractive index μ'_D of the material in the middle prism if a parallel beam of D-light is to enter and emerge from the combination parallel to its base? (b) What is the

function of the second crown glass prism? Why would one expect this combination to produce a spectrum of relatively large angular width?

24. The combination described in Example 5, Sec. 11, is achromatized for C- and F-light. Show that it is not perfectly achromatized for G-light.

25. (a) Calculate the refracting angle of a prism of No. 76 flint glass if it is to be used with a 5.0°-prism of No. 123 crown glass to provide a combination that is achromatic for C- and F-light. (b) If a parallel beam of white light is sent through this combination when it is set in the position of minimum deviation for C- and F-light, what will be the deviation of the C-, the D-, the F- and the G-rays?

26. The ratio $(\mu_F - \mu_C)/(\mu_D - 1)$ is defined as the dispersive power ω of a substance for the spectral region between the Fraunhofer C- and F-lines. (a) Prove that, for a prism having a small refracting angle, the dispersive power of the prism material is equal to $(\delta_F - \delta_C)/\delta_D$, where δ_D evidently is the minimum deviation for a line near the middle of the spectral region considered. (b) Compute the dispersive powers of No. 123 crown glass and No. 76 dense flint glass. (c) Two prisms of the crown and of the flint glass have small refracting angles so chosen that each prism produces the same minimum deviation in D-light; compare the angular widths $\delta_F - \delta_C$ of the spectra which the two prisms will produce.

CHAPTER TWO

FORMATION OF IMAGES

Since reflecting and refracting surfaces are employed as parts of optical instruments for the purpose of forming images, it is essential to investigate the rules and other predictions concerning image formation that are deducible from the basic generalizations of geometric optics, and to determine the extent to which these deductions are compatible with experience and realizable in practice. It will be found helpful if we first gain clear ideas as to what is meant by an optical image.

12. Object-Point and Image-Point. Any object that is visible, either because it is self-luminous or because light is being reflected from it diffusely, may be regarded, for purposes of study, as a collection of luminous object-points distributed over a surface. From any particular object-point, rays of light diverge in all directions. If, as in Fig. 25 (a), two or more of these rays from an object-point S are intercepted by an optical system which, by reflection or refraction, changes the course of the rays so that they converge and eventually intersect one another at some point S', the rays will

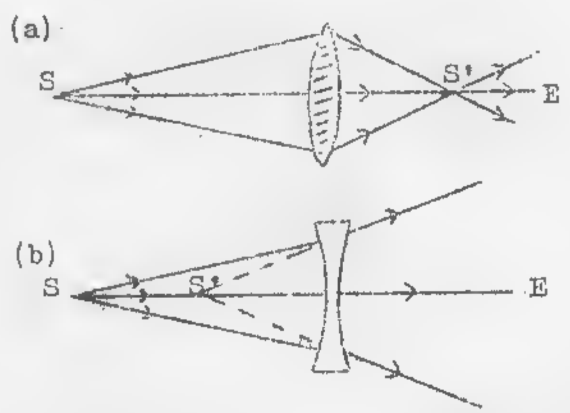


Fig. 25. In (a) the pencil of rays emitted at the object-point S is refracted at the two surfaces of the lens so as to produce a real image-point at S'. In (b) the image-point S' is virtual, for it is not the rays themselves, but only their directions produced backward, that pass through S'. In the case of either lens, an observer stationed on the right-hand side at E would see the image S', not the object S.

appear to the eye which afterwards receives them to have actually originated at this point of intersection. Any such point at which two or more of the rays intersect is called a real image-point of the object-point. On the other hand, if the rays diverging from an object-point are intercepted by an optical system which, although it changes the courses of the rays, does not render them convergent, then the rays cannot intersect to form a real image-point. Nevertheless, as can be seen in Figs. 25(b) and 26, the rays will appear to the eye which receives them to have diverged from a point or points generally different in location from the object-point. Such a point, from which two or more of the rays merely appear to diverge without having actually passed through it, may be conveniently thought of as a virtual image-point. The essential difference between a real and a virtual image is that the former can be received on a screen and made visible, whereas the latter cannot.

An obvious requirement for ideal image formation is that the rays from any particular object-point either all converge to and afterwards diverge from a single real image-point, or else all appear to diverge from a single virtual image-point.

By a real image of an object of finite size is meant the aggregate of the real image-points of the component object-points. Similarly, a virtual image of a finite object is the aggregate of the virtual image-points of the component object-points. Now, any finite image might be regarded as perfect, in the ordinary sense of the word, if a one-to-one correspondence existed between the points of the image and object and if, moreover, the spatial distribution of the image-points were precisely the same as that of the object-points; or, to restate the latter requirement, if the magnification, which is defined as the ratio of any linear dimension of the image to the corresponding dimension of the object, were exactly the same for every part of the image. However, we shall see that it is not possible to produce such a literally perfect image of a finite object except under special circumstances of very limited practical value. Thus there has arisen in optical practice the concept of an optically perfect image, which meets the somewhat more modest requirements that points, lines and planes of the object have corresponding to them points, lines and planes of the image. Any departure from these requirements for an optically perfect image is called aberration.

Formation of Images by Mirrors

The mirrors used in optical instruments are either silvered on the front surface or are made of black glass, thus eliminating all refraction of light with its attendant complications. That this is advantageous was foreseen by Newton when he invented the first reflecting telescope in 1668.[1]

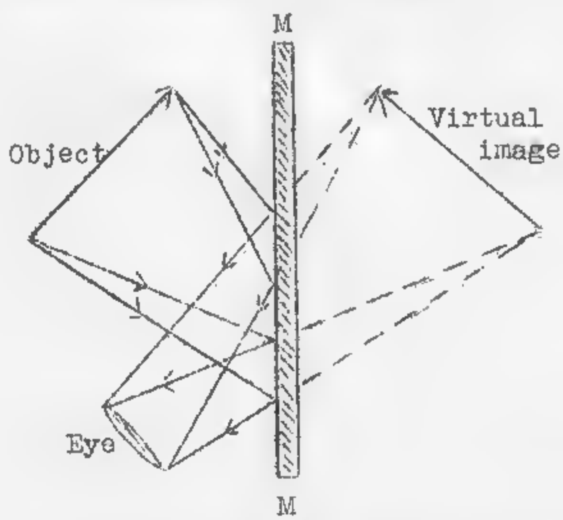


Fig. 26. Character of the image formed by a plane mirror as determined by an actual graphical construction based on the laws of reflection.

Experiment and theory show that a flat mirror, if it is accurately plane, forms images that are entirely free from aberration, no matter how large the object may be or where it is placed in front of the mirror. But, as also can be easily demonstrated or deduced from the laws of reflection (Fig. 26), the image formed by a plane mirror always is virtual, always is at

[1] Newton, Opticks (ed. 4 reprinted, 1931), p. 102 ff. The first reflecting telescope actually was proposed by James Gregory in 1663, but he had not succeeded in constructing the instrument practically.

the same distance behind the mirror as the object is in front of it, and always has a magnification equal to unity. Thus the utility of any plane mirror as part of an optical instrument is extremely limited. Curved mirrors, on the other hand, are generally not free from aberration but have other distinctive properties that make them highly useful for optical purposes. Some of these properties apparently have long been known, for we find it noted in the *Optics* of Euclid that concave mirrors turned toward the sun will ignite objects situated at the image-point,[1] and Archimedes is said to have thus utilized them in the defense of Syracuse against the Romans.

13. *Spherical Mirrors*. Mirrors whose surfaces are portions of spherical surfaces are of great importance, not merely because the geometry involved in their theory is relatively simple, but because they are comparatively easy to construct and test. A spherical surface is attained automatically when two surfaces are rubbed together, since they will fit together in all positions only when they have acquired a uniform curvature. On the other hand, an *aspheric* surface -- that is, one that is neither plane nor spherical, being usually some quadric surface -- can be obtained with precision only by local grinding with continual testing. Aspheric mirrors of relatively crude construction are used in searchlights, motion-picture projectors and other similar devices where the image-forming requirement is not very severe; and aspheric mirrors of the highest accuracy are employed in astronomical telescopes of the reflecting type. But in commercial instruments of precision quality, spherical surfaces are the most practicable and are used almost exclusively.

In Fig. 27, *MM* represents a central cross-section of a *concave spherical mirror* of radius *R*, the center of curvature being at *C*. The point *O* at the

---

[1] Euclid's *Optics*, theor. 30.

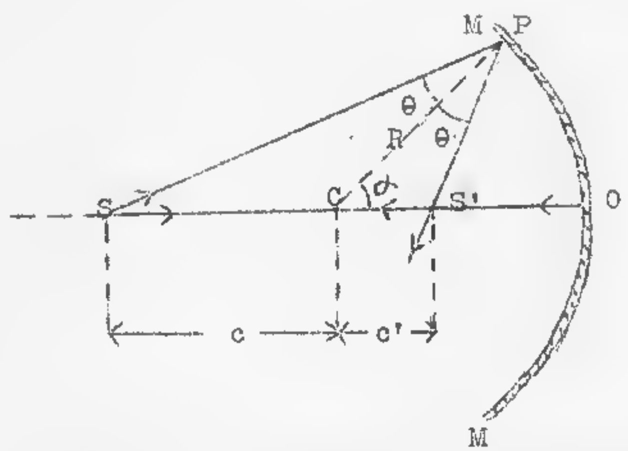


Fig. 27. A concave spherical mirror.

center of the reflecting surface is called the vertex of the mirror, and the straight line imagined drawn through the vertex O and center of curvature C is the axis of the mirror. Suppose that an object-point S is placed on the axis at a distance c from the center of curvature O. Let SO and SP represent two rays diverging from S, the former along the axis so that it is incident on the mirror at O, and the latter to any other point P of the mirror such that the arc PO subtends an angle α at C. After reflection, the two rays intersect and thus form an image-point S' which is on the mirror axis at some distant c' from the center of curvature C. Our problem is, first, to derive an expression that will enable us to compute the position of this image-point S' and, second, to determine whether or not all the other rays from the object-point S are reflected to this same image-point S'; that is, to determine whether c' is the same for all values of the angle α.

Now, any curved surface may be regarded as made up of an infinite number of plane surfaces to each of which the law of regular reflection applies. Thus, in Fig. 27, $\angle \overline{SPC} = \angle \overline{S'PC} = \theta$. By the law of sines,

$$\frac{R}{c} = \frac{\sin(\alpha - \theta)}{\sin\theta} = \sin\alpha\cot\theta - \cos\alpha,$$

$$\frac{R}{c'} = \frac{\sin(\alpha + \theta)}{\sin\theta} = \sin\alpha\cot\theta + \cos\alpha;$$

therefore, $(R/c) - (R/c') = -2\cos\alpha$, or

$$c' = \frac{Rc}{R + 2c \cos \alpha} \qquad (17)$$

This is the desired expression, and from it we learn that c' is not the same for all values of tbe angle α; the larger is the value of α corresponding to a particular incident ray, the nearer to the vertex O does it intersect the axial ray after reflection. Thus the various rays from a given object-point S' are reflected so as to form with the reflected axial ray, not a single image-point S', but a short line of image-points X'X" (Fig. 28). What is more, any two rays reflected from neighboring points P' and P" of the mirror surface intersect each other before they reach the axis. Thus, for a single object-point on the axis there exists a whole family of extra-axial image-points, in addition to the image-points formed along the axis (Fig. 28).

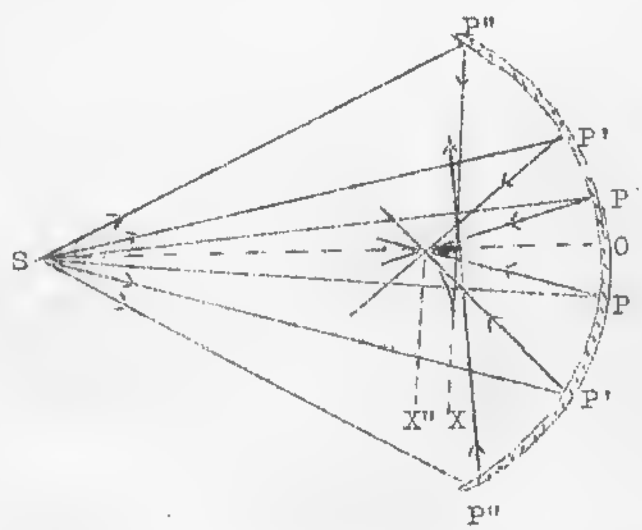


Fig. 28. This diagram illustrating spherical aberration in a spherical concave mirror can be constructed by plotting Eq. (17) or, more simply, by drawing diverging rays from the object-point S and constructing angles of reflection equal to angles of incidence.

Since Eq. (17) can be shown to hold for an object-point located anywhere on the axis of either a concave or a convex (Fig. 29) spherical mirror, it follows that no spherical mirror is entirely free from aberration, and that this would be true even if the mirror surface were perfectly spherical. These predictions from theory are fully confirmed by experiment.

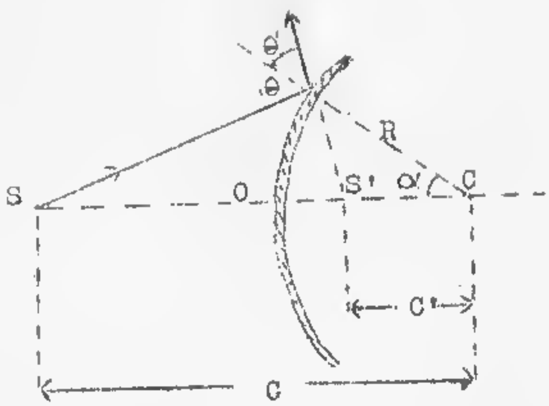


Fig. 29. A convex spherical mirror.

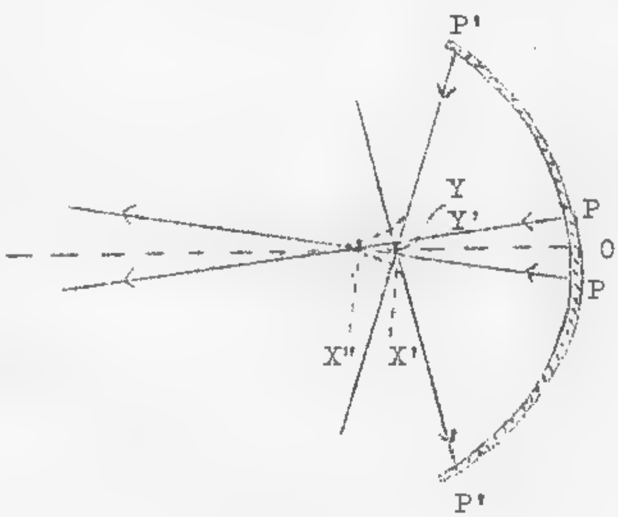


Fig. 30. This diagram differs from that of Fig. 28 in that only the reflected rays from near the margin and near the vertex of the mirror are shown.

The end-points X'X" of the short line of axial image-points in Fig. 28 or 30 are seen from Eq. (17) to correspond to rays reflected, respectively, from the margin of the mirror, for which α is maximum, and from points very close to the vertex O, for which α is very small. Thus the distance X'X" may be regarded as a measure of the aberration of the marginal ray, and is termed the longitudinal spherical aberration. As for the extra-axial image-points, formed by pairs of rays reflected from neighboring points on the mirror and intersecting before they reach the axis, they are found to lie on a curve called a caustic curve, the form of which varies with the position of the object-point S on the axis. Figures 28 and 30, it must be remembered, are drawn in a single plane, whereas the mirror actually is a portion of a spherical surface. But, by imagining either figure to be rotated about the axis OC, we can invisage the complete situation; namely, a solid cone of light diverging from the object-point S to illuminate the whole face of the mirror, and any two neighboring rays of this cone so reflected that they intersect on the caustic surface which was traced out by the caustic curve during the imagined rotation of the diagram. The cusp of the caustic surface is at X" and is evidently the image-point formed by rays corresponding to very small values of the angle α.

In Fig. 30, imagine a small screen held between the mirror and the object-point so as to intercept the reflected rays. The disk of intercepted light evidently would be smallest in diameter when the screen was held in the plane YY', where the rays reflected from the margin of the mirror cut the caustic surface. Held nearer to the mirror, the screen would intercept a larger disk having a bright edge; held farther away, also a larger disk, but in this case with a bright center. The smallest disk, at YY', is called the circle of least confusion. It is the nearest approach to a single image-point obtainable with a spherical mirror.

These ideas regarding spherical aberration also pertain when the image-points are virtual, instead of real, except that they cannot be demonstrated by placing a screen at the image-points; the reflected light does not actually pass through a virtual image-point and hence nothing is there to illuminate a screen.

14. The First-Order Theory of Spherical Mirrors; Paraxial Rays.

If the radius R of a spherical mirror is very large, or if an opaque plate containing a small opening, called a stop, is placed centrally between the mirror and the object-point (Fig. 31), then only those rays that correspond to small values of the angle α will reach the mirror and be reflected from it. If the largest value of α does not exceed 2° or 3°, the resulting image is found to be quite well defined,[1] for the only rays now incident on the mirror are those that are reflected so as to intersect near the point X'' (Figs. 28 and 30). This result is predicted by eq. (17); for, by Maclaurin's theorem,

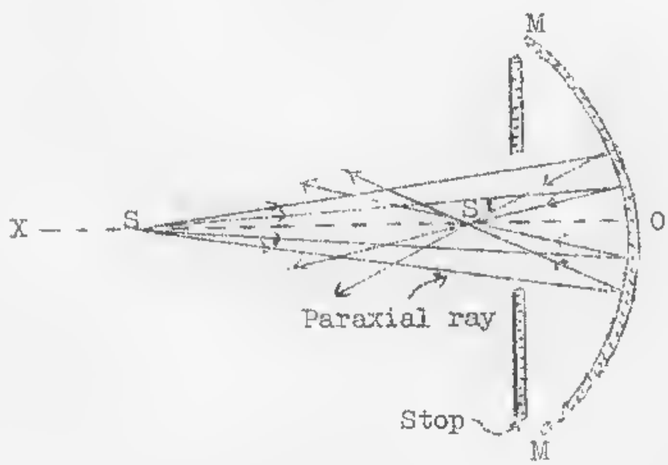


Fig. 31. If only paraxial rays are allowed to reach the mirror, the spherical aberration is greatly reduced, although at the expense of the brightness of the image.

[1] This case appears to have been first discussed by Alhazen, Opticae thesaurus, Bks. 4-6. Alhazen also clearly had an understanding of longitudinal spherical aberration in spherical mirrors.

$$\cos \alpha = 1 - \frac{\alpha^2}{2!} + \frac{\alpha^4}{4!} - \frac{\alpha^6}{6!} + \cdots, \tag{18}$$

and hence, when α is small, Eq. (17) reduces approximately to

$$c' = \frac{Rc}{R + 2c} . \tag{19}$$

Thus, when α is sufficiently small, c' has practically the same value for all the rays from a given object-point on the axis. These rays that produce a sharp image are seen not only to make very small angles with the axis of the mirror but to lie close to the axis throughout their length. Any ray that fulfills both of these conditions is said to be paraxial.

In our further treatment of spherical mirrors we shall find it convenient to employ a rectangular coordinate system, with the vertex O of the mirror taken as the origin, and the axis of the mirror on the reflecting side taken as the positive X-axis. In harmony with this convention, the radius of curvature of any concave mirror is to be regarded as positive, and that of any convex mirror, as negative. If the distances OS and OS' from the vertex to the object-point and image-point, respectively, are denoted by x and x', so that c and c' (Fig. 27) become $x - R$ and $R - x'$, respectively, then Eq. (19) can be transformed into the simple and useful relation,

$$\frac{1}{x} + \frac{1}{x'} = \frac{2}{R} . \tag{20}$$

Thus, for the comparatively sharp images that can be obtained with a spherical mirror when only paraxial rays are utilized, the sum of the reciprocals of the object distance and the image distance is equal to twice the curvature of the mirror.

<u>Example 6</u>. A concave mirror of radius of curvature 0.20 m is provided with a stop suitably placed to exclude rays that are not paraxial. Determine the position and character of the image-point when the object-point is located on the axis and at a distance from the mirror of (a) 0.30 m, (b) 0.15 m, (c) 0.05 m.

<u>Solution</u>. The mirror is concave and therefore $\underline{R}$ in Eq. (20) is always positive: (a) $(1/0.30) + (1/x') = 2/0.20$, or $x' = 0.15$ m. Since this image distance is positive, the image-point is on the same side of the mirror as the object-point, and hence is real; the reflected rays actually pass through and diverge from it, as if the object-point itself were there.

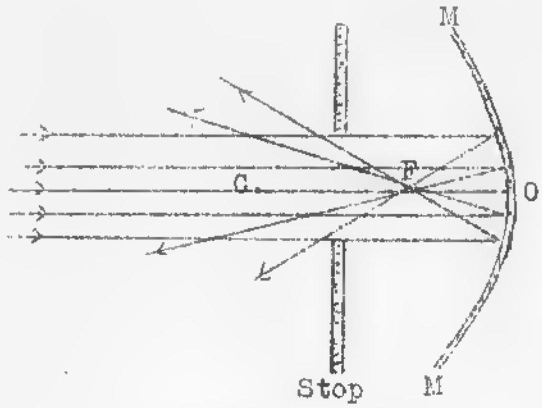
(b) $(1/0.15) + (1/x') = 2/0.20$, or $x' = 0.30$ m. This image-point also is real. Comparison of (a) and (b) shows that the image-point and object-point in these two cases have simply exchanged places, which could have been anticipated, in view of the principle of reversibility of light rays (Chap. I).

(c) $(1/0.05) + (1/x') = 2/0.20$, or $x' = -0.10$ m. Since this image distance is negative, the image-point is behind the mirror, and therefore is virtual; the reflected rays diverge from the mirror, but if produced backward, intersect at a point on the axis 0.10 m behind the mirror.

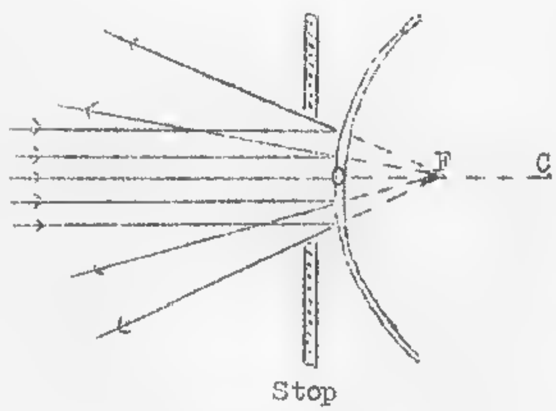
<u>Example 7</u>. Solve Example 6 for the case of a <u>convex</u> mirror of radius -0.20 m, the other conditions being the same as before.

<u>Solution</u>. (a) $x' = -0.075$, (b) $x' = -0.060$ m, (c) $x' = -0.033$ m. In every case the image is behind the mirror and is virtual.

Instead of describing a particular mirror in terms of its curvature $1/R$, it is often convenient to employ another constant of the mirror called the <u>focal length</u>, which is obtained by considering the limiting case of an object, such as the sun, that is practically at an infinite distance from the mirror. If the object is at infinity on the axis, $1/\underline{x}$ is zero and $\underline{x}'$ is equal to $R/2$. All the incident rays are then parallel (Why?), and after reflection those that are paraxial converge to, or appear to diverge from, a point $\underline{F}$ midway between the vertex and center of curvature of the mirror (Fig. 32). This point $\underline{F}$ is called the <u>focal point</u> of the mirror, and the distance $\underline{OF}$, usually denoted by $\underline{f}$, is known as the <u>focal length</u>. By substituting $\underline{f}$ for $R/2$ in Eq. (20), we obtain the mirror equation



(a) Concave mirror



(b) Convex mirror

Fig. 32. Illustrating the concepts of focal point F and focal length f.

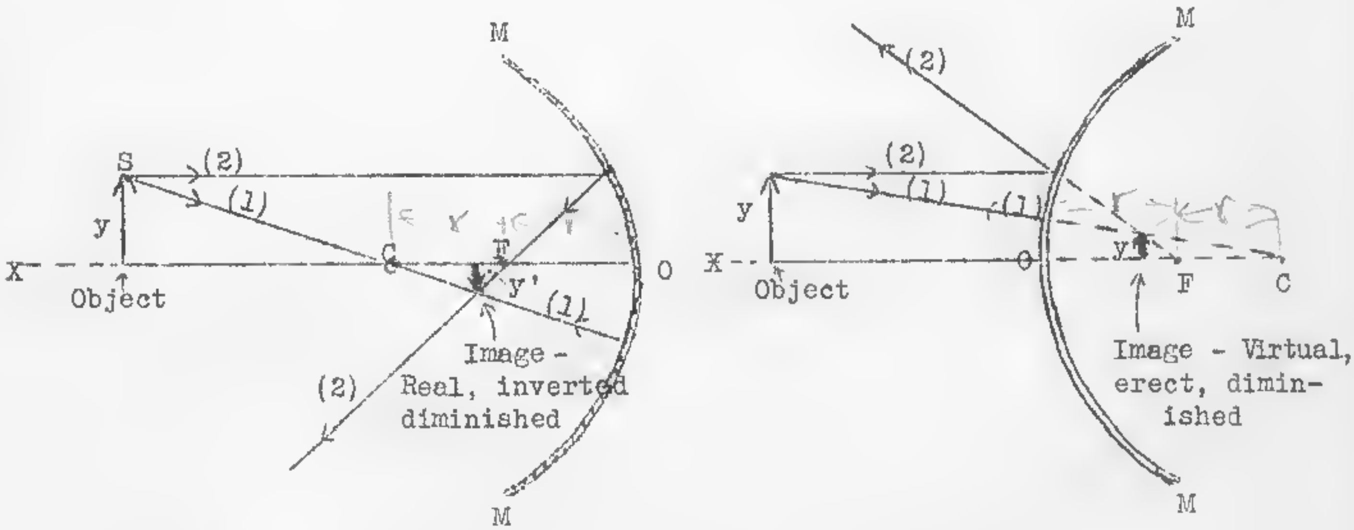
for paraxial rays in the useful alternative form,

$$\frac{1}{x} + \frac{1}{x'} = \frac{1}{f} . \qquad (21)$$

The reciprocal of the focal length, namely, $1/f$, is termed the dioptry, or "focal power," of the mirror. A commonly employed unit of dioptry is 1 m^{-1}, which is called 1 diopter. For example, a convex mirror of radius of curvature -0.50 m has a dioptry of -4.0 diopters.

15. Images of Finite Objects Formed by Paraxial Rays; Magnification. So far only axial object-points have been considered. If the object-point is not on the axis, its image will also be extra-axial and can always be located graphically, by drawing a bundle of rays diverging from the object-point and tracing the corresponding reflected rays with the help of the law of reflection. In practical cases where the object-point is close to the axis and only paraxial rays are allowed to reach the mirror, Eq. (20) or (21) usually is adequate for locating the position of the image-point. However, errors of sign are sometimes made in this computation, so it is better not to rely on it solely, but to check the result by graphical construction. This can be done quickly if it will

be remembered that at least two of the rays diverging from any object-point on or close to the mirror axis have paths after reflection that are immediately predictable: (1) the ray that passes through, or is directed toward, the center of curvature C, since it is always reflected on itself; (2) the ray that is parallel to the axis, since it is always reflected so as to pass, or appear to pass, through the focal point F (Fig. 33).



(a) Concave mirror (b) Convex mirror

Fig. 33. A rapid graphical method for determining the position and character of an image formed by paraxial rays.

If an object-point is at a small distance y from the mirror axis, the foregoing construction gives the distance y' of its image-point from the axis (Fig. 33); this enables one to compute the ratio y'/y, which is known as the lateral magnification. A formula for this ratio can also be deduced. From the geometry of, say, Fig. 33(a),

$$\frac{-y'}{y} = \frac{R - x'}{x - R} = \frac{x'(\frac{R}{x'} - 1)}{x(1 - \frac{R}{x})},$$

and this, in view of Eq. (20), reduces to

$$\frac{y'}{y} = -\frac{x'}{x}\ . \tag{22}$$

Thus the lateral magnification is equal to the negative ratio of the image distance and object distance. The negative sign in Eq. (22) indicates that the image will be inverted if x and x' have the same sign, and erect if they have opposite signs.

Example 6. Verify the information given in Table III by making the appropriate graphical constructions, and also by employing Eqs. (20), (21), and (22) whenever possible.

Table III. Classification of the images formed by spherical mirrors

Position of object	Position of image	Magnification y'/y	Character of image
Concave mirrors			
At ∞	At F	0	Real
Between ∞ and C	Between F and C	$< 0, > -1$	Real, inverted
At C	At C	$= -1$	Real, inverted
Between C and F	Between C and ∞	< -1	Real, inverted
At F	At ∞		
Between F and O	Between -∞ and O	> 1	Virtual, erect
Convex mirrors			
At ∞	At F	0	Virtual
Between ∞ and O	Between F and O	$> 0, < 1$	Virtual, erect

16. Concerning Third-Order Aberrations. Eq. (17) holds strictly only for object-points situated on the axis of a spherical mirror, so that at best it gives a good approximation when all the points of the object and image lie very close to the axis. By setting cos α equal to unity in Eq.(17), the first-order theory of Secs. 14 and 15 is obtained; it holds only for paraxial rays and shows that the images formed by such rays practically meet the requirements for an optically perfect image (Sec. 12).

In practice, however, a large part of the image-forming rays may be nonparaxial, and questions arise as to what extent the resulting images depart from optical perfection, and what can be done to a mirror system to reduce or eliminate these departures. Now, in view of the Maclaurin theorem Eq. (18), if cos α in Eq. (17) were set equal, not to unity, but to $1 - \frac{\alpha^2}{2}$, the resulting equation should provide the basis for a more accurate theory, although still applicable only to object-points on or very close to the mirror axis. In fact, Ludwig von Seidel (1821-1896), by proceeding along such lines, and also taking into account object-points distinctly off the axis and the rays from them that never intersect the axis, worked out in 1855 the so-called third-order theory for both reflecting and refracting surfaces. The quantitative deductions and applications of this theory are so extensive and detailed as to be mastered only after prolonged study and experience with optical instruments, and hence are far beyond the scope of this book. The third-order theory predicts, and observations confirms, that an image formed by a spherical mirror is subject to five different types of aberration. Four of these types -- coma, astigmatism, curvature of the image and distortion of the image ---occur, or become important, only for extra-axial object-points. The remaining type -- spherical aberration -- has already been shown in Sec. 13 to occur for object-points lying on the axis. These aberrations are not due to lack of sphericity or to mechanical defects in the mirror surface, for all of them could exist in a mirror that was perfect in these respects. However, they can often be minimized by choosing a mirror of suitable radius of curvature, by introducing stops at suitable places, by changing the location of the object, or by employing a mirror of nonspherical form.

17. Nonspherical Quadric Mirrors. Although no curved mirror, spherical or otherwise, will produce an image that is entirely free from spherical aberration for every position of the object-point on the axis, such a perfect image is possible with certain curved mirrors for a limited number of positions of the object-point. In fact, this is true of a spherical mirror for the single case of an object-point located at the center of curvature, but the case lacks utility because the image-point and object-point coincide. Of real utility in this respect, however, are the following quadric surfaces, each of which can easily be shown to produce an image that is entirely free from spherical aberration for certain definite positions of the object-point.[1]

(a) An object-point located at either focus of an ellipsoidal mirror has a single, real image-point at the other focus. Although ellipsoidal mirrors have found limited use in theatrical lighting, in a special type of projection lantern and in the now obsolescent Gregorian telescope,[2] they are not well adapted for use in optical instruments because other aberrations exist for object-points that are even a short distance from the focus.

---

[1] See Prob. 5, Chap. One.

[2] For a brief description of various types of telescopes, see D. Gill and A. S. Eddington, article "Telescope," Encyclopaedia Britannica, ed. 14.

(b) When an object-point is located at the focus of a paraboloidal mirror, all the reflected rays are parallel to the axis; conversely, an object-point a great distance away on the axis has a single real image-point at the focus. Paraboloidal mirrors are used in headlights and searchlights, and in all modern reflecting telescopes.[1]

(c) An object-point located at one focus of either a convex or a concave hyperboloidal mirror has a single virtual image-point at the other focus. A small convex hyperboloidal mirror is used in conjunction with the large paraboloidal mirror in the Cassegrain telescope.[1]

Formation of Images by Refracting Surfaces

Refracting surfaces have also long been used to produce images. Magnifying glasses seem to have been used by the ancient Chaldeans; "burning" glasses are described in Greek and Roman writings, and convex lenses have been found in Pompeii and among Roman ruins in England. The fact that a segment of a glass sphere will produce magnified images was known to Alhazen[2], who appears also to have been the first to recognize that the eye is, in modern words, a tiny camera with a transparent lens producing an image on the retina. Before the end of the thirteenth century tools for grinding spherical surfaces had come into use, and with them the first spectacles made their appearance. Some three centuries later Kepler worked out many valuable

---

[1] For a brief description of various types of telescopes, see D. Gill and A. S. Eddington, article "Telescope", Encyclopaedia Britannica, ed. 14.

[2] Alhazen, Opticae thesaurus, vii, 48.

and interesting properties of refracting surfaces[1] while developing the theory of the telescope and, as we already know (Chap. I), did this without benefit of the exact law of refraction or of the laws of image formation familiar to us today. Kepler was aware of the complications now classed as spherical aberrations, and both he and Descartes suggested that they could be reduced by making the surfaces of lenses hyperbolic, instead of spherical, in shape. But hyperbolic surfaces are difficult to construct; moreover, Newton soon showed that the most troublesome defects of the optical instruments then in use were to be attributed less to spherical aberration than to effects that could not be corrected simply by resorting to a lens of another shape. Nowadays lenses with spherical surfaces, or sometimes cylindrical surfaces, are still the most widely used and provide geometrical optics with its most important problems.

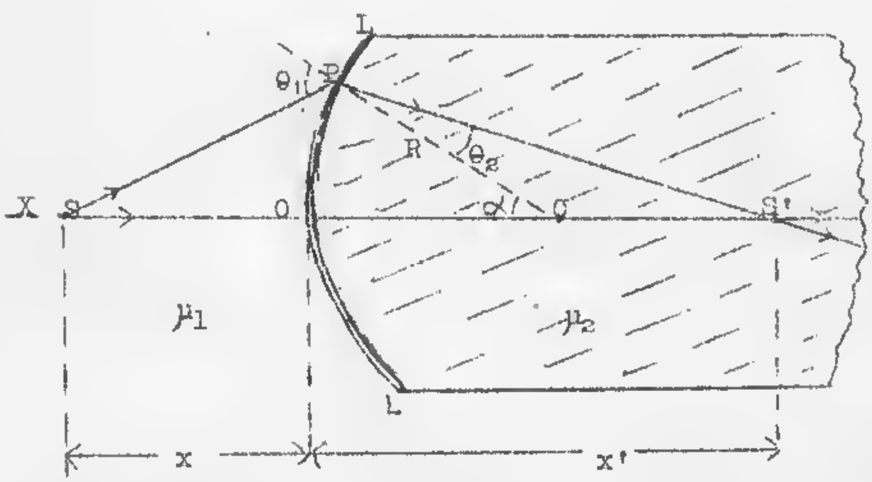


Fig. 34. Refraction at a convex spherical interface; for example, at the spherical end-surface of a thick glass rod situated in a gas. In this diagram, the object is in the medium of smaller refractive index, and *R* is negative in sign.

18. *Formation of Images by a Single Spherical Interface.* Let *LL* in Fig. 34 be the central cross section of a convex spherical surface that separates a medium of

[1] Kepler, *Dioptrice*. For a brief account of Kepler's work on refracting surfaces see *H. Crew, *The Rise of Modern Physics* (Williams & Wilkins, 1935), pp. 100-102.

refractive index μ_1 on the left from one of refractive index μ_2 on the right. Let R be the radius of curvature of the refracting surface, the point C being the center of curvature. Suppose that an object-point S is situated on the axis of symmetry OC, at a distance x from the vertex O. Let SO and SP represent two rays diverging from the object-point, the former along the axis so that it is incident normally on the refracting interface at O, the latter to any point P on the interface such that the arc PO subtends an angle α at the center of curvature C. After refraction, the two rays intersect at some point S' on the axis and at a distance x' from O. Our problem is to determine how the object distance x and image distance x' are related. As in Sec. 14, we shall employ rectangular coordinates, with the vertex O taken as the origin. The axis of the interface in the medium from which the light comes just before striking the interface is taken as the positive X-axis.

If θ_1 and θ_2 are the angles of incidence and refraction of the ray penetrating the interface at P, then, according to the law of refraction,

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2 . \qquad (23)$$

In triangles SPC and CPS', by the law of sines,

$$\frac{\sin \theta_1}{\sin \alpha} = \frac{x - R}{PS} \quad \text{and} \quad \frac{\sin \theta_2}{\sin \alpha} = \frac{x' - R}{PS'} \qquad (24)$$

where the angles θ_1, θ_2, α are all taken as positive quantities. By the sign convention above, x and $\overline{PS}$ are positive, and x', R and $\overline{PS'}$ are negative. Hence $\frac{x - R}{\overline{PS}}$ and $\frac{x' - R}{\overline{PS'}}$ are both positive quantities, since $x - R$ is a positive and $x' - R$ a negative quantity. By eliminating $\sin \theta_1$ and $\sin \theta_2$ from Eqs. (23) and (24), we obtain, after rearrangement,

$$\mu_1 \left(\frac{1}{R} - \frac{1}{x}\right) \cdot \frac{x}{PS} = \mu_2 \left(\frac{1}{R} - \frac{1}{x'}\right) \cdot \frac{x'}{PS'} \qquad (25)$$

The quantities $\overline{PS}$ and $\overline{PS'}$ also can be expressed in terms of x, x', R and α, by means of the cosine law; thus

$$\overline{PS} = \left\{R^2 + (x - R)^2 + 2R(x - R)\cos\alpha\right\}^{1/2}$$
$$= x\left\{1 - \frac{2R^2}{x}\left(\frac{1}{R} - \frac{1}{x}\right)(1 - \cos\alpha)\right\}^{1/2}$$

and

$$\overline{PS'} = \left\{R^2 + (x' - R)^2 + 2R(x' - R)\cos\alpha\right\}^{1/2}$$
$$= x'\left\{1 - \frac{2R^2}{x'}\left(\frac{1}{R} - \frac{1}{x'}\right)(1 - \cos\alpha)\right\}^{1/2}$$

Substituting these expressions for $\overline{PS}$ and $\overline{PS'}$ in Eq. (25) we have finally,

$$\mu_1\left(\frac{1}{R} - \frac{1}{x}\right)\left\{1 - \frac{2R^2}{x}\left(\frac{1}{R} - \frac{1}{x}\right)(1 - \cos\alpha)\right\}^{-1/2}$$
$$= \mu_2\left(\frac{1}{R} - \frac{1}{x'}\right)\left\{1 - \frac{2R^2}{x'}\left(\frac{1}{R} - \frac{1}{x'}\right)(1 - \cos\alpha)\right\}^{-1/2} \qquad (26)$$

This equation is rigorously exact for an object-point located anywhere on the axis of symmetry of any spherical interface.

Example 9. Show that there are two values of the object distance x for which α disappears from Eq. (26); namely, in the trivial case where the object-point is situated at the center of curvature C, and in the case when

$$\frac{1}{x} + \frac{1}{x'} = \frac{1}{R} \qquad (27)$$

Except for the two positions of the object-point found in Example 9, the angle α is always present in Eq. (26). In general, then, the various rays diverging from a given object-point on the axis of a spherical interface are refracted so as to intersect the axial ray, not at a single point S', but in a short line of axial image-points. This is the same phenomenon of spherical aberration as is encountered in the case of reflection (Sec. 13).

Equation (26), unlike the corresponding equation for spherical mirrors, contains the refractive indexes μ_1 and μ_2, a fact that becomes especially significant when we recall that the refractive index of a substance varies with the color of the light (Secs. 4 and 9, Chap. I) and hence that the image distance x' must vary with the color as well as with the angle α. In other words, if the light from a given object-point on the axis is <u>heterochromatic</u> -- that is, composed of more than one, distinct color -- then even those rays that make a given angle α with the axis will fail to meet in a common point after refraction; each heterochromatic ray will be dispersed at the interface into as many different rays as there are distinct colors involved, and these will cross the axial ray <u>SOC</u> (Fig. 34) at as many different points. If the light from the object-point is white, the rays corresponding to any definite value of α will be refracted so as to form with the white axial ray a short line of image-points. This general type of aberration, which is due to the dispersion of light and therefore is not encountered in mirrors, is called <u>chromatism</u>, or <u>chromatic aberration</u>. One obvious although often impractical way to eliminate chromatism is to employ monochromatic light.

19. <u>The First-Order Theory for a Spherical Interface</u>. If the rays allowed to reach a spherical refracting interface from an axial object-point are all paraxial, then $\alpha \cong 0$ and, in view of the Maclaurin theorem, Eq. (26) becomes, after rearrangement,

$$\frac{\mu_2}{x'} - \frac{\mu_1}{x} = \frac{\mu_2 - \mu_1}{R} . \qquad (28)$$

This special but extremely important equation contains the refractive indexes, but not α; hence we can conclude that the image formed by paraxial rays refracted at a spherical interface are subject to chromatism, but not to spherical aberration. These predictions are confirmed by experience; only when the rays from a given axial object-point are both monochromatic and paraxial will they be observed practically to converge to or diverge from a single image-point, real or virtual.

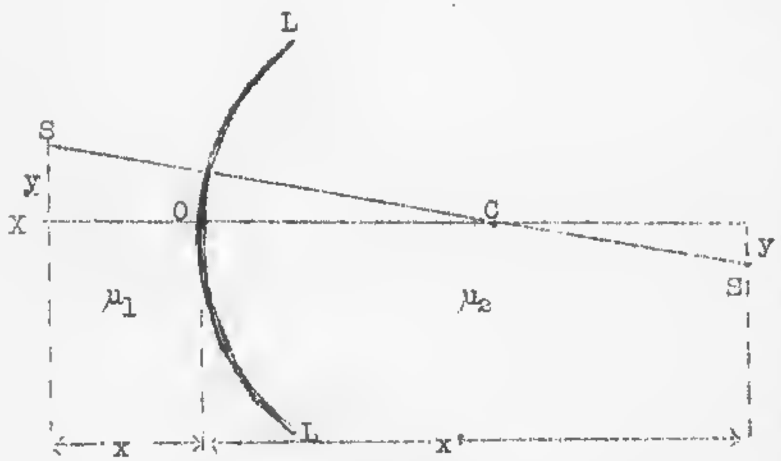


Fig. 35. The lateral magnification is y'/y. Notice that the ray SC passing through the center of curvature is incident normally on the interface and therefore suffers no deviation.

Suppose that the object-point S, instead of being on the axis of symmetry, is located at a small distance y from the axis, so that its rectangular coordinates are (x,y). Then, if y is not too great, the image-point S' of S will be found to be at (x',y') where x' is determined by Eq. (28). An expression for the lateral magnification y'/y can be obtained by writing the equation - $-y'/y = (R - x')/(x - R)$, which follows from the geometry of Fig. 35, and combining it with Eq. (28). The desired result is

$$\frac{y'}{y} = \frac{\mu_1 x'}{\mu_2 x} . \qquad (29)$$

Notice that this expression, unlike Eq. (22) for mirrors, does not contain the negative sign. Thus the image in the present case of refraction will be erect or inverted, accordingly as x and x' have the same or opposite signs.

The theory of image formation which has been treated in this section is known variously as the first-order theory, since it involves setting powers of the angle α higher than the first equal to zero, and as the Gauss theory, after Karl Friedrich Gauss (1777-1855), who was the first to develop it.[1]

---

[1] Gauss, Dioptrische Untersuchungen (1840).

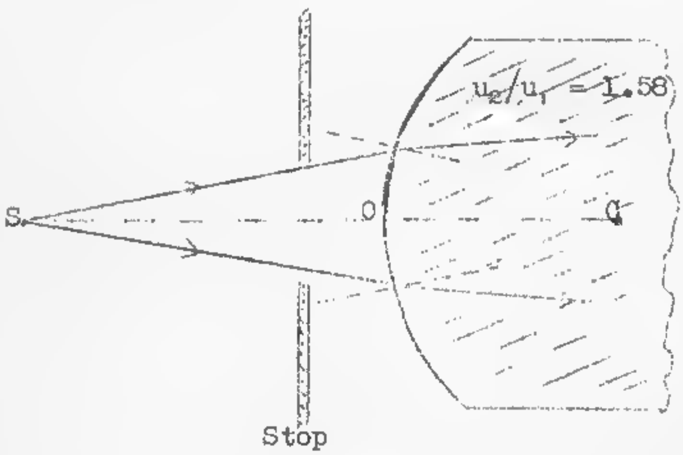


Fig. 36. Example 10, for the case where $\underline{x}$ is 15 cm.

Example 10. One face of a large block of No. 188 light flint glass is ground as shown in Fig. 36 to form a spherical convex surface of radius of curvature 12.0 cm. A small source of yellow D-light is placed in the air in front, and on the axis of symmetry, of the spherical interface. If a stop is arranged in each instance so as to exclude the nonparaxial rays, what is the location, magnification and character of the image formed by the interface when the object distance is (a) 90.0 cm, (b) 41.0 cm, (c) 32.0 cm, (d) 20.7 cm, (e) 15.0 cm?

Solution. Here $\underline{R}$ is -12.0 cm and, from Table II, Chap. I, μ_2/μ_1 is 1.580; therefore Eqs. (28) and (29) become, respectively, $(1.58/x') - (1/x) = -.0483$ and $y'/y = x'/1.58x$.

(a) $(1.58/x') - (1/90.0) = -.0483$, or $x' = -42.5$ cm; $y'/y = -42.5/(1.58 \cdot 90.0) = -.30$. The image in this instance is located in the glass, on the axis and at a distance of 42.5 cm from the vertex. The image is reduced, inverted and real; the refracted rays converge to and pass through the image, after which they diverge as if from an actual source placed at the point.

(b) $x' = -66.2$ cm; $y'/y = -1.02$ The image is slightly enlarged and is inverted and real.

(c) $x' = -92.5$ cm; $y'/y = -1.83$. The image is enlarged, inverted and real.

(d) $x' = -\infty$. The rays after refraction are parallel to the axis.

(e) $x' = 86.1$ cm; $y'/y = 3.63$. The image in this instance is enlarged, erect and virtual. The rays are still divergent after the refraction and hence never actually meet (Fig. 36); but the paths of the refracted rays, produced backward, meet in front of the interface, so this is the region in which they appear to have originated.

20. Images Formed by a Plane Interface. Since Eqs. (28) and (29) do not become indeterminate when $\underline{R}$, the radius of curvature, becomes infinite, they may be used to predict the location and character of the images formed by paraxial rays refracted at a plane interface. The equations become in

this special case,

$$\frac{\mu_2}{x'} - \frac{\mu_1}{x} = 0 \text{ and } \frac{y'}{y} = 1 \tag{30}$$

Thus the sign of x' will always be the same as that of x, and the magnification will always be unity; in other words, an image formed by refraction at a plane surface will always be virtual, erect and the same size as the object. If the object is situated in the medium of smaller refractive index ($\mu_1 < \mu_2$), as when an object in air is viewed by a swimmer under water, the image will be in the air and farther from the surface than the object; but if the object is in the medium of larger refractive index ($\mu_1 > \mu_2$), as when an object under water is observed from the air, the image will be in the water and nearer to the surface than the object. If the image-forming rays are not confined to the paraxial region, the image will be afflicted with other aberrations in addition to chromatism. For example, when a small sphere immersed in water is viewed from a point in the air not directly above the sphere, the image is indistinct and also distorted, the vertical diameter appearing shortened as compared with the horizontal.[1)]

Example 11. The development of parallel sets of equations for the phenomena of reflection and refraction becomes unnecessary if the assumption is made that reflection is a special case of refraction for which $\mu_2 = -\mu_1$. Thus show that, for the special case of reflection, (a) Snell's law of refraction [Eq. (2), Chap. I] reduces to the law of reflection, $\theta_2 = -\theta_1$; (b) Eqs. (29), (28) and (27) reduce, respectively, to Eqs. (22), (20) and (17).

---

1) As has already been mentioned (Sec. 16), the third-order theory developed by Seidel includes consideration of object-points that are distinctly off the axis, and of the rays from them that do not intersect the axis; and, for such points and rays, the four third-order types of aberration known as coma, astigmatism, curvature of the field and distortion are found to occur or to become of importance. When the theory is extended to include powers of α higher than the third, still other, higher-order types of aberration are predicted. Moreover, if the light is heterochromatic, chromatism also is present, this being true even when the first-order theory is applicable (Sec. 18). Chromatism is by far the most troublesome of all the types of aberration, and corrections are made for it in all but the crudest of refracting systems.

21. <u>Systems of Coaxial Spherical Surfaces</u>. Practically every optical instrument consists of a series of refracting mediums separated by spherical surfaces whose centers of curvature lie on a common line called the <u>axis</u> of the system. The paths of selected rays in such a system can be traced through each interface in turn by applying the law of refraction and taking into account the geometry involved. In this way the real or virtual image formed by the first interface is located, this image is then taken as the object for the second interface, and so on until the final image formed by the system as a whole has been determined.

The foregoing method of <u>ray-tracing</u> becomes unnecessary if stops are introduced into the system so as to eliminate all but the paraxial rays or if the aberrations have been otherwise sufficiently corrected; for then Eqs. (28) and (29) hold, and may be applied directly to each interface in turn. If it is found that the image formed by a particular interface lies <u>beyond</u> the next interface, then the object distance for this next interface must be regarded as negative in applying Eq. (28) to it (Fig. 37).

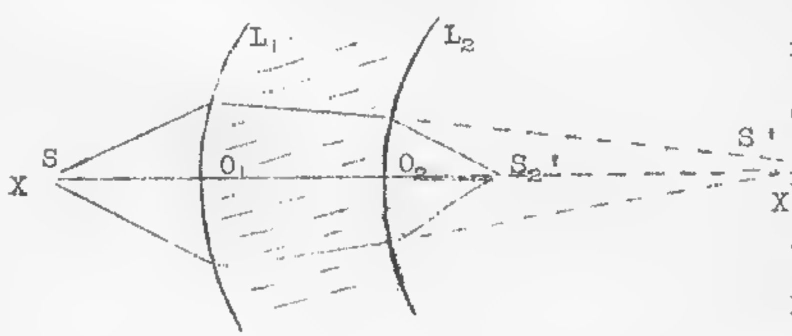


Fig. 37. In a case where the rays emerging from one interface are convergent and also are intercepted by the next interface before they intersect, then the object distance for the latter interface is $\overline{O_2S_1'}$ and hence must be regarded as negative.

A simple <u>lens</u> is a portion of a transparent medium, such as glass, bounded by two coaxial spherical surfaces or by one spherical surface and a plane surface (Fig. 38). Such a system obviously may be treated by the step-by-step procedure already described, no new equations therefore being necessary. However, lenses must be dealt with so frequently that special formulas for them are advantageous; and these can easily be obtained by applying Eqs. (28) and (29) successively to the two faces of a lens in general, thus rendering it unnecessary in subsequent computations to employ this more tedious step-by-step method.

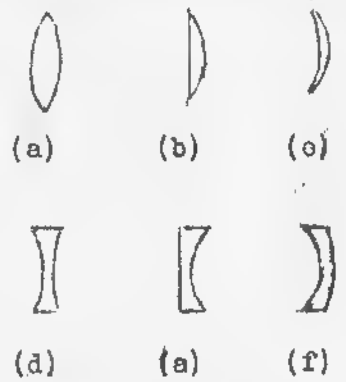


Fig. 38. Common types of lenses.

In Fig. 39, let $\mu'(=\mu_2/\mu_1)$ be the refractive index, relative to the surrounding medium, of the material composing the lens. Let $\underline{R}_1$ and $\underline{R}_2$ be the radii of curvature of the front and back faces of the lens, and let $\underline{t}$ be the distance between the faces at the axis. Suppose that the object-point $\underline{S}$ is at a distance $\underline{x}_1$ from the vertex $\underline{O}_1$ of the front face. The corresponding image-point $\underline{S}_1'$, if the front face were the only refracting surface present, would be at a distance $\underline{x}_1'$ from the vertex $\underline{O}_1$. This distance $\underline{x}_1'$ is given by Eq. (28), which here becomes

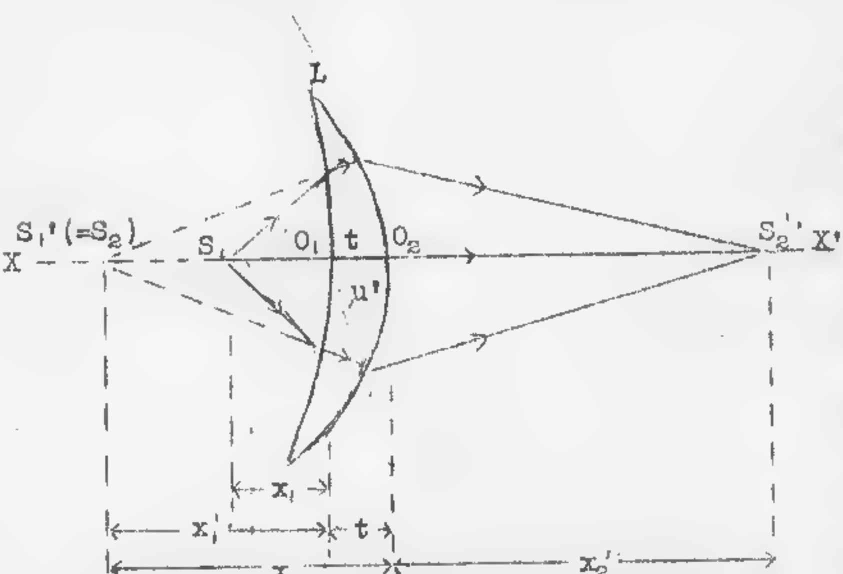


Fig. 39. Derivation of the equation for a simple lens. The subscripts $_1$ and $_2$ refer to two different rectangular coordinate systems having $\underline{O}_1$ and $\underline{O}_2$ as their respective origins.

$$\frac{\mu'}{x_1'} - \frac{1}{x_1} = \frac{\mu' - 1}{R_1}, \qquad (31)$$

since $\mu_2/\mu_1 = \mu'$. If we now regard the image-point S_1' as the object-point S_2 for the back face of the lens, then the object distance x_2, measured from the vertex O_2, is equal to $x_1' + t$. Moreover, the light is now passing from the interior of the lens to the surrounding medium, and the refractive index of the latter, relative to the lens material, is $1/\mu'$. Thus Eq. (28), applied to the back face, becomes

$$\frac{(1/\mu')}{x_2'} - \frac{1}{x_1' + t} = \frac{(1/\mu') - 1}{R_2}$$

or

$$\frac{1}{x_2'} - \frac{\mu'}{x_1' + t} = \frac{1 - \mu'}{R_2} . \qquad (32)$$

Eqs. (31) and (32) suffice to give the relation between image distance and object distance for any simple lens, provided the rays are paraxial.

If the object-point is not on the axis but is a short distance from it, at (x_1, y_1), then the corresponding image-point formed by refraction at the front face alone will be at a distance y_1' from the axis, given by $y_1'/y_1 = (1/\mu')(x_1'/x_1)$, from Eq. (29). Similarly, the final image is at a distance y_2' from the axis given by $y_2'/y_1' = \mu' x_2'/(x_1' + t)$. But the product of the left-hand members of these two equations is seen to be y_2'/y_1, and this is the lateral magnification produced by the lens as a whole. Thus the latter is equal to the product of the lateral magnifications produced at the two faces, or

$$\frac{y_2'}{y_1} = \frac{x_1'}{x_1} \cdot \frac{x_2'}{x_1' + t} . \qquad (33)$$

We see that when the mediums on both sides of the lens have the same refractive index, as in the present case, the refractive index does not appear in the expression for the magnification.

22. Thin Lenses. If the distance t between the faces of a lens is negligibly small in comparison with x_1' in Eqs. (31), (32) and (33), the lens is said to be thin. In this case t can be set equal to zero in Eq. (32), and x_1' eliminated between it and Eq. (31), thus yielding one simple equation that is sufficient for locating the image; namely,

$$\frac{1}{x_2'} - \frac{1}{x_1} = (\mu' - 1)\left(\frac{1}{R_1} - \frac{1}{R_2} \right).$$

Inasmuch as the origins of coordinates O_1 and O_2 coincide in the lens when t is assumed to be zero (Fig. 40), the subscripts $_1$ and $_2$ lose their significance, and the foregoing equation may be written as

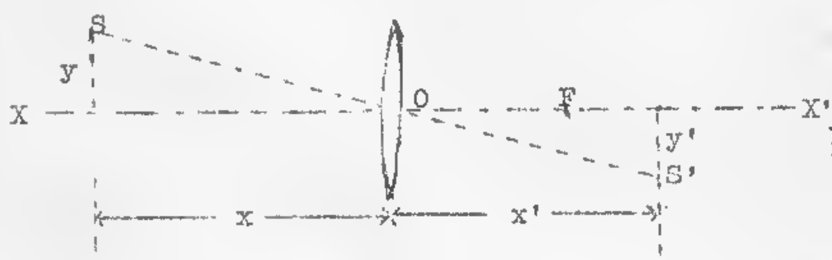


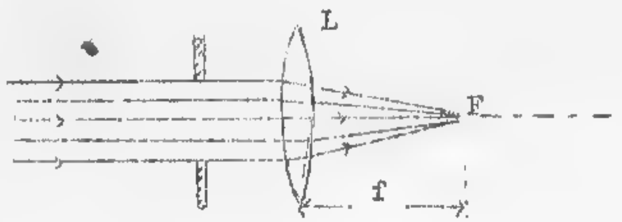
Fig. 40. One type of thin lens.

$$\frac{1}{x'} - \frac{1}{x} = (\mu' - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \quad (34)$$

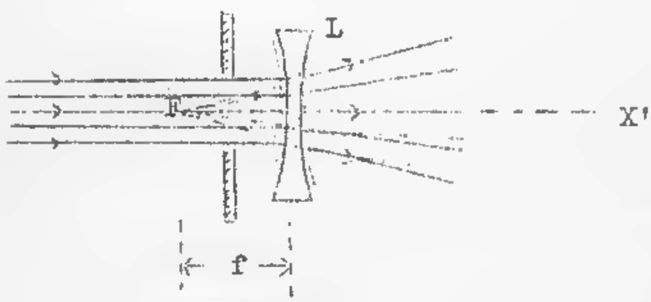
The expression for the lateral magnification obtained by setting t equal to zero in Eq. (33) is

$$\frac{y'}{y} = \frac{x'}{x}. \quad (35)$$

Thus, for a thin lens, the lateral dimensions of the image and object are in the same ratio as the image and object distances (Fig. 40). Eqs. (34) and (35) are applicable only if a lens can be regarded as thin. If the magnitude of the error introduced by neglecting the thickness t is too large, then Eqs. (31), (32) and (33) should be used instead.



(a) A converging lens.



(b) A diverging lens.

Fig. 41. The focal point is at F and the focal length is f.

When the object-point is at infinity on the axis the incident rays are parallel to the axis, and those in the paraxial region converge to, or appear to diverge from, a single image-point F which is called the focal point (Fig. 41). The corresponding image distance OF is the focal length f of the lens, and its reciprocal, $1/f$, is the dioptry D. Setting $1/x$ in Eq. (34) equal to zero, we obtain

$$\frac{1}{f} = D = (\mu' - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \quad (36)$$

The paraxial equation which relates the image and object distances for a thin lens, accordingly can be written in the alternative forms

$$\frac{1}{x'} - \frac{1}{x} = (\mu' - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \frac{1}{f} = D. \quad (37)$$

The importance of the concepts of focal length and dioptry evidently lie in the fact that they are independent of the object and image distances, being constant for a given lens and given surrounding medium.

A particular lens is said to be convergent if it increases the convergence of the rays incident upon it, and this will be the case if the focal point F is situated on the other side of the lens from the incident light and hence is a real image-point. It is said to be divergent if it decreases the convergence of the incident rays, in which case the focal point is on the same side as the incident light, and hence is virtual. As can easily be deduced from Eq. (36), if the surrounding medium is air or any other substance of smaller refractive index than the lens material ($\mu' > 1$), the lens will be convergent or divergent accordingly as it is thicker at the center, as in Fig. 38(a), (b), (c), or thicker at the edges, as in Fig. 38(d), (e), (f). But if the surrounding medium has the larger refractive index ($\mu' < 1$), the opposite will be true; for instance, a double concave lens is convergent when immersed in the medium of larger refractive index.

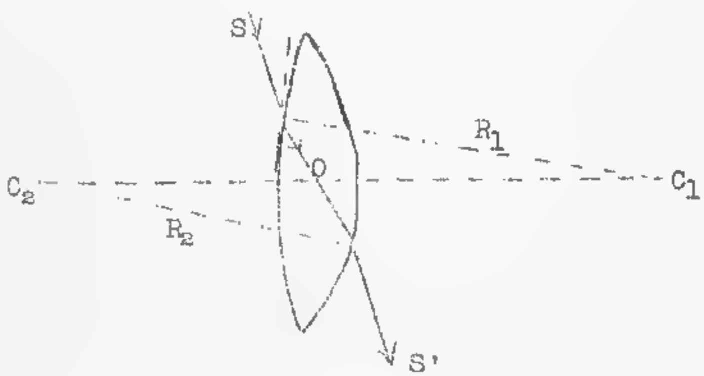
Example 12. The thin double convex lens shown in Fig. 40 is made of glass of refractive index 1.50, relative to air. The radii of curvature of the faces are of magnitudes 0.30 and 0.20 m. (a) What are the focal length and dioptry when the surrounding medium is air? (b) Determine the position and character of the image of a small object placed on the axis and 0.50 m from the lens, it being assumed that a suitably placed stop excludes all but the paraxial rays. (c) Make the foregoing computations for a converging meniscus lens [Fig. 38(c)] made of the same glass and having radii of curvature of the same magnitude.

Solution. (a) In Eq. (36), μ' is 1.50, R_1 is -0.30 m and R_2 is 0.20 m; therefore, f is -0.24 m and D is $-4.2\ m^{-1}$, or -4.2 diopters. The result is the same when the light is considered to be incident on the opposite face, in which case R_1 is -0.20 m and R_2 is 0.30 m.

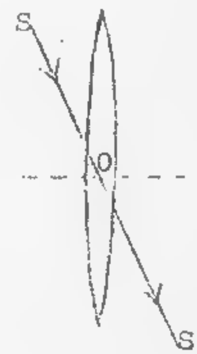
(b) From Eq. (37), $(1/x') - (1/0.50) = -4.2$, or $x' = -0.45$ m; and from Eq. (35), $y'/y = -0.45/0.50 = -0.9$. Thus the image is real, inverted and slightly reduced.

(c) For the converging meniscus lens, μ' is 1.50, R_1 is 0.30 m and R_2 is 0.20 m; therefore, f is -1.2 m and D is -0.83 diopter. When x is 0.50 m, x' is 0.85 m and y'/y is 1.7; thus the image is virtual, erect and enlarged.

A graphical method similar to that employed with mirrors (Sec. 15) enables us to verify the position and nature of the image when the object does not lie entirely on the axis. It is based on the facts that (1) any ray which passes through the optical center O of a lens emerges undeviated and, if the lens is thin, without appreciable lateral displacement (Fig. 42), and (2) any incident ray parallel and close to the axis is refracted by the lens so as to pass through, or appear to diverge from, the focal point F.



(a) A thick lens



(b) A thin lens

Fig. 42. (a) A ray $\overline{SOS'}$ will emerge from a lens undeviated if the planes tangent to the faces at the points of incidence and emergence are parallel (Example 1, Chap. I). Since the radii R_1 and R_2 drawn to these points also will be parallel to each other, $\overline{OC_1}/\overline{OC_2} = R_1/R_2$. Thus the fixed distance C_1OC_2 between the centers of curvature of the faces is divided in the fixed ratio R_1/R_2, and O, called the optical center of the lens, is a fixed point on the axis. (b) If the lens is thin, the aforementioned tangent planes will be very close together and any ray through O will be straight, having no appreciable lateral displacement as well as no deviation.

Summary of the Convention as to Signs

1) The positive direction at a reflecting or refracting surface is taken as the direction from which the light rays come just before striking the surface in question. For example, the positive direction at the first refracting surface is the direction from the surface to the object.

2) The radius of curvature of a surface is considered positive if the center of curvature lies in the positive direction from the surface in question.

3) For thin lenses or for mirrors the focal length is defined as the image distance corresponding to an infinite object distance. Under the above system this means that:

a) The focal length of a concave or converging mirror is positive;
b) The focal length of a convex or diverging mirror is negative;

c) The focal length of a converging lens is negative;
d) The focal length of a diverging lens is positive.

4) Since the dioptric power of a lens or mirror is defined as the reciprocal of the focal length in meters, this means that in this system:

a) A converging mirror has a positive dioptric power;
b) A diverging mirror has a negative dioptric power;
c) A converging lens has a negative dioptric power;
d) A diverging lens has a positive dioptric power.

Please note that contrary to this convention most opticians refer to a converging lens or mirror as having a positive focal length or dioptric power, while a diverging lens or mirror has a negative focal length or dioptric power.

5) In problems involving a single refracting surface, μ_1 refers to the object space, where the object may either be a physical object or the image from a previous optical system.

6) In working optics problems, it is always wise to check geometrically your algebraic solutions.

23. Combinations of Lenses. In treating a system consisting of a combination of two or more lenses a method similar to that outlined in Sec. 21 can be followed. One first determines the image-point of the first lens treating it as if it were acting alone. The image-point of the first lens is then taken as the object-point of the second lens and its image point determined. This procedure can then be carried through step-by-step until the position of the final image-point of the combination has been determined.

In the special case where the lenses are in contact and can be considered thin, a simple relationship holds. For example consider three lenses in contact, of focal lengths respectively f_1, f_2 and f_3. Applying Eq. 37 to each of the three lenses, we find

$$\begin{aligned} \frac{1}{x_1'} - \frac{1}{x_1} &= \frac{1}{f_1} \\ \frac{1}{x_2'} - \frac{1}{x_2} &= \frac{1}{f_2} \qquad (38) \\ \frac{1}{x_3'} - \frac{1}{x_3} &= \frac{1}{f_3} \end{aligned}$$

where the subscripts designate the respective image and object distances of the three lenses. Because the image-point of the first lens is the object-point of the second, etc., we have $x_1' = x_2$ and $x_2' = x_3$. Substituting these values into Eqs. (38) and adding we obtain

$$\frac{1}{x_3'} - \frac{1}{x_1} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}$$

But the focal length F of the combination is given by

$$\frac{1}{x_3'} - \frac{1}{x_1} = \frac{1}{F}$$

Hence

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} \tag{39}$$

24. Achromatic Lenses. In Sec. 11, Chap. I, it was shown that two prisms of different kinds of glass can form an achromatic combination. Such a combination produces deviation for any two chosen colors in a beam of light without producing dispersion. The problem of eliminating chromatic aberration in a lens is obviously analogous to the problem of constructing an achromatic prism. This is accomplished by combining into one lens a convex lens of crown glass and a concave lens of flint glass as shown in Fig. 43. The dispersion produced by one lens is then exactly neutralized by the other lens for any two chosen colors. The condition is reached if the image-points of C and F light coincide. In other words the focal length of the combination should be the same for C light as for F light. Hence from Eq. (39),

Fig. 43. An achromatic lens.

$$\left(\frac{1}{f} + \frac{1}{f'}\right)_C = \left(\frac{1}{f} + \frac{1}{f'}\right)_F \tag{40}$$

where the unprimed quantities refer to the first lens and the primed quantities to the second lens. Substituting from Eq. (37) for $1/f$ and $1/f'$, we obtain

$$\frac{1}{F} = (\mu_C - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + (\mu_C' - 1)\left(\frac{1}{R_1'} - \frac{1}{R_2'}\right) = (\mu_F - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + (\mu_F' - 1)\left(\frac{1}{R_1'} - \frac{1}{R_2'}\right) \quad (40)$$

where μ_C and μ_F refer to the indices of refraction for C and F light respectively. If the radii of curvature of the inner surfaces of the two lenses are the same, then R_2 is equal to R_1', and after rearranging terms we obtain,

$$(\mu_F - \mu_C)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + (\mu_F' - \mu_C')\left(\frac{1}{R_2} - \frac{1}{R_2'}\right) = 0 \quad (42)$$

Eq. (41) then states the condition that the lens shall be achromatic for C and F light, and gives the focal length F of the combination. Achromatic lenses are used in the construction of all high-grade optical instruments.

Example 13: Given that an achromatic combination consisting of a concave lens of No. 188 flint glass and a convex lens of No. 123 crown glass shall have a focal length of -30 cm for C and F light. If the outer surface of the concave lens is a plane compute the radius of curvature of the surface common to the two lenses and of the outer surface of the convex lens.

25. Simple Optical Instruments. In the use of most simple optical instruments such as telescopes, microscopes, simple magnifying glasses, etc., the human eye usually forms part of the optical system. The essential components of a human eye are indicated in Fig. 44. The outer front surface of the eye is covered with a tough, transparent membrane C known

Fig. 44

as the cornea. The cornea together with the lens of the eye L produce a real, inverted image of an object, upon a sensitive membrane R known as the retina. An opaque diaphragm, called the iris, regulates the size of the pupil which is the aperture through which the light passes into the eye. The size of the pupil is varied through muscular action, which action is governed by the intensity of the light which strikes the eye. Muscular action also changes the shape of the lens of the eye and thus its focal length. A person with normal eyes can see objects in sharp focus for distances which vary from infinity to approximately 15 cm from the eye. Usually, however, sustained viewing of objects as close as 15 cm will strain the eye muscles and cause fatigue. The process of focusing the eye on objects of varying distances is known as accommodation; the power of accommodation usually decreasing with increasing age.

The detail of an object is more clearly seen the closer the object is to the eye, for the image on the retina increases with decreasing object distance. However, the power of accomodation places a limit on this distance of approach, and because of the muscular strain which occurs for very close distances, the distance of most distinct vision is usually taken to be about 25 cm. We shall call this distance D, and arbitrarily take it as the minimum object distance for normal vision.

Simple Microscope or Magnifying Lens

As can be seen in Fig. 45 the accommodation of the eye can be effectively increased simply by viewing an object through a convex lens which is placed close to the eye. The eye will then see the virtual image of the object formed by the converging lens, and in general the image will

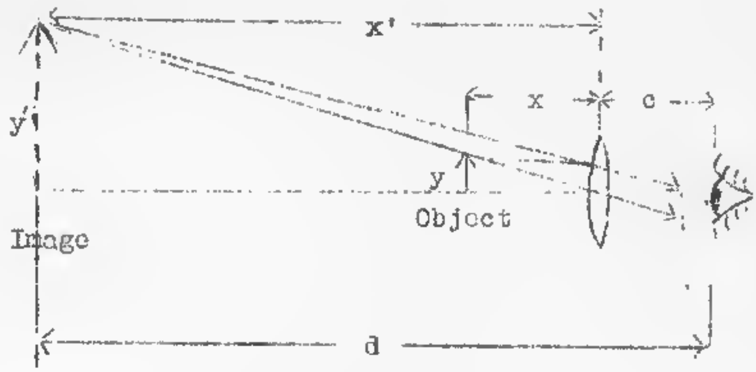


Fig. 45.

appear larger than would the object alone if it were viewed without the aid of the lens. Hence a converging lens may be called a magnifying glass.

The magnifying power M of a magnifying glass is defined as the ratio of the angle subtended at the eye by the image to that subtended by the object alone if it were viewed at the distance of most distinct vision D.

One must be careful to distinguish between magnifying power as here defined and lateral magnification as defined in Sec. 22, which is simply the ratio of the actual linear dimension of an image measured lateraly to the corresponding dimension of the object.

The magnifying power of a lens is then given by

$$M = \frac{\theta}{\theta_o} \qquad (43)$$

where θ and θ_o are the angles subtended at the eye by the image and object, respectively, as defined above. If the lateral dimensions of the object and image are y and y' respectively, then,

$$y' = d \tan \theta \qquad \text{and} \qquad y = D \tan \theta_o$$

where d is the distance from the image to the eye (Fig. 45) and D is the distance of most distinct vision (25 cm). Since in general θ and θ_o are small angles we may take the angles themselves for the tangents and we obtain

$$M = \frac{y'D}{yd} \tag{44}$$

The ratio y'/y may be computed by using Eq. (35) and (37), p. 27 and 28, and we obtain

$$\frac{y'}{y} = \frac{f - x'}{f}$$

where from Fig. 45, x' = d - c; hence,

$$\frac{y'}{y} = \frac{f - d + c}{f} \tag{45}$$

Since the lens is usually held close to the eye, $c \ll d$, and we find,

$$\frac{y'}{y} = \frac{f - d}{f} \tag{46}$$

Substituting into Eq. (44),

$$M = \frac{D(f - d)}{df} \tag{47}$$

Now if the position of the lens is such that the image is formed at the distance of most distinct vision D, then d = D, and

$$M = \frac{f - D}{f} \tag{48}$$

Since the focal length f of a convex lens according to the sign convention is a negative number, if we let $|f|$ represent the absolute value of f, then Eq. (48) becomes

$$M = \frac{|f| + D}{|f|} \tag{49}$$

If the image is formed at infinity then in Eq. (47) d becomes infinite and the equation reduces to

$$M = -\frac{D}{f} = \frac{D}{|f|} \tag{50}$$

The focal length f of a magnifying glass is usually small compared with D, so that the values of M in Eqs. (49) and (50) do not differ much from one another, and the magnifying power of the lens does not vary greatly as the image distance is changed.

Compound Microscope

It is clear from the above discussion that a magnifying glass in order to have high magnifying power must have a very short focal length and must be beld very close to the object and to the eye. These inconveniences and the additional fact that it is difficult to obtain a lens of very short focal length sufficiently free of aberrations place a practical limit of about twenty on the magnifying power of a simple magnifying glass.

For higher magnifying powers an instrument known as a compound microscope is employed. It consists esaentially of two lenses. The first lens, known as the objective, is a convex lens which produces a real and greatly enlarged image of the object. The second lens, known as the eye-piece or ocular, is simply a magnifying glass used for viewing the real image produced by the objective. A compound microscope is shown sebematically in Fig. 46.

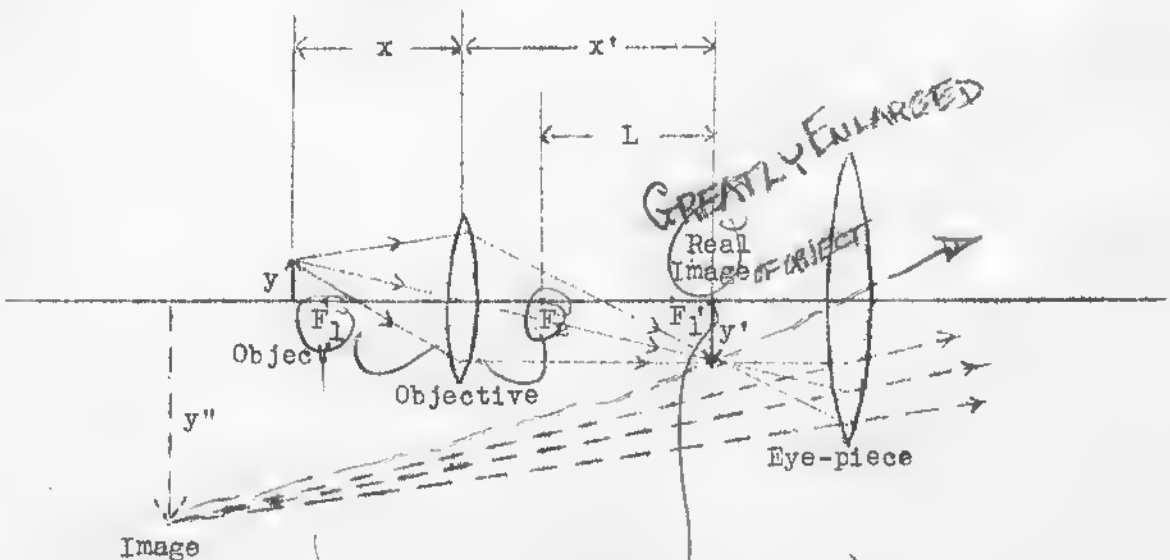
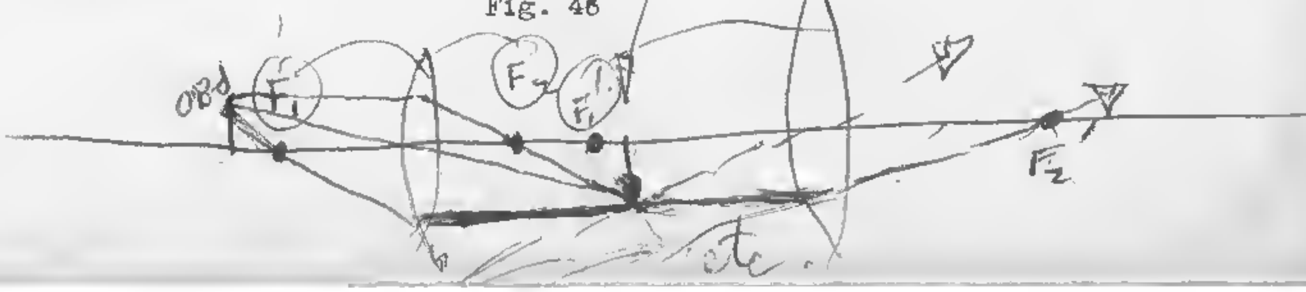


Fig. 46



The magnifying power of a compound microscope is defined as the product of the magnifying powers of the objective and the eye-piece. The magnifying power of the objective, since it produces a real image, is equal simply to its lateral magnification which from Eq. (35) is equal to x'/x. Hence if M is the magnifying power of the compound microscope, and if M' is the magnifying power of the eye-piece, we see that,

$$M = \frac{x'}{x} M' \qquad (51)$$

Let F_1 and F_2 be the two focal points of the objective and F_1' one of the focal points of the eye-piece (Fig. 46). From the relation $\frac{1}{x'} - \frac{1}{x} = \frac{1}{f}$, Eq. (37), p. 28, applied to the objective, it is easily shown that $\frac{x'}{x} = \frac{f_1 - x'}{f1}$. If now we let L represent the distance between F_2 and the position of the real image formed by the objective, and take L as a positive quantity, then $L = -x' + f_1$ and $\frac{x'}{x} = + \frac{L}{f_1}$. Substituting this value for x'/x into Eq. (51), and since from Eq. (50), $M' = - \frac{D}{f_2}$, we obtain,

$$M = - \frac{LD}{f_1 f_2} \qquad (52)$$

as the magnifying power of the compound microscope of Fig. 46. It is the usual practice to construct microscopes of such dimensions that L = 18 cm and since we have previously taken D = 25 cm, we obtain,

$$M = - \frac{450}{f_1 f_2} \qquad (53)$$

The objective of a high-grade microscope consists of a combination of several lenses, in some cases as many as ten, designed so that correction is made for chromatism, spherical and other aberrations. The focal length of the objective for a high-power instrument may be only 2 mm or less. Five is a common figure for the magnifying power of the eye-piece.

Telescopes

An astronomical telescope in its simplest form consists essentially of a converging objective lens of long focal length and an eye-piece for viewing the real image formed by the objective (Fig. 47). If the eye-piece is so adjusted that the image as seen by the eye is at infinity, then the focal point F_1' of the eye-piece will coincide with the real image. The lines drawn in Fig. 47 are not rays of light but are merely construction lines.

The magnifying power M of a telescope is defined as the ratio of the angle θ' subtended by the real image at the eye-piece to the angle θ subtended by the real image at the objective, hence,

$$M = \frac{\theta'}{\theta} . \qquad (54)$$

And since both θ and θ' are in general small, this may be written,

$$M = \frac{l_1}{l_2} \qquad (55)$$

where l_1 and l_2 are the distances from the image to the objective and eye-piece respectively. It is customary to assign a negative sign to the magnifying power of a telescope which produces an inverted image such as that shown in Fig. 47. The negative sign is taken case of by the sign convention for l_1 is a

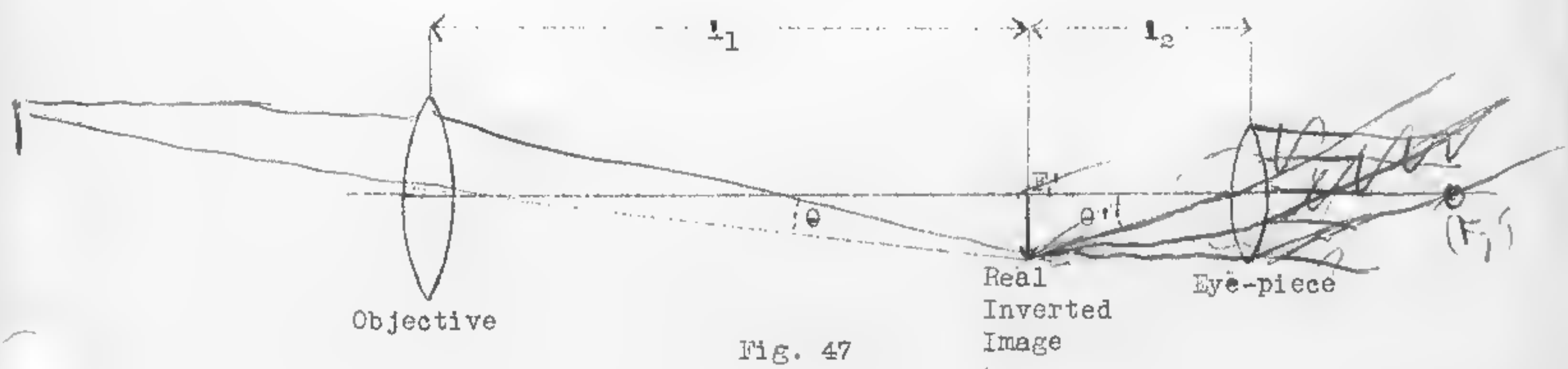


Fig. 47

negative and l_2 a positive quantity. It will be noted that if the object is far away the ratio θ'/θ is practically equal to the ratio of the angular dimension of the image seen by the eye through the telescope to that seen by the unaided eye.

While an inverted image is not a disadvantage in a telescope used for astronomical work, it is decidedly so when terrestrial objects are viewed. Various means have been devised for producing upright images as for example by the insertion of a third lens known as an erecting lens. The erecting lens is placed behind the real image formed by the objective at such a distance as to form a second real image of approximately the same size as the first. The second real image is erect and is viewed in the usual manner by means of the eye-piece. Since the minimum distance between an object and a real image formed by a convex lens is four times its focal length, the use of an erecting lens will increase the length of a telescope by an amount equal to four times the focal length of the erecting lens. This increase in length makes the telescope unwieldy and is therefore not often employed. A second method of producing an erect image is simply to use a diverging instead of a converging lens as the eye-piece. A telescope so constructed is known as a Galilean telescope, and is the usual way in which the compact opera glass is constructed.

Various types of eye-pieces consisting of more than one simple lens are often employed, but we shall not discuss these in detail.

Spectroscope

For precise measurement of the angle of deviation of a ray of light through a prism an instrument known as a spectroscope is employed. The essential features of a spectroscope are represented diagrammatically in

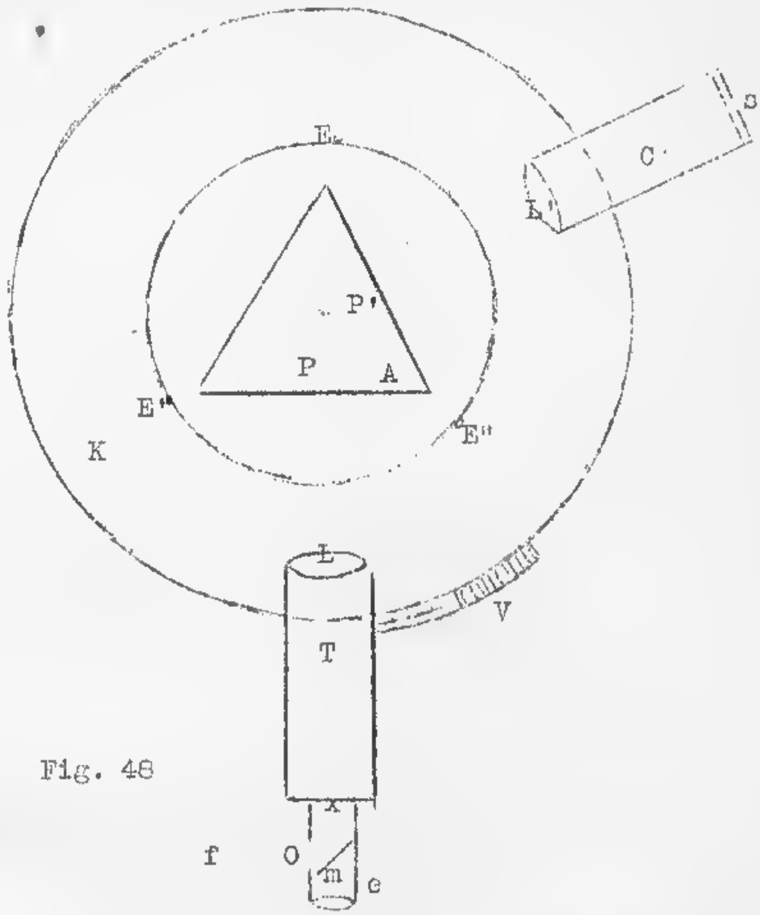


Fig. 48

Fig. 48. A circular table K, the edge of which is graduated in degrees, is supported upon a mounting which carries also a telescope T and a so-called collimator C. The latter consists merely of a tube carrying a slit s so mounted that it may be placed in the principal focal plane of a lens L'. The object of this arrangement is to make it possible to regard s as an infinitely distant source of light, for waves which originate at s become plane waves after passing through the lens L'. The telescope T is mounted so as to rotate about the axis of the table K. The angular position of the telescope with reference to the graduations on the table may be read with the aid of a vernier V attached to the telescope. Attached to the circular table is a

second smaller circular plate called the prism table, which may be leveled by means of the leveling screws E. This table carries the prism P. The telescope may be clamped to the mounting, and the circular table, with the attached prism table, rotated. The rotation with reference to the fixed telescope may then be read with the aid of the vernier V. A small piece of plane transparent glass m is inserted in the eye-piece e so as to make an angle of 45° with the axis of the telescope. The purpose of this arrangement is to make it possible to illuminate the cross hairs at x by throwing a beam of light from a flame or other source f into the eye-piece through the circular opening O, and thence, after reflection from the surface of m, down the axis of the telescope tube. An eye-piece provided with the opening O and the glass plate m is called a Gauss eye-piece.

PROBLEMS

1. A small object is located on the axis and 35 cm from the vertex of a concave spherical mirror of dioptry 10 m^{-1}. Determine the location, lateral magnification and character of the image formed by the paraxial rays.
Ans. $x' = 14$ cm, $y'/y = -0.4$, real image.

2. How far from a concave spherical mirror of radius of curvature 1 m must a small object be placed in order that its image may have a lateral magnification (a) of 3, (b) of -3? What is the character of the image in each case?

3. A plane mirror may be regarded as a spherical mirror of infinite radius. (a) Find the focal length of a plane mirror. (b) What do Eqs. (20) and (22) reduce to in the case of a plane mirror, and what information do these resulting equations give concerning image formation by plane mirrors?

4. A small object is situated on the axis and at a distance of 1.0 ft from a convex spherical mirror of focal length -10 in. How far from the object must a plane mirror be placed in order that the images formed by the two mirrors lie adjacent to each other in the same plane? Ans. 8.7 in.

5. A concave spherical mirror of curvature 2.5 m^{-1} and a plane mirror are placed 1.0 m apart, facing each other. A luminous particle is placed on the common axis of the two mirrors, at a distance of 10 cm from the plane mirror. Compute the diatance behind the plane mirror, and the lateral magnification, of each of the first three images.

6. Two convex spherical mirrors A and B of radii of curvature numerically equal to 12 in. and 7.8 in., respectively, are arranged 9.8 in. apart, facing each other. An object 4 in. long is placed at right angles to the common axis of A and B, and at a distance of 3.9 in. from A. Determine the position, length and character of the image formed by paraxial rays that have been reflected first at B, and then at A.

7. Letting ξ and ξ' denote the object distance and image distance, respectively, measured from the focal point of a spherical mirror, show that the equation for paraxial rays becomes $\xi\xi' = f^2$, which is known as the Newtonian form of the mirror equation for paraxial rays.

8. By the longitudinal magnification of an image is meant the quantity dx'/dx, where dx' is a very short element of length of the image in the direction of the axis and dx is the corresponding element of length of the object. Prove that, if the image forming rays are paraxial, the longitudinal magnification produced by a spherical mirror is equal to $-x'^2/x^2$.

9. A hemispherical, polished metal bowl of radius of curvature 4.0 ft is placed concave side upward on the floor of a room having a 10-ft ceiling. A marble dropped from the ceiling traverses the axis of the bowl. Compute the velocity of the image of the marble formed by the bowl at the moment when the distance of the marble below the ceiling is (a) 4.0 ft, (b) 8.0 ft, (c) 9.0 ft.

10. Substitute $1 - (\alpha^2/2)$ for $\cos\alpha$ in Eq. (17) and employ the resulting third-order equation to compute the magnitude of the longitudinal spherical aberration that exists when an object-point is located at infinity on the axis of a spherical concave mirror of dioptry 10 m^{-1}, the diameter of the face of the mirror being 10 cm.

11. Solve Example 10, Sec. 19, for the case where the object-point is situated in the glass, instead of in the air, the other data being as originally given.

12. Solve Example 10, Sec. 19, for the cases where the refracting interface is concave toward the air, instead of convex, and the object-point is situated in (a) the air, (b) the glass.

13. Solve Example 10, Sec. 19, for the case where the light from the object is a mixture of violet G-light and yellow D-light.

14. A vat filled to a depth of 5.2 ft with water is viewed from directly overhead. What displacement will the bottom of the vat appear to undergo if enough water is removed to reduce the depth to 2.6 ft?

Ans. 0.6 ft downward

15. A small source of light is in the air at such a distance above a vessel containing ethyl alcohol that the image of the source produced by reflection at the surface of the alcohol appears to coincide with the bottom of the vessel. When enough alcohol is added to the vessel to increase the depth by 7.0 cm, it is found that the source must be raised 12 cm in order to bring its image again into coincidence with the bottom of the vessel. Compute the refractive index of the alcohol. Ans. 1.4.

16. Prove that if the object-point is located on the axis and at a distance equal to $R(\mu_1 + \mu_2)/\mu_1$ from the vertex of any spherical refracting interface, (a) the image distance will be equal to $R(\mu_1 + \mu_2)/\mu_2$, (b) the lateral magnification for a small object placed at the object-point in question will be $(\mu_1/\mu_2)^2$.

17. In the case of refracting surfaces, as with mirrors (Sec. 14), the image distance corresponding to an infinite object distance is termed the <u>focal length f</u>. (a) Show that $f = \mu_2 R/(\mu_2 - \mu_1)$ and hence that Eq. (28) may be written in the alternative form, $(\mu_2/x') - (\mu_1/x) = \mu_2/f$. (b) Show that a given refracting surface has two focal lengths, depending upon which side of the interface the object is placed, and that each of these is subject to chromatism. (c) Compute the focal length of the interface in Example 10, Sec. 19, for the conditions described there.

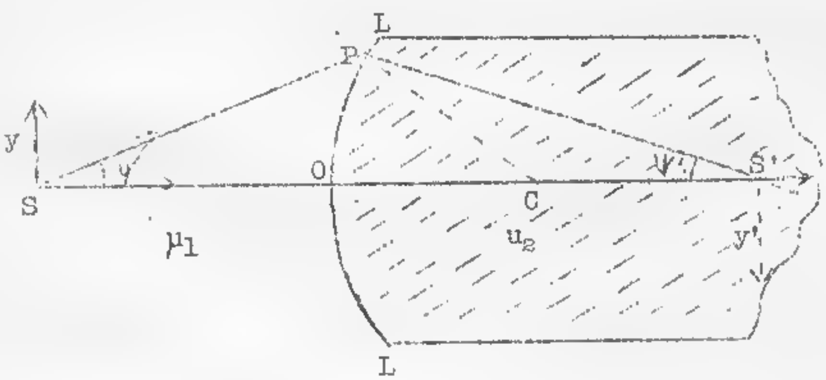


Fig. 49 (a). According to the Abbe sine condition, $\mu_1 y \sin\psi = \mu_2 y' \sin\psi'$ or, for paraxial rays, $\mu_1 y \psi = u_2 y' \psi'$.

18. Considering the general case of an object-point located anywhere on the axis of symmetry of any spherical refracting interface [Fig. 49 (a)], prove with the help of the law of sines and Snel's law of refraction that $\mu_1 \overline{CS} \sin\psi = \mu_2 \overline{CS'} \sin\psi'$ where ψ and ψ' are the angles which the incident and refracted rays make with the axis. (b) Given a second object-point located at a distance <u>y</u> off the axis from the first, show that $y'/y = \overline{CS'}/\overline{CS}$. (c) Hence show that $\mu_1 y \sin\psi = \mu_2 y' \sin\psi'$. Developed independently in 1873 by Ernst Abbe (1840-1905) and Helmholtz, this equation was later derived by Clausius and others from the first law of thermodynamics, and hence ranks as one of the most fundamental principles of optics. (d) If all the rays intercepted by the refracting surface are paraxial, show that the Abbe sine condition then reduces to $\mu_1 y \psi = \mu_2 y' \psi'$, which is known variously as the <u>Lagrange law</u> and the <u>Smith-Helmholtz law</u>. (e) Derive the Lagrange law directly from Eq. (29), the expression for the lateral magnification.

19. Make a table for converging and diverging thin lenses similar to Table IV for mirrors; indicate in it the position, lateral magnification and character of the image when the numerical value of the object distance is infinity, between infinity and 2f, 2f, between 2f and f, f, between f and zero. (b) Point out the similarities that exist between the images formed by concave mirrors and converging lenses, and between those formed by convex mirrors and diverging lenses. (c) Would it be appropriate to designate all concave mirrors as converging, and all convex mirrors as diverging? (d) Show that whenever a real image is formed by a thin lens or a mirror, it is always possible to interchange the object and image.

20. Construct graphs showing the image distance x' and the lateral magnification y'/y plotted as functions of the object distance x for (a) a thin converging lens, (b) a thin diverging lens.

21. Prove that, for any thin converging lens, the distance between the object-point and the corresponding real image-point is never less than four times the focal length of the lens.

22. The quartz composing a certain thin plano-convex lens has a refractive index, relative to air, of 1.61 at 18° C and for ultraviolet light of a certain specified "color". The curvature of the one face of the lens is 0.040 cm^{-1}. Compute the focal length and dioptry in air, for the temperature and "color" specified.

Fig. 50. Problem 23.

23. When a certain thin lens is placed either at L_1 or at L_2 in Fig. 50 a real image of S is formed at the same point S'. (a) Find the focal length and power of the lens. (b) Is the lens convergent or divergent?

24. A thin double convex lens is made of No. 123 crown glass. If the focal length in air is 30.0 cm for red C-light, what will it be for violet G-light?

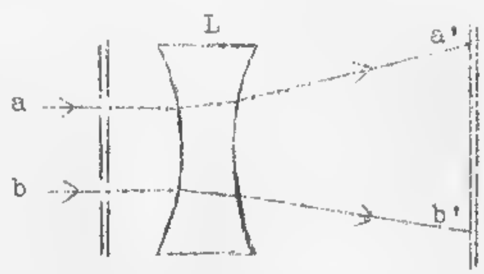


Fig. 51. Problem 25.

25. In Fig. 51, sunlight passes through two openings a and b in a screen placed in front of thin double concave lens L and illuminates a second screen at a' and b'. The distance between the second screen and the lens is 15 cm. The distances ab and a'b' are 5.0 cm and 7.5 cm respectively. Find the focal length and dioptry of the lens.

26. Prove that the longitudinal magnification dx'/dx produced by any thin lens is equal to the square of the lateral magnification.

27. Find the focal length of the double convex lens described in Example 12, Sec. 22, when it is immersed in (a) water, (b) a liquid of refractive index 1.7.

28. A thin double convex lens is made of glass of refractive index 1.5, relative to air. The curvatures of the two faces have the same numerical value, 1/R. If the lens is sealed to the end of a tube containing water, what is the image distance when paraxial rays parallel to the axis are incident on the lens (a) from the air, (b) from the water?

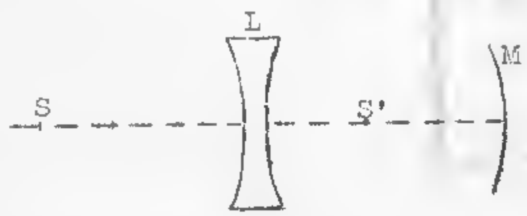


Fig. 52. Problem 29.

29. In Fig. 52, a double concave lens L of unknown dioptry and a concave spherical mirror M of radius of curvature 60 cm are arranged 65 cm apart on a common axis. When a small object S is placed on the axis and 50 cm in front of the lens, a sharp real image is formed at S', 45 cm in front of the mirror. Find (a) the dioptry of the lens, (b) the lateral magnification produced by the system.

30. Find the longitudinal magnification produced by the system described in Problem 29.

31. In Problem 29, suppose that the refractive index of the lens material is 1.5, relative to air, and that the curvatures of the lens faces have the same numerical value C. Where would the image of S be formed if the whole lens-mirror system were under water?

32. Two thin lenses of focal lengths f_1 and f_2 are mounted coaxially and at a distance d apart. (a) Show that the image of an object at infinity on the axis will be located at a distance from the second lens equal to $f_2(f_1 + d)/(f_1 + f_2 + d)$. (b) Hence show that if the distance between the lens is small compared with the focal lengths, the equivalent focal length of the system is $[(1/f\ ) + (1/f_2)]^{-1}$, and the equivalent power is $D_1 + D_2$.

33. Two thin converging lenses, each of focal length of magnitude 12 cm, are mounted on a common axis. (a) If the lenses are in contact with each other, what is the equivalent focal length of the system? Find the position, lateral magnification and character of the image of an object situated 20 cm in front of one of the lenses for the case where the lenses are (b) in contact, (c) 21 cm apart, (d) a distance apart equal to the numerical sum of their focal lengths.

34. Determine the location and character of the image formed when a narrow pencil of parallel rays of yellow D-light is incident normally (a) on a No. 123 crown glass sphere of diameter 12 cm, (b) on a hemisphere of the same material and diameter.

35. An achromatic combination is formed of a converging and a diverging lens of different kinds of glass. If the combination is to be converging which lens must have the greater dispersive power? Why?

36. An achromatic combination consists of a No. 123 crown glass converging lens of focal length -20 cms, and a No. 76 dense flint glass lens. Find the focal length of the flint glass lens required to produce a combination achromatic for C and F light. Find also the focal length of the combination in D light.

37. It is desired to construct a converging achromatic combination (for C and F light) of -50 cm focal length. The common surface between the converging lens of No. 123 crown glass and the diverging lens of No. 76 flint glass, has a radius of curvature equal to 25 cm. Find the radius of curvature of each of the other faces of the two lenses.

38. The focal length of a magnifying glass is -3 cm. An observer whose eye is accomodated for a distance of 25 cm views an object through this glass. How far from the glass should the object be placed? What is the magnifying power of the glass?

39. What is the magnifying power of a ball 3 cm in diameter, made of No. 123 crown glass, if D light is employed?

40. With a certain compound microscope two objectives are supplied, of focal lengths 3 mm and 8 mm, and two eye-pieces of magnifying powers, 5 and 7.5 respectively. What possible magnifying powers are obtainable with this microscope and what are the focal lengths of the eye-pieces?

41. The objective of a certain telescope has a focal length of -30 cm. When the telescope is focussed at infinity its magnifying power is -4.0. What is the focal length of the eye-piece? If the telescope is focused on an object 1.5 meters away what then will be its magnifying power?

42. A luminous object situated on the axis and 40 cm from a certain thin lens in air, is found to produce a sharp, real image at a distance of 25 cm from the lens. (a) Compute the focal length and dioptry of the lens. (b) Should this lens be classified as converging or diverging? If the length of the object in a direction normal to the axis is 3 cm, what will be the position, size and character of the image when the distance of the object from the lens is (c) 30 cm, (d) 20 cm, (e) 15 cm, (f) 5 cm?

Ans. (a) -15 cm -6.5 diopters.

CHAPTER THREE

PHOTOMETRIC CONCEPTS AND MEASUREMENTS

I° Intensity of illumination is directly proportional to the number of candles, or lamps, or emiting points, which illuminate a card or other plane object.

II° It is inversely proportional to the square of the distance of the illuminated plane from the luminous body.

III° It varies as the sine of the angle of incidence.

The three fundamental principles of photometry as stated by J. H. Lambert in 1760. Translated from the Photometria, Part I, Bk. III, Sec. 226, p. 105.

The foundations of scientific photometry were laid as early as the eighteenth century by the French physicist, Pierre Bouguer[1] (1698-1758), and the German physicist, mathematician and astronomer, Johann Heinrich Lambert[2] (1728-1777). Yet only in comparatively recent years have attempts been successful to introduce photometric measurements into

---

[1] Bouguer first described his methods in the Essai d'optique sur la gradation de la lumière (Paris, 1729). His complete work, Traité d'optique sur la gradation de la lumière (Paris, 1760), was published after his death.

[2] J. H. Lambert, Photometria, sive de mensura et gradibus luminis, colorum et umbrae (Augsburg, 1760). An annotated German translation of this treatise will be found in Ostwald's Klassiker der Exakten Wissen-Schaften (Engelmann, Leipzig, 1892), Nos. 31-33. Brief digests of both Bouguer's and Lambert's work appear in *E. Mach, The Principles of Physical Optics, tr. by J. S. Anderson and A. F. A. Young (Dutton, 1926), pp. 13-20 and in *A. Wolf, A History of Science, Technology and Philosophy in the Eighteenth Century (Macmillan, 1939) pp. 167-170. Whereas Bouguer was primarily a good experimenter, but one who restricted himself to observations and drew from these only the more obvious inferences, Lambert gave a complete solution for each problem and, by critically investigating his fundamental postulates, created the concepts of photometry and built them into a mathematical system.

the art of lighting, and thus to place this art on the quantitative basis characteristic of genuine physical science. Today the science of providing and directing light to supply a great variety of human needs has reached a high stage of development, although we still have far to go before architects, decorators and the consumers of light come to take full advantage of this knowledge in conserving human eyesight and life, and in promoting economical lighting. Defective lighting has been conservatively estimated to be the cause of at least 15 percent of industrial accidents, and is one of the causes of defective vision.[1]

The science of lighting is complicated in that it involves not only the physics of light generation, measurement and control, but also the physiology and psychology of vision, and the economics of light production and consumption. Photometry, the part that deals with measurements of light, is essentially a branch of geometrical optics, although additional facts and concepts enter because its concern is with measurements that are all based ultimately on judgements of the visual effects of light. Because some of the important concepts and units used in photometry have been developed comparatively recently, and unsystematically, in connection with specific engineering needs, the treatments of them in much of the literature appeal to the physicist as still lacking to some extent the logical structure and consistent terminology characteristic of most branches of physics.[2]

---

1. A study of more than half a million individuals made about 1934 by the Eyesight Conservation Council of America shows that 20 percent of grade-school children have defective vision, and that the proportion is 40 percent by the time they graduate from college.

2 As far as has seemed to be practicable in a physical treatment, the definitions, terms and symbols employed in this chapter are those given in the "I.E.S. Report of Committee on Nomenclature and Standards", Transactions of the Illuminating Engineering Society 28, 263 (1933).

1. Radiant Flux and Spectral Radiant Flux. Light may be regarded physically as a flow of radiant energy U, this radiant energy being expressible in terms of the joule or any other ordinary energy unit. The quantity dU/dt, which is the time-rate of flow of radiant energy, is termed the radiant flux P. Radiant flux may be expressed in watts, or any other ordinary power unit; that is, the word flux, as it is used here, is equivalent to the term power. In photometry, we shall generally find it simpler to deal in terms of radiant flux, rather than radiant energy, because visual effects do not depend merely on the quantity of radiant energy entering the eye but on its time-rate of entrance.

Most light sources are heterochromatic; that is, the radiant flux emitted by the source consists of various, distinctly different, colors. Moreover, if such a source is operated under constant conditions, each color present is found to be associated with a constant fraction of the total radiant flux emitted by the source. This assertion may be verified by passing the light from the source through a prism, so as to form a spectrum, and determining the relative flux arriving in each region of this spectrum by means of a sensitive thermopile. If the source is the filament of an electric lamp, or any other incandescent substance kept at some temperature exceeding about 800° K, the spectrum is found to be continuous — that is, to exhibit all conceivable colors in a continuous variation from extreme violet to extreme red — and the radiant flux is found to be distributed among these colors in some regular manner (Fig. 1) which remains fixed as long as the temperature of the source is kept constant. In a flux-distribution curve of the type shown in Fig. 1 we shall denote any point (color) on the abscissa by λ, and any corresponding ordinate of the curve by P_λ. The quantity P_λ is called the monochromatic, or spectral, radiant flux.

In Chapter 4 we shall see that λ can be interpreted as a "wave-length in vacuum". An unique value of λ exists for any given color or, more specifically, for any given line of the spectrum (Chap. 1, Sec.10), and this value varies continuously from line to line, from one end of the spectrum to the other. For the visible portion of the spectrum, with which we are concerned in photometry, λ varies from about 4×10^{-5} cm, for extreme violet, to about 7×10^{-5} cm, for extreme red. In Chapter 1 we saw how the discovery of the Fraunhofer lines made it possible thenceforth to refer to certain definite lines of the spectrum, rather than having to speak in terms of vaguely defined colors. Of still greater utility is the designation of spectral lines by their wave-lengths, for a value of λ exists for <u>every</u> point in the spectrum and it varies continuously from point to point.

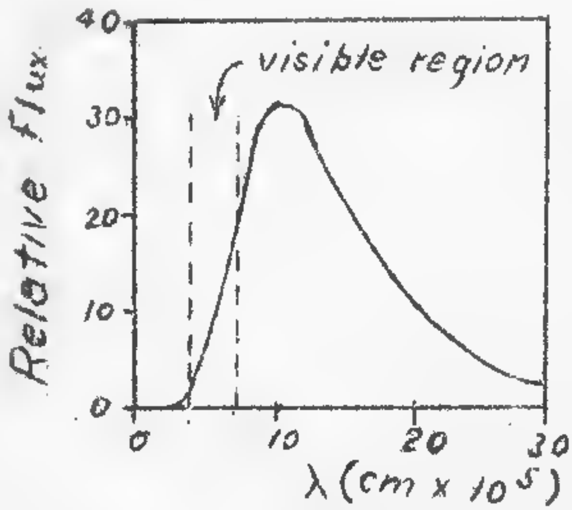


Fig. 1. The distribution of radiant flux in the spectrum of a tungsten-filament lamp operated at 2450°K. In photometry we are concerned only with the flux emitted in the visible region of the spectrum, which is the narrow region included between the two dotted ordinate-lines.

We can now define monochromatic radiant flux P_λ precisely, by saying that the radiant flux which a source emits in the infinitesimal wave-length range from λ to $\lambda + d\lambda$ is given by $P_\lambda d\lambda$. Thus the total radiant flux P emitted by any source is given by the relation

$$P = \int_{\lambda = 0}^{\lambda = \infty} P_\lambda \, d\lambda \; ; \qquad (1)$$

that is, P is proportional to the total area beneath the flux-distribution curve (Fig. 1). The portion of the radiant flux that is in the visible region of the spectrum evidently is given by

$$P = \int_{\lambda = 4 \times 10^{-5} \text{ cm}}^{\lambda = 7 \times 10^{-5} \text{ cm}} P_\lambda \, d\lambda , \qquad (2)$$

this being proportional to the part of the area in Fig. 1 that is beneath the distribution curve and between the two dotted ordinate-lines.

A source whose radiant flux is limited to a very narrow range of wave-lengths $\Delta\lambda$ is said to be monochromatic. It is impossible to have a source that is monochromatic in the literal sense of the term; that is, one which emits radiant flux of a single wave-length λ. But many sources, such as the sodium burner, are practically monochromatic and, for them,

$$P = \int_{\lambda}^{\lambda + \Delta\lambda} P_\lambda \, d\lambda ,$$ where $\Delta\lambda$ is finite in value, but very small.

2. The Luminosity Curve. Experience shows that visual effects — more specifically, sensations of brightness — depend not only on the magnitude P of radiant flux entering the eye but also markedly on the colors present; the eye is not equally sensitive to all wave-lengths, being strikingly selective in its response. For instance, a monochromatic source that is emitting yellow D-light at the rate of 50 watts appears much brighter to the eye than does a monochromatic source that is emitting violet G-light at the rate of 50 watts.

In order to see just how the sensation of brightness varies with color, or wave-length, consider a number of monochromatic sources, each of a different color, but all emitting radiant flux of the same magnitude, say, 50 watts. If these sources are arranged in order of increasing wave-lengths from violet to red, the average observer looking at each in turn will

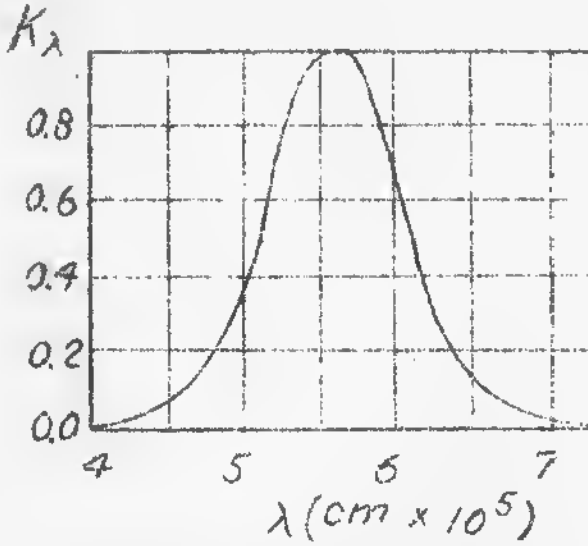


Fig. 2. The international luminosity curve adopted by the International Commission on Illumination [Report of the I.C.I. Congress (1924)].

report that the brightness varies in the manner shown in Fig. 2, where judgments of relative brightness are plotted as a function of wave-length λ. Even skilled observers differ slightly in their judgments of relative brightnesses, but the performance of the "average eye" can be determined by averaging the judgments of a large number of competent observers. Thus the data given in Table I are obtained, and from them the luminosity curve shown in Fig. 2 was plotted. As the data and curve show, the normal eye is most sensitive to light of wave-length 5.55×10^{-5} cm, which is yellowish-green in color. Light of this color has the highest possible visual efficiency, and hence would be the ideal "cold light" if it were not for the fact that a yellowish-green color is unsatisfactory for purposes of general illumination.

The factor K_λ in Table I and Fig. 2 is a pure number which is called the relative luminosity factor for light of wave-length λ. Its value for the yellowish-green color producing the

Table I. Relative luminosity factors K_λ for various wave-lengths λ.*

λ (cm)	K_λ
4.00×10^{-5}	0.0004
4.50	0.038
5.00	0.323
5.50	0.995
5.55	1.000
5.60	0.995
6.00	0.631
6.50	0.107
7.00	0.0041
7.50	0.0001

*Abstracted from Bureau of Standards Scientific Paper No. 475, p. 174.

maximum visual effect is arbitrarily made unity. The reciprocal of K_λ obviously gives the relative magnitude of radiant flux required to produce a given brightness sensation.

3. Luminous Flux. As was shown in Sec. 1, light confined to an infinitesimal range of wave-lengths from λ to $\lambda + d\lambda$ has a radiant flux of magnitude $P_\lambda\, d\lambda$. The brightness sensation which this radiant flux will produce in the normal eye is proportional, therefore, to $K_\lambda P_\lambda\, d\lambda$. If this expression be integrated over all wave-lengths, a quantity is obtained that clearly is a measure of the total visual effect produced by all the light from a source. This quantity which measures the efficacy of radiant flux for producing visual sensation is called the luminous flux, symbol F; its defining equation is

$$F = C_{max} \int_{\lambda = 0}^{\lambda = \infty} K_\lambda P_\lambda \, d\lambda , \qquad (3)$$

where C_{max} is a constant of proportionality the value of which depends on the units employed to express F and P_λ. The limits of integration in Eq. (3) may always be made 0 and ∞, instead of the values of λ corresponding to the ends of the spectrum; this is because K_λ is zero for all values of λ that are not within the visible spectrum, and because P_λ will be zero for any wave-lengths that are entirely missing in the light from a particular source.

Although luminous flux F is the time-rate of flow of radiant energy evaluated according to its capacity to produce visual sensation, the unit in terms of which it is measured is not one of the ordinary power units, as is the case with radiant flux P, but a special photometric unit called the lumen (abbreviation "lu"). Logically, we should expect to define the lumen by means of Eq. (3), the defining equation for luminous flux; but, for reasons of a historical character and of convenience in measurements, the unit actually is defined in terms of a certain standard light-source and another photometric concept which we will discuss in Sec. 5. Therefore, it must suffice to say at this point that 1 watt of monochromatic radiant flux of the yellowish-green color ($\lambda = 5.55 \times 10^{-5}$ cm) which produces the maximum visual effect is found by the most reliable experiments to be equivalent to 621 lumens of luminous flux. This means that, in Eq. (3), F is 621 lu when K_λ is 1 and $\int_0^\infty K_\lambda P_\lambda d\lambda$ is 1 w; hence, that C_{max} is 621 lu·w^{-1}. Eq. (3) may accordingly be rewritten in the form,

$$F = 621 \int_{\lambda = 0}^{\lambda = \infty} K_\lambda P_\lambda \, d\lambda \qquad (4)$$

This equation is applicable to monochromatic light of any color, or to any heterochromatic light for which the flux-distribution curve (Fig. 1) is known, the only restrictions being that F must be expressed in lumens, and $P_\lambda d\lambda$ in watts.

4. Standard Light-Sources. At the time when the comparison of light-sources first became a scientific problem, it was the practice to describe any source as equivalent to some number of ordinary candles. Later, in the interests of accuracy, candles of specified composition, dimensions and rate of consumption were adopted as standards. But candles as standards could be only approximately uniform, at best, and they were eventually replaced by various kinds of flame lamps. Flame lamps are subject to the same limitations as candles, but in lesser degree; hence they served as satisfactory standards until about the time the carbon-filament electric lamp came into use, when they were found to be less constant in their emission of light than the modern sources which were to be compared with

them. This resulted in the adoption of an international standard[1] consisting of groups of carbon-filament lamps of specified construction and operation which are deposited at the various national physical laboratories throughout the world. These groups of lamps constitute the primary standards, and working standards are calibrated by comparison with them. Since electric lamps are not reproducible, which makes them unsatisfactory as a physical standard, a new reproducible standard has since been developed that consists of a "black body" made of fused thoria and immersed while in use in a bath of pure, freezing platinum.[2] A long series of tests carried out at the National Bureau of Standards, Washington, has shown that each time this source is set up anew, it emits light which is the same, as compared with the electric lamp standards, within 1 part in 1000. This is about the limit of accuracy at present required in the most precise photometry. Fortunately, the color of the light emitted by the new standard is practically identical with that of the old standard carbon-filament lamp,[3] which greatly facilitates visual comparison of the two standards.

5. Luminous Intensities of a Point-Source. Every source of light is of course finite in size, and hence emits flux from every point of its surface. In a great many practical situations, however, we deal with light-sources whose dimensions, while finite, are negligibly small in comparison

---

[1] Hyde, *Transactions of the Illuminating Engineering Society* 2, 426 (1907); *Bureau of Standards Bulletin* 3, 65 (1907).

[2] This primary standard was suggested by Waidner and Burgess, *Electrical World* 52, 625 (1908). For an account of the experimental work involved, see Wensel and co-workers, *Bureau of Standards Journal of Research* 6, 1103 (1931).

[3] *Science*, Aug. 1, 1930, p. 109.

with the distances from which we observe them. A source thus observed or used may be regarded as a point-source, a concept that we can often employ to great advantage as a means of simplifying our calculations. Another complication encountered in dealing with light sources arises from the fact that the flux they emit is usually not uniformly distributed in all directions, and this makes it necessary to devise a way to express the flux emitted in any specified direction.

Let S in Fig. 3 be a source of light which is small enough to be considered as a point-source, and let ASA' be an imaginary cone of very small solid angle $\Delta\omega$ with its apex at S. This imaginary cone contains a part ΔF of the luminous flux F emitted by S. A measure of the luminous flux emitted by L in the direction Lx can be gotten by forming the ratio $\Delta F/\Delta\omega$ and taking this ratio in the limit when $\Delta\omega$ becomes vanishingly small. The resulting quantity is called the luminous intensity, I, of the point-source S in the direction specified. Its defining equation evidently is

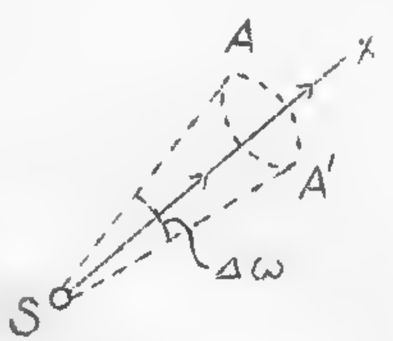


Fig. 3. Definition of luminous intensity.

$$I = \lim_{\Delta\omega \to 0} \frac{\Delta F}{\Delta\omega} = \frac{dF}{d\omega} \qquad (5)$$

In words, the luminous intensity of a point-source, in any specified direction, is equal to the luminous flux emitted per unit solid angle in that direction.[1]

---

[1] Similarly, the radiant intensity, or steradiance, J, of a point-source is defined as the radiant flux P emitted per unit solid angle ω in a specified direction. Its defining equation evidently is $J = dP/d\omega$, and the unit usually employed is the watt per steradian.

The steradian, the most convenient and widely employed unit of solid angle, is defined as the solid angle subtended at the center of a sphere of unit radius by unit area of the spherical surface. Thus the solid angle ω subtended at the center of a sphere of radius r by a portion of the spherical surface of area A is given by the equation, $\omega = A/r^2$ steradians. A complete sphere evidently subtends at its center a solid angle of 4π steradians.

The unit of luminous intensity in most general use is the candle (abbreviation "ca") which was defined originally as the luminous intensity, in a horizontal direction, of a standard candle-source.[1] If a certain light-source is said to have a luminous intensity of 10 ca in a certain direction, this means that its luminous intensity in the direction specified is 10 times that of a standard candle-source in the horizontal direction. Today the magnitude of the unit called the candle is preserved by means of the groups of carbon-filament lamps maintained at the various standardisation laboratories (Sec. 4), a specific fraction of their average luminous intensity in a definite direction being defined as the international candle.[2]

6. The Unit of Luminous Flux. A unit of luminous flux (Sec. 3) can now be obtained simply by writing Eq. (5) in the integral form,

$$F = \int I \, d\omega; \qquad (6)$$

for we already have units for luminous intensity I and solid angle ω, and hence Eq. (6) immediately yields a derived unit of luminous flux, namely, 1 unit of luminous intensity x 1 unit of solid angle. Thus, if I is expressed in candles and ω is expressed in steradians, the unit of luminous flux F is the candle·steradian. It is this unit that is called the lumen. One candle·steradian, or lumen, is evidently equal to the luminous flux emitted in a solid angle of magnitude 1 steradian by a point-source that has a luminous intensity of 1 candle in every direction within that solid angle.

---

[1] Because the candle is so widely used as a unit of luminous intensity, the expression "candlepower" is often used as a substitute for the correct term, luminous intensity. It is a misleading substitute, for luminous intensity is not simply power, but is luminous flux per unit solid angle.

[2] In accordance with an agreement effected in 1909 between the National Bureau of Standards, the Laboratoire Central d'Electricité (France) and the National Physical Laboratory (England), and adopted in 1921 by the International Commission on Illumination.

Suppose that the luminous intensity of a particular point-source is measured for each of a large number of different directions in space, and that the average value I_{ms} is computed. Then, in view of Eq. (6), the total luminous flux emitted by this point-source is given by

$$F = I_{ms} \int_{\omega = 0}^{\omega = 4\pi} d\omega = 4\pi I_{ms} \qquad (7)$$

The quantity I_{ms} is called the <u>mean spherical luminous intensity</u> or, less aptly, the "mean spherical candlepower". For example, measurements show that a photoflash lamp has a mean spherical luminous intensity of 3.6×10^5 candles at the moment of maximum output[1]; hence the luminous flux which it emits at that moment is 4.5×10^6 candle·steradians, or lumens.

Eq. (6) can also be employed to compute the luminous flux emitted within a cone having its apex at the point-source and enclosing a solid angle ω which is smaller than 4π steradians. If the size of the cone is described by specifying the <u>plane</u> angle ϕ made by the surface of the cone with its axis of symmetry, then, as can easily be shown, Eq. (6) may be conveniently expressed in the form

$$F_\phi = \int_{\omega_1 = 0}^{\omega_2 = 2\pi(1 - \cos\phi)} I \, d\omega \qquad (8)$$

Of course, this equation can be used to compute F_ϕ only if the relation of the luminous intensity I of the point-source to the direction ϕ is representable by a mathematical function. Fortunately, this is possible, at least approximately, in the case of several types of sources.

---

1 A wealth of similar data for many different types of light sources and fixtures, together with treatments of all the various practical phases of lighting problems, will be found in H. H. Higbie, <u>Lighting Calculations</u> (Wiley, 1934).

Example 1. Certain sources, such as a Photoflash lamp or a highly rarefied gas rendered luminous by the passage of an electric discharge (Chap.), have approximately the same luminous intensity I_o in all directions. Show that, for them,

$$F_\phi = 2\pi I_o(1 - \cos \phi) \qquad (9)$$

Example 2. In the case of a lighting unit consisting of an incandescent bulb in a white-enameled metal reflector of the shape shown in Fig. 4, the distribution of luminous intensity is described fairly closely by the equation $I_\phi = I_o \cos \phi$, where ϕ is the plane angle between the ray of luminous intensity I_ϕ and the axial ray of maximum luminous intensity I_o. Derive expressions for (a) the luminous flux F_ϕ emitted within a cone, the surface of which makes a plane angle ϕ_1 with the axial ray, and (b) the total luminous flux F emitted by the lighting unit.

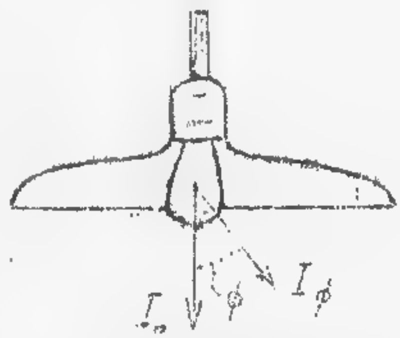


Fig. 4. Incandescent lamp with metal reflector.

7. Illumination. We shall now turn to a consideration of surfaces that receive light from point-sources, for illuminated surfaces are of great importance in determining our ability to see and it is necessary to have a way to calculate the luminous flux incident on them. Suppose that ΔA is the area of a small element of a surface and that ΔF is the luminous flux incident on this element. The ratio $\Delta F/\Delta A$ is termed the average illumination E_{av} on the element, and the limit that this ratio approaches as ΔA is made to approach zero is defined as the illumination E at a point in the surface element ΔA; that is

$$E = \lim_{\Delta A \to 0} \frac{\Delta F}{\Delta A} = \frac{dF}{dA} \qquad (10)$$

Two units of illumination in common use are the lumen per square meter, which is usually called the lux, and the lumen per square foot.

Example 3. Measurements made with an illuminometer (Sec. 10) reveal that the illumination on a certain circular table top of area A and radius R has a maximum value E_o at the center of the table and a minimum value E_R at the edge. Assuming that the illumination decreases uniformly in all directions radially from the center to the edge, derive expressions for (a) the total flux incident on the table top and (b) the average illumination.

Solution. (a) Since the illumination is the same at all points equidistant from the center of the table top, the area may be conveniently divided, by means of concentric circles, into annular zones of infinitesimal width dr. If r is the mean distance of any annulus from the center, the illumination E on the annulus is

$$E_0 - \frac{E_0 - E_R}{R} r,$$

and the area dA of the annulus is $2\pi r dr$. Therefore, the total flux incident on the table top is, in view of Eq. (10),

$$F = \int_{r=0}^{r=R} E dA = \frac{1}{3} A(E_0 + 2 E_R).$$

(b) The average illumination is $\frac{1}{3} (E_0 + 2 E_R)$.

We shall now derive an equation that will enable us to compute the illumination produced on a surface by a single point-source. Any particular element ΔA of the illuminated surface subtends some solid angle $\Delta\omega$ at the source S (Fig. 5). If the luminous intensity of S throughout this solid angle has the average value I_{av}, the flux ΔF incident on ΔA is, by Eq. (6) $I_{av} \Delta\omega$. Hence the average illumination $\Delta F/\Delta A$ on ΔA is $I_{av} \Delta\omega/\Delta A$.

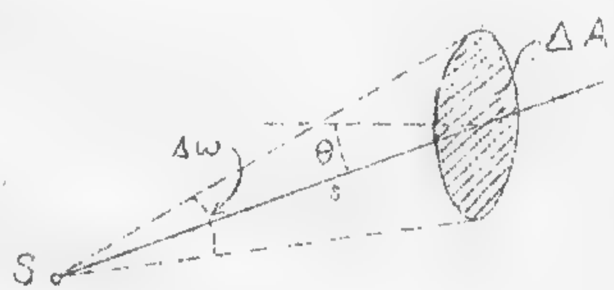


Fig. 5. A surface element ΔA illuminated by a single point-source S.

Suppose that the distance of the element ΔA from the source S is s, and that the element is so oriented that the normal to it makes a plane angle θ with the incident light rays. The solid angle $\Delta\omega$ then has the magnitude $\Delta A \cos\theta/s^2$, and the expression for the average illumination on the element becomes

$$E_{av} = \frac{\Delta F}{\Delta A} = \frac{I_{av} \cos\theta}{s^2} \qquad (11)$$

By taking the limits of the members of this equation, we obtain

$$E = \frac{I \cos \theta}{s^2} \qquad (12)$$

in which I is the luminous intensity of the point-source in the direction of the line connecting it to the point on the surface where the illumination E is produced.

Eq. (12) expresses three of the most frequently employed laws of photometry; namely, that the illumination at a point on a surface

(i) is inversely proportional to the square of the distance between the point and the source producing the illumination; (This is the inverse-square law for illumination),

(ii) is directly proportional to the cosine of the angle between the normal to the surface at the point and the line connecting the point with the source of the illumination; (This is the cosine law for illumination),

(iii) is directly proportional to the luminous intensity of the source in the direction of the point.

The first of these three generalizations — the inverse-square law — was clearly formulated by Kepler[1], although only by an appeal to intuition, and qualitative statements of all three were given by Leonardo da Vinci in his book on painting[2]. They were all advanced by Lambert[3] who treated them as already known, but nevertheless derived them theoretically and tested them experimentally.

---

1 Kepler, Ad Vitellionem Paralipomena (Frankfurt, 1604).

2 Über die Malerei (German ed. by Ludwig, Vienna, 1882), p. 308.

3 Lambert, Photometria, Part I, Chap. 1. Lambert's geometrical demonstration of the cosine law is used in many modern textbooks of physics.

Because Eq. (12) is so frequently employed in photometric measurements and calculations, it is essential to note that its derivation involves the assumptions that the light travels in straight lines from the point-source to the surface illuminated and that none of its energy is converted into other forms during the transmission. Hence the equation and the three laws implicit in it apply accurately only to situations in which: (a) the dimensions of the source are negligibly small compared with the distance between the source and the point illuminated[1]; (b) every part of the medium between the source and the point illuminated has the same refractive index, since otherwise refraction and partial reflection of the light would occur (Chap. 1); (c) the medium does not absorb any appreciable quantity of the luminous energy.

Example 4. A lighting unit of the type described in Example 2 is installed at a height h above the center of a floor. Assuming that h is large compared with the dimensions of the lighting unit, and that the walls and ceiling of the room are covered with a light-absorbing material, develop an expression for the illumination at a point on the floor which is at a horizontal distance r from the center of the room.

Solution: For a lighting unit of this type the luminous intensity I_ϕ of any ray making an angle ϕ with the vertical is $I_0 \cos \phi$. From Eq. (11), $E = I_\phi \cos \phi / s^2 = I_0 \cos^2 \phi / (h/\cos \phi)^2 = I_0 \cos^4 \phi / h^2 = I_0 h^2/(h^2 + r^2)^2$.

Example 5. Show that the lumen per square foot, which is a unit of illumination defined by Eq. (10), may be defined alternatively as the illumination on a surface, all points of which are at a distance of 1 ft from a point-source having a luminous intensity of 1 candle in every direction. Because the unit can be defined in this way, illuminating engineers usually call it the "foot-candle", which is a misleading term since it incorrectly implies that illumination is the product of the luminous intensity of the source and its distance from the surface.

---

[1] If an extended source is concentrated essentially in a plane at right angles to the line connecting its center to the point illuminated, and the length s of this connecting line is 15 times the maximum diameter of the source, the error incurred by employing Eq. (12) is about 0.2 percent. If the length s is 5 times the maximum diameter of the source, the error is about 1 percent, but even this is less than the experimental error involved in illumination measurements unless the photometrist is very experienced and employs equipment of high precision. When a source is entirely too large to be treated as a point-source, the illumination it produces can be computed by imagining the area of the source divided into infinitesimal parts, and then finding the total illumination due to all these parts by the method of the integral calculus.

8. Measurement of Luminous Intensity. We have deferred consideration of how to measure the luminous intensities of a point-source (Sec. 5) until after our discussion of illumination, for this measurement, and most photometric measurements, ultimately depends upon a determination of the illumination at a surface.

The first effective photometer was constructed by Bouguer.[1] A few years later Lambert[2] used an instrument that was very similar in principle to the type now known as the Rumford shadow photometer. Today, the method widely employed consists in illuminating two white, diffusely reflecting screens, the one with light from the source to be measured, the other with light from a sub-standard source of known luminous intensity, and comparing the brightness of these reflecting surfaces by means of the eye. The eye cannot be used to estimate degree of brightness accurately, but it can be used to judge equality of brightnesses, provided that the photometer is so arranged that the reflecting screens are presented to the eye side by side with the finest possible line of demarcation. A second essential provision is that the color-quantity of the light from the two sources be nearly the same, for otherwise the impression of color contrast is so strong that the observer will be unable to duplicate closely his judgment of equality of brightness, and different observers will make widely different judgments. Any method that involves judgment of brightness is, therefore, practically limited to the comparison of light-sources of the same type, though of different power; for example, any two evacuated tungsten-filament lamps or any two carbon-filament lamps.

---

1 Essai d'optique sur la gradation de la lumière (Paris, 1729).

2 Photometria, Part I, Chap. 2.

It is also important for the accurate comparison of two sources that only the light directly emitted from each source shall reach its respective screen, and that a method be provided for varying the illumination of the screens according to some known law. The simplest way to vary the illumination is to move each source toward or away from its respective screen along a line normal to the screen at its center, always keeping the distance between screen and source large enough for the latter to be treated as a point source. Then the law of variation of illumination is the inverse-square law (Sec. 7). Suppose that the sources have luminous intensities of magnitudes $\underline{I}_1$ and $\underline{I}_2$, respectively, and that their respective distances from the screens have to be made $\underline{s}_1$ and $\underline{s}_2$ in order for the screens to appear equally bright. Then, by Eq. (12)

$$\underline{I}_1/s_1^2 = \underline{I}_2/s_2^2, \qquad (13)$$

provided it can be assumed that equality of brightness of the two reflecting screens implies equality of illumination $\underline{E}$. Since this assumption is never safe, the following <u>substitution method</u> is employed in all accurate photometry. A third "comparison" source of constant luminous intensity is placed at a fixed distance from one of the reflecting screens. The two sources of luminous intensities $\underline{I}_1$ and $\underline{I}_2$ are then placed one at a time in front of the other screen and their respective distances $\underline{s}_1$ and $\underline{s}_2$ from this screen are in each case adjusted until the brightness appears to match that produced on the first screen by the comparison lamp. The illuminations produced on the second screen must then have the same value $\underline{E}$, and hence Eq. (13) is applicable. If either $\underline{I}_1$ or $\underline{I}_2$ is the luminous intensity of the sub-standard source, the luminous intensity of the other lamp can be computed at once.

9. The Lummer-Brodhun Photometer Head. In carrying out the visual, direct comparison method described in the preceding section, the instrument almost universally used is the Lummer-Brodhun photometer head[1], shown diagramatically in Fig. 6. The two reflecting screens are simply the opposite sides of a single, matt white screen ss. They are viewed through an eyepiece at e with the aid of two plane mirrors, or two total reflection prisms, placed at M and M'. In order to bring the two sides of the screen ss into immediate juxtaposition, as seen by the eye at e, the phenomenon of total reflection (Chap. I) is made use of in the design of the cube PP'. This cube consists of two right-angled prisms, P and P', with adjacent faces that are made to come into as perfect contact as possible in certain places, but not to come into contact in other places; this is accomplished by having definite parts of the contacting face of one prism either etched, sand-blasted or curved, so that when the two prisms are firmly pressed together, only certain portions make complete contact with the face of the other prism. The light coming to the eye at

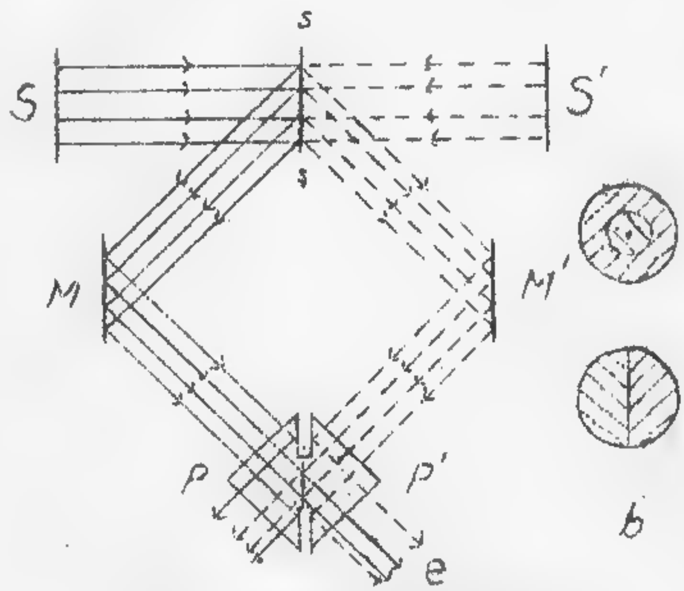


Fig. 6. The Lummer-Brodhun equality-of-brightness photometer head. Two different types of field are shown at (b).

[1] O. Lummer and E. Brodhun, Zeitschrift für Instrumentenkunde 9, 23 (1889). Although this instrument was invented by William Swan in 1859 (Transactions of the Royal Society of Edinburgh, Vol. 21), accurate photometry was not needed in his time and the instrument was not brought into use until Lummer and Brodhun invented it independently in about 1888. The instrument employs the same principle, but in a much more refined form, as the "grease-spot" photometer, designed in 1844 by Robert Wilhelm vonBunsen (1811-1899) for use in determining the luminous intensities of electric arcs.

e through the portions which make complete contact is light that comes from the source S, undergoes reflection at M and then is transmitted through the cube PP' just as if there were no interface in it. On the other hand, the light coming to the eye at e from the portions of the interface where an air film exists, is composed entirely of rays that have come from the source S' by way of the reflector M', and then have undergone total reflection at the surface of the air film in PP'. Hence, if the two sides of the screen ss are exactly similar surfaces and M and M' are exactly similar reflectors, it is only necessary to set the screen ss at such a point between the sources S and S' that the whole surface PP' [Fig. 6(b)] appears uniformly bright, with the line of demarcation between its portions made as nearly invisible as possible, and then to apply Eq. (13). However, since the instrument cannot be perfectly symmetrical, it is best to employ the substitution method (Sec. 8), or else to rotate the photometer head through 180° about an axis passing through the screen ss, thus interchanging the two sides of the screens and also the reflectors, and take the mean of the setting before and after reversal as the correct setting.

10. Measurement of Illumination. Problems of illuminating engineering more frequently involve the measurement of illumination (Sec. 7) than any other photometric quantity, and this has led to the development of various types of portable illumination photometers, or illuminometers.[1] Although the degree of precision required in these illuminometers is not high, the more accurate of them always involve three parts: a diffusely reflecting test screen which can be placed at the spot where the illumination is to be measured; a similar screen installed internally in the

---

[1] For descriptions of the various types, see J. W. T. Walsh, Photometry (1926), Chap. 12. This excellent and comprehensive treatise deals with all phases of photometric measurements and instruments, and includes an extensive bibliography.

instrument and illuminated by a battery lamp in such a way that its brightness can be varied; and some device, such as a small Lummer-Brodhun cube (Sec. 9), for judging when the brightness of the internal screen has been adjusted to match that of the external test screen. The scale of an illuminometer is usually calibrated to give readings directly in terms of the lumen per square foot (foot-candle) or the lux (meter-candle). This calibration, which should be checked at frequent intervals can be carried out by producing calculable amounts of illumination on the test screen with the aid of a sub-standard source of known luminous intensity.

Recently it has been found possible to develop various extremely compact forms of nonvisual or physical illuminometers; that is, illuminometers that do not involve visual comparisons of brightnesses. They consist essentially of either a selenium cell or a copper-oxide photovoltaic cell (Chap.) connected to a sensitive galvanometer, the dial of which is calibrated to give direct readings of illumination. Although the light-sensitive materials used in these physical illuminometers have wave-length versus response curves that differ somewhat from the luminosity curve for the eye (Fig. 2), it is possible to make the former curve agree with the latter fairly well by having the light pass into the instrument through a suitably chosen filter.

11. Brightness of an Extended Source. Thus far we have been concerned with point-sources whereas modern lighting problems often involve sources of dimensions so large compared with the distances from which they are measured that neither the concept of luminous intensity (Sec. 5) nor the law of inverse squares (Sec. 7) is directly applicable. The use of sources of larger area, although of lower brightness, is the tendency in modern lighting; for example, incandescent lamps are now often enclosed in diffusing glass globes of large surface area, the filaments of the modern lamps

being so bright that direct view of them would be unbearable. Vision may be impaired if the brightness of a surface, or the contrast between its brightness and that of the background, is excessive. Now, to be able to judge the equality of brightness of two adjacent surfaces of the same color is not difficult, as we have already noted in Sec. 8; but if we are to be able to express a particular brightness quantitatively, we must define the concept in terms of physical operations and determine its relation to other physical quantities.

Imagine the surface of an extended source divided into elements, each small enough in area to be treated as a point-source. Let $\Delta\sigma$ in Fig. 7 be the area of such an element, let θ be the plane angle between the normal to it and the direction of the observer, and let ΔI be the luminous intensity of the element in the direction θ. The <u>brightness</u> B at a point in the element $\Delta\sigma$ and in the direction θ is defined by the equation

$$B = \lim_{\Delta\sigma \to 0} \frac{I}{\Delta\sigma \cos\theta} = \frac{dI}{d\sigma \cos\theta} \qquad (14)$$

Evidently, $\Delta\sigma \cos\theta$ is the area $\Delta\sigma$ projected on a plane perpendicular to the direction θ. Hence, in words, the brightness in a particular direction at a point in any surface element is the ratio of the luminous intensity in the given direction and the area of the element projected on a plane perpendicular to the given direction. Brightness at a point is expressed in <u>candles per unit area</u> of emitting surface perpendicular to the direction of view; that is to say, in lumens per steradian per unit area of normal emitting surface.

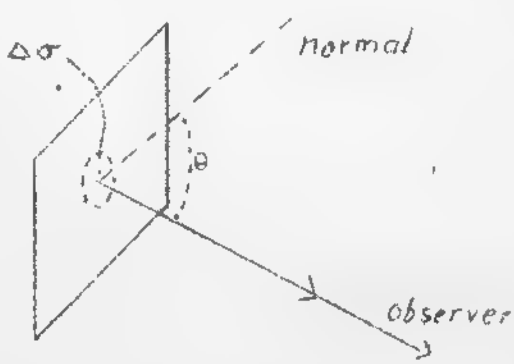


Fig. 7. Definition of brightness.

12. Highly Diffusing Sources. The brightness of a source may vary from point to point of its surface and, at any point, may vary with the direction from which the point is observed. However, many light sources are in use for which the brightness at any given point on the surface is practically independent of the angle of view. This is true, for example, of a deep, narrow cavity in a piece of incandescent metal, or of opal or alabaster glass from which transmitted light is emerging. Any such surface is said to be highly diffusing and, in the ideal case where the brightness at each point is entirely independent of the angle of view, is termed perfectly diffusing. Now, as an inspection of Eq. (14) will show, if the brightness B at any point of a surface does not change with the angle of view θ, this must be because the luminous intensity dI at any point varies directly as $\cos \theta$; that is, the luminous intensity in any direction θ is equal to the luminous intensity in the direction of the normal to the surface element multiplied by $\cos \theta$. This statement, that the luminous intensity is proportional to the cosine of the angle of view, is known as the Lambert cosine law. The law applies only approximately to most actual surfaces and hence it is always best to specify angles of observation in giving data on brightness.

However, an advantage results when the brightness is nearly enough the same for every angle of observation to make it possible to regard the surface as perfectly diffusing, for then a simple relation is found to exist between the brightness at any point and the luminous flux per unit area, $dF/d\sigma$ emitted in all directions from that point. To find this relation consider that the luminous flux dF (emitted by a surface element $d\sigma$ (Fig. 7)) through a solid angle $d\omega$ in the direction θ is, in view of Eq. (5), $Id\omega$; and this, per unit area of the emitting surface, is

$$\frac{d}{d\sigma}\,dF = \frac{dI}{d\sigma}\;d\omega$$

Introducing the brightness into this expression by means of Eq. (14), we have

$$\frac{d}{d\sigma}\,dF = B\cos\theta\;d\omega \qquad (15)$$

Now $d(dF/d\sigma)$ expresses the luminous flux per unit area emitted through an infinitesimal solid angle in the direction θ, whereas we wish our final result to be an expression for $dF/d\sigma$, the luminous flux per unit area emitted in all directions from $d\sigma$; that is, in the whole solid angle 2π steradians through which $d\sigma$ is emitting light. An easy way to approach this integration of Eq. (15) is to imagine a hemispherical surface of radius s described about $d\sigma$ as a center (Fig. 8), since all the light emitted by $d\sigma$ must pass through this hemispherical surface. The light is emitted symmetrically about the normal to $d\sigma$ and hence we may choose as the element of area of the hemisphere an infinitesimal ring, all portions of which make the same plane angle θ with the normal. If the width of the ring subtends a plane angle dθ then this width must be $s\,d\theta$. The length of the ring is $2\pi \cdot s \sin\theta$. Therefore, the area of the ring is $2\pi s^2 \sin\theta\, d\theta$. This expression divided by s^2 gives the solid angle $d\omega$ that the ring subtends at $d\sigma$. Hence a substitution for $d\omega$ may be made in Eq. (15), yielding

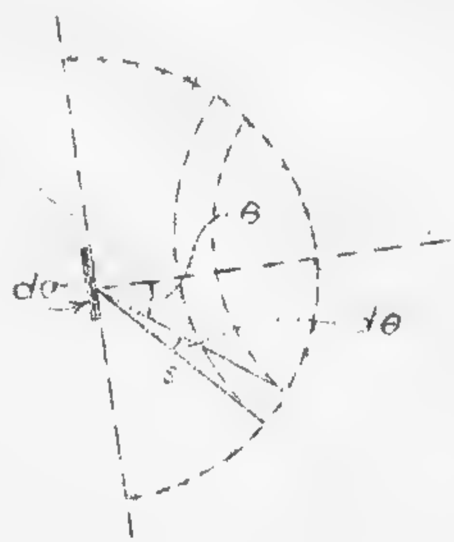


Fig. 8. Method of computing the flux-brightness at a point.

$$d\left(\frac{dF}{dA}\right) = 2\pi\, B \cos\theta \sin\theta\, d\theta .$$

If this equation be integrated from $\theta = 0$ to $\theta = \pi/2$, thus covering the entire surface of the hemisphere, we obtain for the total luminous flux per unit area of $d\sigma$,

$$\frac{dF}{d\sigma} = 2\pi B \int_{\theta=0}^{\theta=\pi/2} \sin\theta \cos\theta\, d\theta = 2\pi B \left[\frac{\sin^2\theta}{2}\right]_0^{\pi/2}$$

or

$$\frac{dF}{d\sigma} = \pi B. \tag{16}$$

The quantity $dF/d\sigma$ is termed the flux-brightness at a point on a perfectly diffusing surface, and we have found it to be equal to π times the brightness at the point. If B is expressed in, say, candles per square centimeter, then $dF/d\sigma$, or πB, will be expressed in candle steradians per square centimeter; that is, in lumens per square centimeter. This unit of flux-brightness, $1\ \text{lu} \cdot \text{cm}^{-2}$, is in practice called the lambert. When a highly diffusing surface has the same brightness at every point, as is usually the case, then Eq. (16) obviously becomes, by integrating, $F = \pi B\sigma$, where σ is the total area of the surface.

Example 6. When a 100-w gas-filled tungsten-filament lamp is operated under the conditions for which it was designed, its light-source efficiency (defined as the ratio of the luminous flux from a self-luminous source to the power required to maintain it) is found to be $12.5\ \text{lu} \cdot \text{w}^{-1}$. A certain globe of diameter 30 cm made of diffusing glass, surrounds such a lamp. The surface of this globe is found by direct measurements to have a flux-brightness that is fairly uniform from point to point and for various angles of view, the average value being .35 lamberts. Show by computation that (a) the luminous flux transmitted by the globe is 9.9×10^2 lu, (b) the percentage of luminous glux transmitted by the globe is 79 percent, and (c) the average brightness of the surface of the globe is $0.11\ \text{ca} \cdot \text{cm}^{-2}$.

As Example 6 illustrates, a transparent material from which light is emerging in all directions more or less in accordance with the Lambert cosine law may be treated as if it itself were the source of light. The flux-brightness of such a secondary sourse evidently is equal to τE, where $\underline{E}$ is the illumination, or flux per unit area, incident on the rear surface of the material from the primary source, while τ is the fraction of this incident luminous flux that is transmitted by the material so as to emerge diffusely from the front surface. If $\underline{E}$ is expressed in lumens per square centimeter, the flux-brightness $\tau \underline{E}$ will be expressed in lamberts.

A surface that <u>reflects</u> light diffusely may also be regarded as a secondary source. No reflecting surface is perfectly diffusing although a material such as white blotting paper provides a fair approximation. The flux-brightness of a diffusely reflecting surface evidently is equal to $\rho \underline{E}$, where $\underline{E}$ is the illumination on the surface and ρ is the fraction of the incident luminous flux that is reflected diffusely.

The fractions τ and ρ are called, respectively, the <u>transmission factor</u> and the <u>reflection factor</u>. A third fraction α, called the <u>absorption factor</u>, is defined as the fraction of the incident luminous flux that is absorbed by a material. It is found that the values of ρ, α and τ depend not only on the character of the surface but generally also on the angle of incidence and wave-length of the light. For any given material, $\rho + \alpha + \tau = 1$. Since either ρ or τ for any material is always less than unity, the flux-brightness of any diffusely reflecting or diffusely transmitting surface is always less than the illumination that produces this flux-brightness. This latter statement should serve to emphasize the essential distinction between the concepts of flux-brightness and illumination; the former expresses the luminous flux per unit area emerging at a point of a surface, and the latter expresses the luminous flux per unit area received at a point of a surface from external sources.

If a surface is not highly diffusing, its flux-brightness is given by the expression $\int \underline{B}_\theta \cos\theta \, d\omega$, where $\underline{B}_\theta$ is now a function of the angle of observation θ. The flux-brightness is no longer simply $\pi \underline{B}$, nor for a transmitting or a reflecting surface is it simply $\tau \underline{E}$ or $\rho \underline{E}$, and consequently the concept loses most of its utility. However, in practical work with poorly diffusing surfaces, it is customary to make use of the idea of apparent flux-brightness, a poorly diffusing surface which has a brightness $\underline{B}_\theta$ when observed from a particular angle θ being said to have an apparent flux-brightness $\pi \underline{B}_\theta$ for the angle of view θ. Its actual flux-brightness is of course smaller or larger than $\pi \underline{B}_\theta$. More specifically, the statement that the apparent flux-brightness at a point of the surface is 1 lambert for a particular direction means that if the surface were perfectly diffusing, it would emit 1 $lu \cdot cm^{-2}$ at this point.

13. Measurement of Brightness and Flux-Brightness. Any illuminometer that is equipped with a detachable external test screen (Sec. 10) can be recalibrated so as to give direct readings of brightness rather than of illumination. The illuminometer, as originally calibrated, gives direct readings of the illumination $\underline{E}_t$ of the external test screen furnished with the instrument, and the corresponding flux-brightnesses can be computed by multiplying these readings $\underline{E}_t$ by ρ_t, the reflection factor for the test screen (Sec. 12). Hence, to measure the flux-brightness of any other diffusing surface, it is only necessary to detach the test screen from the instrument, sight the latter on the surface in question, take the reading $\underline{E}$, and then compute the flux-brightness $\rho_t \underline{E}$. If the brightness is desired it can be computed by means of the relation $\underline{B} = \rho_t \underline{E}/\pi$.

The reflection factor ρ_t may be determined as follows. A point-source of known, high luminous intensity is placed at a measured distance from the illuminometer test screen and, considering this screen as a secondary source, its luminous intensity $\underline{I}_t$ is determined in the usual manner (Sec. 8). At the same time the illumination $\underline{E}_t$ of the test screen is measured with the illuminometer. Then ρ_t can be computed by means of the relation, $\rho_t = \pi \underline{I}_t / \underline{E}_t \underline{A}_t$, where $\underline{A}_t$ is the area of the test screen.

14. Measurement of Total Luminous Flux. The earlier practice in photometry was to rate any primary source in terms of its luminous intensity in some specified direction, the measurement being carried out by the

method described in Sec. 8; but today, ratings in terms of the total luminous flux emitted are of more significance, because of the practice of using modern primary sources in connection with some sort of diffusing globe or its equivalent. The luminous flux can of course be found indirectly, by measuring the luminous intensity in each of a large number of directions about the source and then, say, computing the average value I_{ms} and employing Eq. (7). Many ways to facilitate such computations have been devised, but the process as a whole is laborious and is seldom used. Another possible method is to surround the primary source with a highly diffusing globe of known transmission factor τ, measure the brightness of this globe and compute the total luminous flux; but this method lacks the accuracy that is often required in rating primary sources.

The almost universal practice nowadays is to measure luminous flux with the aid of an important photometric device known as the integrating, or Ulbricht, sphere. It involves a simple principle[1]; namely, if a source of light is placed within a hollow sphere whose internal wall is perfectly diffusing, the same illumination is produced at every point of the wall by light reflected from the remainder of the wall. In truth, the illumination due to the direct light from the source depends upon the distribution of luminous intensity about the source, the position of the source and the size of the sphere. But the illumination E at any point, in so far as it is due to multiple reflections from the rest of the sphere's surface, is (1) uniform over the whole surface and (2) depends only on the total luminous flux F of the source, the reflection factor ρ of the sphere's wall and the radius R of the sphere.

---

[1] This principle was first formulated by Sumpner, in 1892, in connection with an investigation of the reflection factors of various materials. The proposal to use the sphere in photometry was first made by Ulbricht, in 1900, and the theory and technic of the device have been studied extensively since that time. See Bureau of Standards Scientific Paper No. 4.

To prove the first of these two assertions, consider the cross-section of such a hollow sphere shown in Fig. 9. Any infinitesimal surface element $d\sigma$ of the sphere's wall acts as a secondary source because it reflects diffusely the light incident upon it from the source S. Since the wall is assumed to be perfectly diffusing, the element $d\sigma$ has the same brightness B in every direction, and hence its luminous intensity dI in any direction θ is $Bd\sigma \cos\theta$, by Eq. (14). Now any ray reflected from $d\sigma$ at an angle θ is incident on a point P of the wall at the angle θ (Fig. 9), and therefore the illumination dE at P, due to the light from $d\sigma$, is

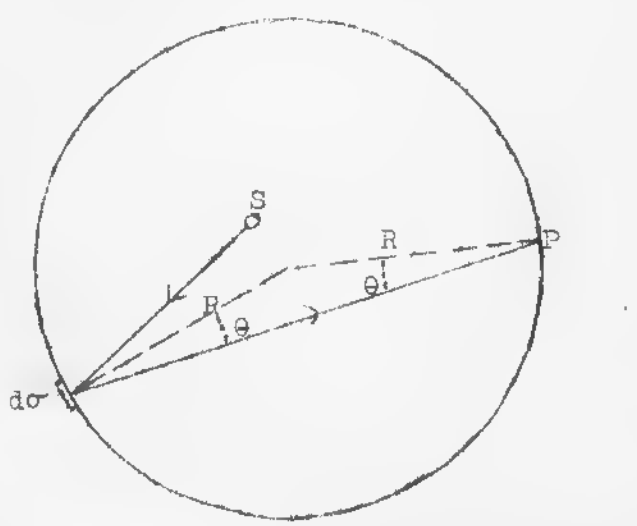


Fig. 9. Theory of the integrating sphere.

$$dE = \frac{dI \cos\theta}{s^2} = \frac{Bd\sigma \cos^2\theta}{(2R\cos\theta)^2} = \frac{Bd\sigma}{4R^2}, \tag{17}$$

where R is the radius of the sphere. Since dE is thus seen not to depend on θ, and hence is independent of the location of P, the conclusion is that $d\sigma$ produces the same illumination at every point P of the sphere's wall. Thus the total illumination E due to light reflected from the whole wall is the same at every point of the sphere's surface.

Next we shall prove that E depends only on F, the total luminous flux from the source S, on ρ, the reflection factor for the sphere's surface and on R. Since E is the same at every point of the wall,

$$E = \frac{\text{Total luminous flux reflected}}{\text{Area of wall}},$$

or

$$E = \frac{F\rho + F\rho^2 + F\rho^3 + \cdots + F\rho^{\infty}}{4\pi R^2}$$

$$= \frac{F\rho}{4\pi R^2}(1 + \rho + \rho^2 + \cdots).$$

Since $\rho^2 < 1$, this becomes

$$E = \frac{F}{4\pi R^2} \frac{\rho}{1 - \rho} \tag{18}$$

This is the relation that we set out to prove. What does $F/4\pi R^2$ represent?

The foregoing theory applies rigorously to a completely diffusing wall in an empty sphere, and hence precautions must be taken in practice to minimize departures from these ideal conditions. The spheres employed are several feet in diameter. They are coated on the inside with a special, highly diffusing paint of large reflection factor and inappreciable selectivity as regards the colors it reflects. Covering a small opening at one place in the sphere's wall is a small, opal glass window W (Fig. 10) which transmits light diffusely and whose transmission factor τ is practically independent of the color of the light.

In order to compare the total luminous flux F_1 of a source with that, F_2, of a sub-standard lamp, it is only necessary to place the lamps in turn within the sphere, to screen the window from the direct light, and to measure in each case the luminous intensity of the outer surface of the window in the normal direction. The ratio of these measured luminous intensities is equal to the ratio E_1/E_2 of the illuminations on the inner surface of the window and, from Eq. (18), $F_1 = (E_1/E_2)F_2$. The measurements of the luminous intensities of the window in the two cases need not be absolute, since only a ratio is involved; any type of photometer head, such as the Lummer-Brodhun, may be employed, although in commercial work it is often the

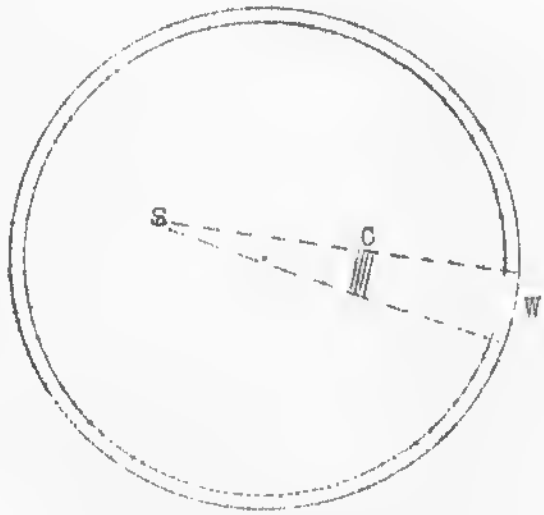


Fig. 10. The window *W* in the integrating sphere is shielded from the direct light by the small screen *c*.

practice to use a physical photometer in which a photoelectric cell, rather than the eye, is the detecting device. The luminous flux F_2 emitted by the substandard lamp must of course be determined by the older method of finding its mean spherical luminous intensity I_{ms} and employing Eq. (7).[1]

[1] For a brief discussion of the difficult subject of heterochromatic photometry, or the comparison of light of different colors, and of spectrophotometry, in which the light beams from different sources are dispersed into spectra and compared wave-length by wave-length, see, for example, *Hardy and Perrin, The Principles of Optics (McGraw-Hill, 1932), pp. 285-292, or *J. W. T. Walsh, article "Photometry", Encyclopaedia Britannica, ed. 14.

Table II. Summary of Radiation and Photometric Quantities*

Quantities	Definition	Unit
Solid angle, ω	$\omega = A/r^2$	Steradian
Radiant Energy, U	Energy transmitted in the form of electromagnetic radiation	Joule, etc.
Radiant flux, P	$P = dU/dt$	Watt, etc.
Monochromatic radiant flux, P_λ	Eq. (1)	Watt per unit wave-length
Radiant intensity, J	$J = dP/d\omega$	Watt per steradian
Monochromatic radiant intensity, J_λ	$J_\lambda = dJ/d\lambda$	Watt per steradian per unit of wave-length
Relative luminosity factor, K_λ	See Sec. 2.	A numeric
Luminosity factor, C_λ	$C_\lambda = F_\lambda/P_\lambda$	Lumen per watt
Luminous flux, F	Eq. (3) or (4)	Lumen
Monochromatic luminous flux, F_λ	$F_\lambda = \frac{dF}{d\lambda}$	Lumen per unit wave-length
Luminous energy, Q	$Q = \int_{t_1}^{t_2} F dt$	Lumen x unit of time
Luminous intensity, I	Eq. (5)	Candle
Illumination, E	Eq. (10)	Lumen per unit area ($1 \text{ lu}\cdot\text{m}^{-2} \equiv 1$ lux; $1 \text{ lu}\cdot\text{ft}^{-2} \equiv 1$ foot-candle)
Brightness, B	Eq. (14)	Candle per unit area
Flux-brightness	$dF/d\sigma$	Lambert
Reflection factor, ρ	F reflected/F incident	A numeric
Absorption factor, α	F absorbed/F incident	A numeric
Transmission factor, τ	F transmitted/F incident	A numeric
Light-source efficiency	F/Operating power	Lumen per watt

* For purposes of future reference, certain quantities are included in the table that have not been explicitly defined elsewhere in this chapter.

PROBLEMS

1. A certain monochromatic source emits 30 w of radiant flux, all in the wave-length region very close to 6.0×10^{-5} cm. Compute the luminous flux emitted.

Ans. 1.2×10^4 lu.

2. Is the total luminous flux emitted by a standard candle-source different in value from that emitted by a wax candle which has a mean spherical luminous intensity of 1 ca?

3. State the conditions under which the following statement is true: "The luminous intensity of a light-source, in a given direction, is the luminous flux incident on a surface placed normal to the given direction divided by the solid angle which this surface subtends at the source."

4. A 40-w evacuated tungsten-filament lamp is found to emit 400 lu of luminous flux when operated under the conditions for which it is designed. Compute its mean spherical luminous intensity.

Ans. 31.8 ca.

5. Compute the luminous energy Q (defined in Table II) emitted in 3.0 hr by the source described in (a) Prob. 1, (b) Prob. 4.

Ans. (a) 3.6×10^4 lu·hr; (b) 1.2×10^3 lu·hr.

6. As was first shown by J. E. Purkyné (1787-1869), for low levels of illumination approaching that of twilight the sensitivity of the eye shifts toward the blue end of the spectrum. It is now known [See Walsh, Photometry (1926), p. 65] that the luminosity curves for low levels of illumination are similar in general shape to that for ordinary levels (Fig. 2) but have their peaks shifted toward the bluish green region, being at about 5.1×10^{-5} cm for twilight illumination. Judged from the point of view of this Purkyné phenomenon alone (a) should one expect a road-sign painted blue or one painted red to be easier to see at twilight? (b) what should be the relative merits of sodium vapor arc lamps and mercury vapor arc lamps for street lighting, where illuminations and brightnesses are low, and also for factory lighting, where illuminations and brightnesses are much higher? The mercury vapor arc lamp is deficient in radiant flux in the red region of the spectrum.

7. Eq. (3) is often written in the form

$$F = \int_0^\infty C_\lambda P_\lambda \, d\lambda ,$$

where C_λ is a quantity termed the luminosity factor for radiant flux of wave-length λ. (a) How is C_λ related to the relative luminosity factor K_λ? (b) Compute the values of C_λ for monochromatic light of wave-lengths 5.55×10^{-5} cm and 6.00×10^{-5} cm, respectively.

8. Eq. (3) may often be conveniently written in the form

$$F = \int_0^\infty F_\lambda \, d\lambda$$

the quantity F_λ being termed the monochromatic luminous flux. (a) Frame an acceptable definition for F_λ that will be analagous to that for E_λ given in Sec. 1. (b) Plot a curve showing the distribution of luminous flux in the spectrum of the tungsten-filament lamp described in Fig. 1.

9. Two incandescent lamps which are at a fixed distance of 200 cm apart have luminous intensities of 16 and 32 ca, respectively, in the direction of the straight line connecting them. Determine the point where a screen placed normal to the connecting line will be equally illuminated by both lamps.

10. A screen and the small lamp which illuminates it are originally 25 in. apart, but when a certain sheet of clear glass is placed between them the lamp must be moved 2.5 in. nearer to the screen in order to produce the same illumination as before. What fraction τ of the incident luminous flux is transmitted by this glass?

11. The illumination on a circular table top of diameter 100 cm is observed to have a maximum value of 550 lux at the center of the table and to decrease uniformly at the rate of 690 lux for each meter of radial distance from the center. (a) Find the total luminous flux incident on the table top. (b) Compute the average illumination and explain why it differs from the arithmetic mean of the values at the center and edge, even though the illumination decreases uniformly from the one point to the other.

Ans. (a) 251 lu; (b) 320 lux.

12. The name phot is often given to the unit of illumination equal to $1 \cdot lu\ cm^{-2}$. (a) To how many foot-candles is 1 phot equivalent? (b) Which is the more convenient unit for most practical purposes, the phot or the milliphot?

13. A small light-source having a luminous intensity of 32 ca in every direction is placed at the focal point of a thin lens of focal length 8 cm. (a) If the lens transmits 85 percent of the luminous flux incident on it, what illumination is produced on a small screen placed in the path of the refracted light and normal to the axis of the lens? (b) In order to produce the same illumination on the screen without the use of the lens, what is the farthest distance that the screen can be from the source?

14. A single lighting unit of the type described in Example 2, Sec. 6, is suspended 12 ft vertically above one corner of a horizontal table top of dimensions 5 x 5 ft. The luminous intensity I_0 of its vertical, axial ray is 1000 ca. (a) Find the illumination at each corner of the table top. (b) If one of these lamps were suspended 12 ft. vertically above each corner of the table, what would be the illumination at the center of the table top? State all the assumptions employed in arriving at the answers.

15. A single lighting unit of relatively small size is to be so designed that it will produce the same illumination $\underline{E}$ at all points of a floor above which it is suspended at a vertical height $\underline{h}$. Prove that the necessary distribution of luminous flux is $I_\theta = \underline{Eh^2}/\cos^3\theta$, where θ is the plane angle between the vertical, axial ray and the ray of luminous intensity $\underline{I_\theta}$.

16. The reflection factor of a certain highly diffusing wall is 0.5. If the illumination at a certain point on the wall is 6 ft-candles (6 lu·ft^{-2}), what is the flux-brightness at this point?

<u>Ans.</u> 3.2 millilambert

17. A certain spherical globe of highly diffusing glass is 15 cm in diameter and contains a 200-w incandescent lamp. The globe transmits 80 percent of the luminous flux emitted by the lamp, and its surface has a fairly uniform brightness from point to point, of average value 0.3 ca·cm^{-2}. Compute (a) the average flux-brightness of the globe's surface, (b) the total luminous flux transmitted by the globe, and (c) the light-source efficiency of the lamp for the conditions under which it was operated.

18. If the globe in Prob. 17 were replaced by one made of the same kind of glass but 30 cm in diameter, what would be the average brightness of its surface?

<u>Ans.</u> 0.07 ca·cm^{-2}.

19. A flux-brightness of 1 lu·ft^{-2} is usually referred to in practice as <u>1 ft-lambert</u>. (a) Show that 1 ft-lambert is approximately equivalent to 1 millilambert. (b) If the reflection factor of a perfectly diffusing screen were unity, what would be the flux-brightness of the screen when the illumination on it is 1 ft-candle?

<u>Ans.</u> (b) 1 ft-lambert.

20. Describe how the calibration of an illuminometer can be checked with the help of a sub-standard source of known luminous intensity and a photometer bench.

21. The external test screen furnished with an illuminometer is removed, the instrument is sighted at an illuminated wall and the scale on the instrument is read. Show that the ratio of this reading to the actual illumination on the wall is ρ/ρ_t, where ρ and ρ_t are the reflection factors of the wall and illuminometer screen, respectively.

22. A certain incandescent piece of metal which is in the form of a sphere of radius $\underline{R}$ is observed to have the same brightness at every point of its surface, regardless of the angle of view. (a) Prove that if this incandescent sphere is viewed from a distance that is large compared with the diameter of the sphere, it will appear as a flat disk of uniform brightness. (b) Prove that the illumination produced by the sphere on a plane surface placed normal to the radius is inversely proportional to the square of the distance $\underline{s}$ of the surface from the center of the sphere, for all values of $\underline{s}$ between $\underline{R}$ and infinity.

23. Photographs of the sun show clearly that it is darker at the edges than at the center, this being explained by the fact that the atmosphere surrounding the sun absorbs more luminous energy in an oblique than in a normal direction. (a) Does this imply that the Lambert cosine law does not apply to the emission of sunlight, or that the brightness is not the same for all points of the sun's surface, or both? (b) When the sun is at zenith, the brightness at the center of its surface is 1.0×10^6 $ca \cdot in^{-2}$. What is the apparent flux-brightness at the center, for this angle of view? (c) Is the foregoing value larger than, smaller than, or equal to the actual luminous flux per unit area emitted at the center?

Ans. (b) 3.1×10^6 $lu \cdot in^{-2}$.

24. Instead of specifying the brightness of an extended source, the practice sometimes is to specify the apparent luminous intensity for a specified distance and direction, this being defined as equal to the luminous intensity of a point-source that would produce the same illumination at that distance. Measurements with an illuminometer show that a certain mercury arc lamp produces an illumination of 32 $lu \cdot ft^{-2}$ on the test screen placed normal to its rays and at a distance of 5 ft vertically below the unit. Compute the apparent luminous intensity.

Ans. 800 ca at 5 ft vertically downward

25. The inside wall of an integrating sphere is coated with paint for which the reflection and absorption factors are ρ and α, respectively. A small incandescent lamp is placed within the sphere but no screen is placed between the lamp and the translucent window. (a) Show that the illumination produced on the inner surface of the window by the light reflected by the wall of the sphere is $E_d(1 - \alpha)/\alpha$, where E_d is the average illumination produced on the wall by light coming directly from the lamp. (b) Hence show that the average illumination produced on the window by both the reflected and direct light is E_d/α, or $E_d/(1 - \rho)$. (c) Is there anything in the theory of the integrating sphere that requires that the lamp be located at the center of the sphere? (d) Is it true that the ratio of the total luminous flux received by the sphere's wall to that emitted by the lamp is $1/\alpha$ and hence larger than unity? Reconcile your answer with the principle of conservation of energy.

26. A 60-w evacuated tungsten-filament lamp of light-source efficiency 10 $lu \cdot w^{-1}$ is placed in an integrating sphere of inside diameter 46 in. The reflection factor of the inside coating of the sphere is 0.95. Compute (a) the luminous flux emitted by the lamp, (b) the illumination on the sphere's wall produced solely by the light reflected from it, (c) the average illumination on the wall produced jointly by the reflected and direct light, (d) the total luminous flux incident on the wall, (e) the luminous flux absorbed by the wall, (f) the brightness of the wall at any point and for any angle of view, and (g) the flux-brightness at any point of the wall.

27. In dealing with photometric problems by dimensional methods, it is useful to introduce two fundamental units in addition to those of length [L], mass [M] and time [T]. The units of luminous flux and of solid angle are a convenient choice, their dimensions being denoted by the symbols [F] and [Ω], respectively. Write the dimensional formula for each of the quantities listed in Table II. (b) Do the two members of Eq. (12) have the same dimensional formula? Explain.

CHAPTER FOUR

INTERFERENCE AND DIFFRACTION

> Suppose a number of equal waves of water to move upon the surface of a stagnant lake with a certain constant velocity, and to enter a narrow channel leading out of the lake; suppose then, another similar cause to have excited another equal series of waves which arrive at the same channel with the same velocity and at the same time as the first. One series of waves will not destroy the other, but their effects will be combined. If they enter the channel in such a manner that the elevations of the one series coincide with the other, they must together produce a series of greater joint elevations; but if the elevations of one series are so situated as to correspond to the depressions of the other, they must exactly fill up those depressions, and the surface of the water must remain smooth -- at least, I can discover no alternative, either from theory or experiment. Now, I maintain that similar effects take place whenever two portions of light are thus mixed, and this I call the general law of the interference of light.
>
> Thomas Young, quotation from
> Bakerian Lecture before the
> Royal Society in 1801.

In our treatment of light in the previous chapters we have found it convenient to confine our attention to the lines along which the light travels. Such lines we have called rays, and therefore up to this point our description of light has been in terms of ray-optics or geometrical optics. Although it is possible in terms of geometrical optics alone to obtain quantitative descriptions of many properties of light, including a number which have found wide application in practice, such a description however must be considered as only an approximation to a true description of light, and limited in its application. There is a large class of phenomena which can be understood only in terms of the wave properties of light. A description of these phenomena which will be treated in the present chapter form the subject matter of what is known as physical optics as distinct from geometrical optics.

Two theories of light. In Sir Isaac Newton's day (1642-1727) two rival theories of light were struggling for recognition. The one, the wave theory, fathered and championed by the Dutch physicist, Christian Huygens (1629-1695), regarded light, like sound, as some sort of a wave motion, the chief difference between the two being, according to his theory, that, while sound is propagated through the agency of ordinary matter, light is a wave motion in some all-pervading medium to which the name of "the ether" was given.

The rival theory, called the corpuscular theory, regarded light as due to the emission from all luminous bodies of minute corpuscles which travel in straight lines and with enormous velocities through space and produce the sensation of light when they impinge upon the retina of the eye. This theory had its most famous and most brilliant advocate in Sir Isaac Newton himself.

Newton's chief reason for rejecting the wave theory lay in the fact that he was unable to understand why, if light is a wave motion, it is always propagated in straight lines past the edges of opaque objects, instead of undergoing diffraction, that is, being bent around such objects, as are sound waves, water waves, and all the other types of waves with which Newton was familiar. What is commonly regarded as the decisive test between the two theories was made in the year 1800 by Thomas Young, and consisted in showing that it is possible to produce with light waves the diffraction phenomena which are to be discussed in this chapter, and which it does not seem possible to account for from the standpoint of the corpuscular theory.

It is the object of the present discussion and of the succeeding experiments to show both theoretically and experimentally that, under suitable conditions, light does bend around corners. More explicitly stated, our aim will be to show that the phenomenon of straight-line propagation is characteristic of any and all types of wave motion, provided only the aperture through which the waves pass is large in comparison with the wave length of the waves. If this proposition can be proved, it will be evident that the fact of the straight-line propagation of light does not furnish any argument against the wave theory, provided the wave length of ordinary light waves is very minute in comparison with the dimensions of ordinary apertures. Before proceeding to this proposition it is necessary to consider further the nature of a wave motion in a medium of indefinite extent, and the conditions for interference in such a medium.

1. Definition of a Wave-Front. Consider S in Fig. 1 to be the point source of a wave motion in an isotropic medium; that is, a medium in which the disturbance is propagated with equal speed in all directions. When the disturbance which originates at S has just reached a, it has also then just reached all other points, such as b, c, and d, which are at the same distance from S. The spherical surface passing through these points is known as the wave front of the disturbance. In general, the wave front may be defined as the

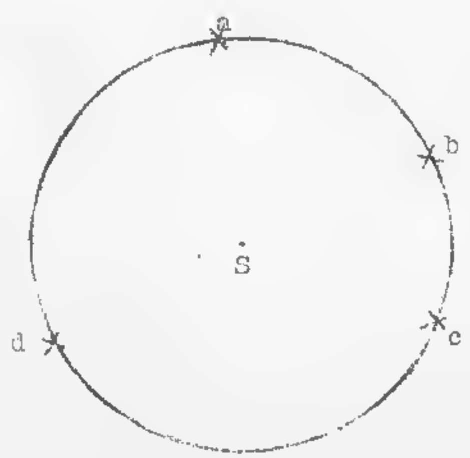


Fig. 1.

surface passing through all the particles which are in the same phase of vibration.

The form of the wave front under the conditions just mentioned is spherical, but conditions may arise in which it has not this form. Further, it will also be shown that under proper conditions a spherical wave may be converging, i.e. concave toward the direction in which it is traveling, instead of diverging, as in the case just considered. If the source is far enough away any small portion of the wave will be sensibly plane. A wave having a plane wave front is called a plane wave.

2. Huygen's Principle -- Construction of a Wave Front. In Sec. 2 of Chap. I it was pointed out that the principle of superposition is fundamental in the study of light. According to this principle the net effect resulting from the action of several rays of light may be determined by considering the sum of the effects of the various individual rays. This principle can be carried over to all considerations in which light is treated as a wave motion. Hence the amplitude[1] of a wave at any point is the vector sum of the amplitudes of all the elementary waves which act at that point. Use is made of this fact in the application of Huygen's principle for the purpose of following the propagation of a wave by constructing a new wave front in terms of a knowledge of the wave front which was present at an earlier instant of time. Huygen's principle was discussed in Chap. 15 of MRW, and we shall here merely summarize the principle as follows: Huygen's principle states that at any instant of time the wave front of a disturbance is the envelope of all the secondary wave surfaces which are due to the action as separate sources of all the various particles that at some previous instant constituted the wave front.

---

1 For an introductory treatment of wave motion see MRW, Chap. 15.

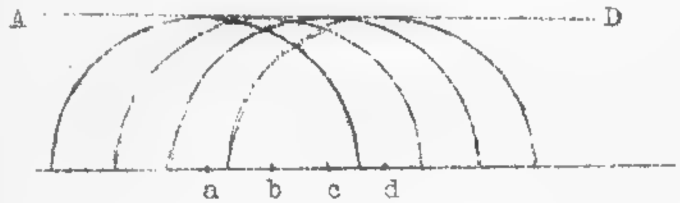


Fig. 2.

For example consider the particles a, b, c, and d in the plane wave represented in Fig. 2. A short time after the disturbance has reached these particles let the spherical wave surfaces due to them have the forms shown in the figure. If the number of these new centers is very large, it is evident from the figure where for clearness only four have been represented, that the disturbance along the envelope AD is very much greater than at any other points, for this envelope represents the locus of points all of which are in the same phase of vibration. It is not immediately apparent that a disturbance should not also be propagated back in the direction from which the incident plane wave moves, but a mathematical analysis, in which account is taken of the magnitudes and directions of the amplitudes of the secondary waves, as well as of the phase relationships between them, shows that the secondary waves from the individual sources do actually destroy one another except at the surface AD. We shall not here attempt a rigorous demonstration of the principle, but merely apply it in the simplified form given in the summarized statement above.

3. Conditions of Interference of Two Wave Trains in a Medium of Indefinite Extent. From the fact that the resultant amplitude at a given point is equal to the vector sum of the amplitudes of all the secondary waves, one would expect that at some points the amplitudes of the secondary waves might add together to produce a disturbance of considerable amplitude whereas at other points their effects might wholly cancel one another.

Experiment shows that such phenomena actually occur, and points such as the former are called points of <u>constructive interference</u>, and the latter points of <u>destructive interference</u>. The variations in amplitude with position which result from the combined action of two or more waves are called <u>interference effects</u>.

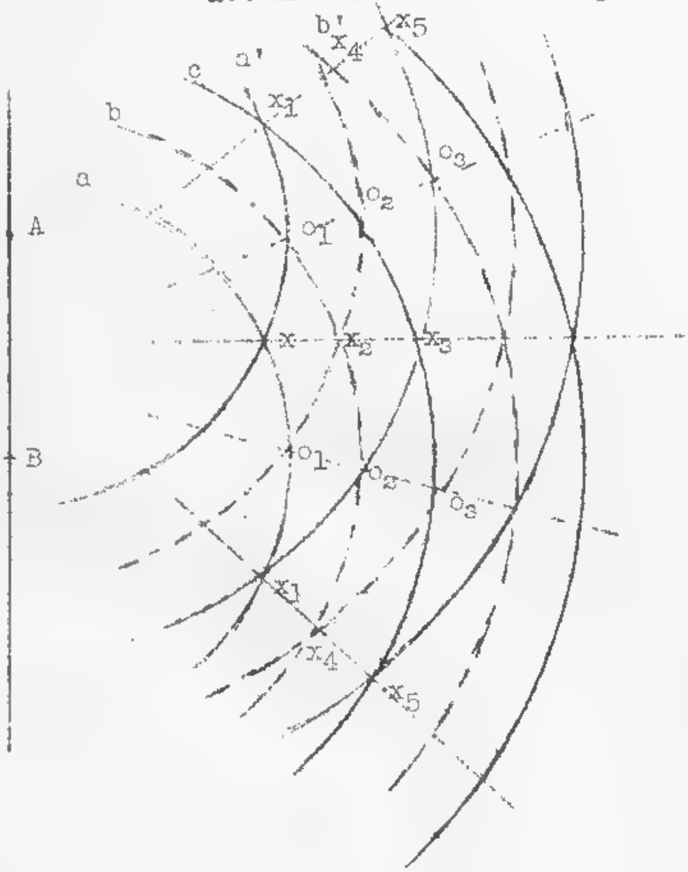


Fig. 3.

Let us consider for example two particles <u>A</u> and <u>B</u> (Fig. 3) vibrating in the same plane, in phase with one another, and with equal amplitudes. From each of these points is propagated a disturbance having a spherical wave front. Let similar wave fronts be constructed for each particle. Thus the circular arcs <u>a</u> and <u>a'</u> represent the position of particles at the same distance from their respective sources and therefore in the same phase. The arcs <u>b</u> and <u>b'</u> represent the wave fronts when the disturbances have traveled one half a wave length farther; that is, each of them represents the locus of a series of particles which are exactly opposite in the phase of their vibration to the particles of <u>a</u> and <u>a'</u>. The arcs <u>c</u> and <u>c'</u> represent the wave fronts when they have traveled a whole wave length beyond <u>a</u> and <u>a'</u>. Their particles are in similar phase to those of <u>a</u> and <u>a'</u> and opposite to those of <u>b</u> and <u>b'</u>.

The particles in the line determined by the points marked x, x_2, x_3 have superimposed upon them vibratory motions of the same phase from both sources. Along this line there is therefore a maximum disturbance or constructive interference. Along the line determined by the points marked ρ_1, ρ_2, ρ_3, on the other hand, the vibrations superimposed are opposite in phase, and there is a minimum disturbance or destructive interference. Further, along the line determined by the points x_1, x_4, x_5 there is again reenforcement. From the construction of the figure it is evident that the condition for a maximum at any point is the existence of a difference in length of path between the point and the sources A and B respectively of some integral multiple of a whole wave length. Thus at x the difference in path is zero wave lengths, at x_1 it is one wave length, etc. Similarly, for a minimum the difference in distance must be an odd multiple of a half wave length. At ρ_1, ρ_2, ρ_3, etc., it is 1/2 wave length. Additional maxima and minima may be found by extending the lines a, a', b, b', etc.

Fig. 3 indicates the interference pattern at a given instant of time. A study of the figure will show, however, that as the two spherical waves originating at A and B travel outward, the lines of maximum disturbance x, x_2, x_3 and x_1, x_4, x_5 and also the lines of minimum disturbance ρ_1, ρ_2, ρ_3, will maintain the same position with respect to A and B at every instant of time. The interference pattern therefore remains fixed in position although the two sets of spherical waves which produce it travel outward from the two centers A and B respectively.

It is important to notice that the lines of minimum disturbance ρ_1, ρ_2, ρ_3, etc., move farther and farther away from the central line of maximum disturbance x, x_2, x_3, the smaller the distance AB becomes in comparison with a wave length. Thus if AB is very large in comparison with a wave length, the line Q_1, Q_2, Q_3 is very close to the line X, X_2, X_3, and similarly the line X_1, X_4, X_5 is close to the line Q_1, Q_2, Q_3. But as AB becomes smaller and smaller these lines diverge more and more. When AB is just equal to a wave length the line X_1, X_4, X_5 is in the prolongation of AB, since it is only points in this line which can then differ by one wave length in their distances from A and B respectively. When AB is equal to a half wave length the line Q_1, Q_2, Q_3 is in the prolongation of AB, and there are then no points of quiescence at all to the right of AB. When AB is less than a half wave length there are no points of quiescence anywhere.

An interference pattern such as that of Fig. 3 can readily be produced with waves on the surface of a liquid or with sound waves.

4. Interference of Light Waves. An experiment showing the interference of light was first performed by Young about 1800. He observed an interference pattern produced on a screen P when sunlight was allowed to pass through a pinhole S in one opaque screen and then through two pinholes S_1 and S_2 in a second screen (Fig. 4). If the distance from S to S_1 is equal to the

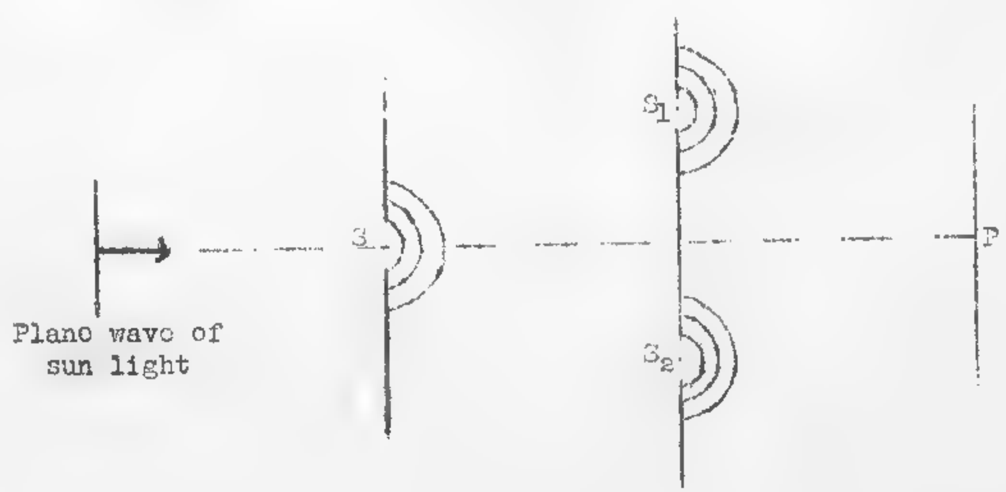


Fig. 4.

distance from S to S_2, then S_1 and S_2 may be considered as two sources of light which are vibrating in phase with one another, and an interference pattern which is just like that indicated in Fig. 3 is produced in the region to the right of S_1 and S_2, and which results in the production of an interference pattern represented by alternate light and dark regions on the screen P.

In the case of light there are certain conditions to be fulfilled in order that interference may be observed which usually are not present in experiments with sound waves. A source of light whether it be the sun, an incandescent body, or whatever else, usually consists of a very large number of atoms each acting separately as a source of light. Now experiment has shown that the length of time during which an individual atom radiates light is of the order of 10^{-8} sec, and since the velocity of light is equal to 3×10^{10} cm/sec, the length of the wave train of the light emitted by an atom is about $3 \times 10^{10} \times 10^{-8}$ cms or 3 meters. This is to be contrasted, for example, with the length of the train of waves of sound emitted by a vibrating string or organ pipe. In this case a wave train can be produced which is indefinitely long simply by causing the string or the organ pipe to emit sound at a definite frequency for an indefinitely long time.

This fact, as will be shown later, has important consequences in the production of interference phenomena in light. A second important characteristic of a light beam is that, since many million atoms usually contribute to its origin, it therefore consists of the superposition of many millions of elementary wave trains, no two of which bear a fixed phase relationship to one another. Two or more wave trains whose phases do not bear a fixed relationship to one another are called incoherent, and wave trains whose phases do bear fixed relationships to one another are called coherent. Hence two elementary wave trains representing light which originates in two different atoms are incoherent with respect to one another.

Now it is clear from the discussion of Sec. 3 that in order that an interference pattern will be produced the waves emanating from the two sources must be coherent. A coherency of the waves can be achieved in the case of light by a device such as that represented in Fig. 4. For consider the light emitted by a single atom of the source. It enters the pinhole S_1 and then emerges on the other side of the screen as a spherical wave if the pinhole is sufficiently small. This spherical wave strikes both the pinholes S_1 and S_2, and produces two new spherical waves of light which, since they are both derived from the same elementary wave-train, do bear a definite phase relationship to one another, and are therefore coherent and capable of forming an interference pattern. The same will be true of all the other elementary wave trains in the beam. The interference pattern is produced then by the combination of two beams which have taken different paths but which have originated in the same atom. Obviously therefore if two independent point sources of light are substituted for S_1 and S_2 in Fig. 4 no interference pattern will result.

In Sec. 3 it was stated that a point of destructive interference will result if the difference in length of the two paths is an odd multiple of a half wave length. Now the fact that wave trains of light are limited in length places an upper limit on the difference in length of the two paths, for in order that two wave trains can combine to produce interference the difference in path length cannot exceed an amount about equal to the length of a wave train itself. Actually the path difference must be shorter than the length of a wave train, and experimentally the greatest permissible difference in path length is found to be somewhat less than one half meter.

In order that a clearly defined interference pattern will be formed, the light must also be monochromatic or nearly so, since otherwise the difference in path length measured in units of half a wave length varies with the color of the light. For example, if white light is used a point of destructive interference for one color may coincide with a point of constructive interference for a second color. It is apparent also that if the path difference is very small, of the order of a few wave lengths, then fairly sharp patterns will be formed even though the light is not monochromatic.

Measurement of Wave length by Interference

By measurements made on an interference pattern the wave length of the light used can be determined directly in a very simple manner. In Fig. 5 let S_1 and S_2 represent two coherent point sources of light such as those indicated in Fig. 4, and let the point O on the screen represent the position of the central bright fringe or central maximum. Then

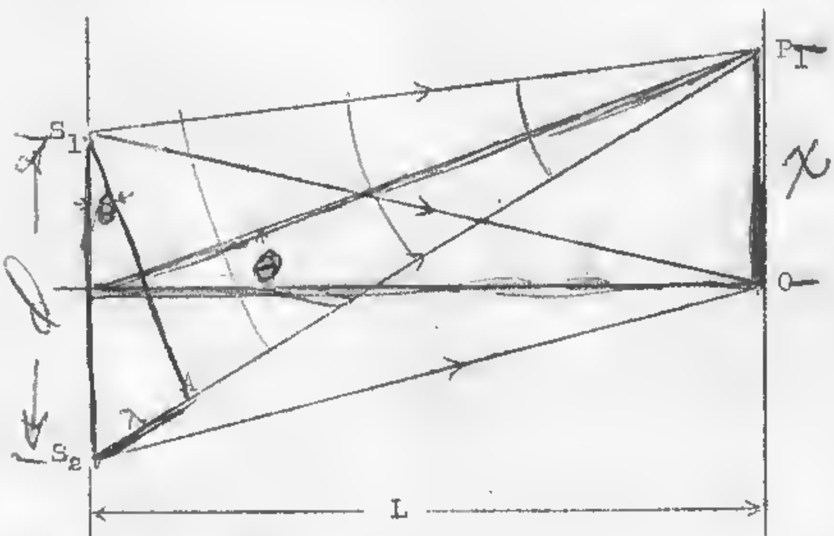


Fig. 5

$$S_1O = S_2O$$

$\theta \cong \frac{\lambda}{d} \cong \frac{x}{L}$

Let P_1 represent the position of the next bright fringe or point of maximum disturbance. This will occur when the difference in path length measured to the two sources is one wave length. Hence

$$S_2P_1 = S_1P_1 + \lambda$$

Choose the point A so that $S_1P_1 = AP_1$; then S_2A is equal to λ. If now the distance S_1S_2 between the two point sources is made large compared with λ, and if L is large compared with S_1S_2, then from Fig. 5,

$$\theta = \frac{\lambda}{S_1S_2} = \frac{OP_1}{L}$$

Letting the distance between S_1 and S_2 be d, then the separation between the central bright fringe and the one adjacent to it, becomes

$$OP_1 = \frac{\lambda L}{d}$$

Furthermore bright fringes will occur at all points where the difference in path length is an integral number of wave lengths, hence the distance x from the central maximum to the nth bright fringe is

$$x = \frac{n\lambda L}{d} \qquad (1)$$

Thus by making very simple measurements the wave length of the light can be determined.

Various devices have been employed for obtaining interference patterns from a single point source of light. Two of these devices, first used by Fresnal, are known as Fresnel's mirrors and Fresnel's bi-prism. In the former light from a point source S (Fig. 6) is reflected from two plane mirrors, M_1 and M_2, whose surfaces make a very small angle with each other. Two virtual images S_1 and S_2 are formed, which serve as the two sources which produce the interference pattern on

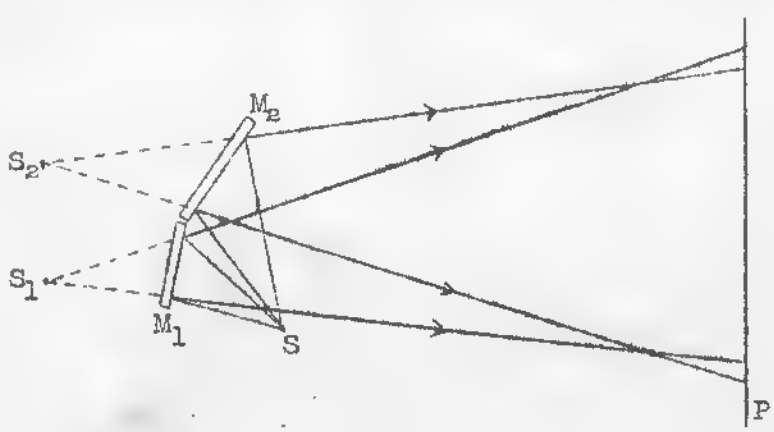


Fig. 6

The screen P. A similar result is obtained by means of the bi-prism (Fig. 7) which consists essentially of a double prism with a very small refracting angle. Again the light reaching the screen P apparently comes from the two

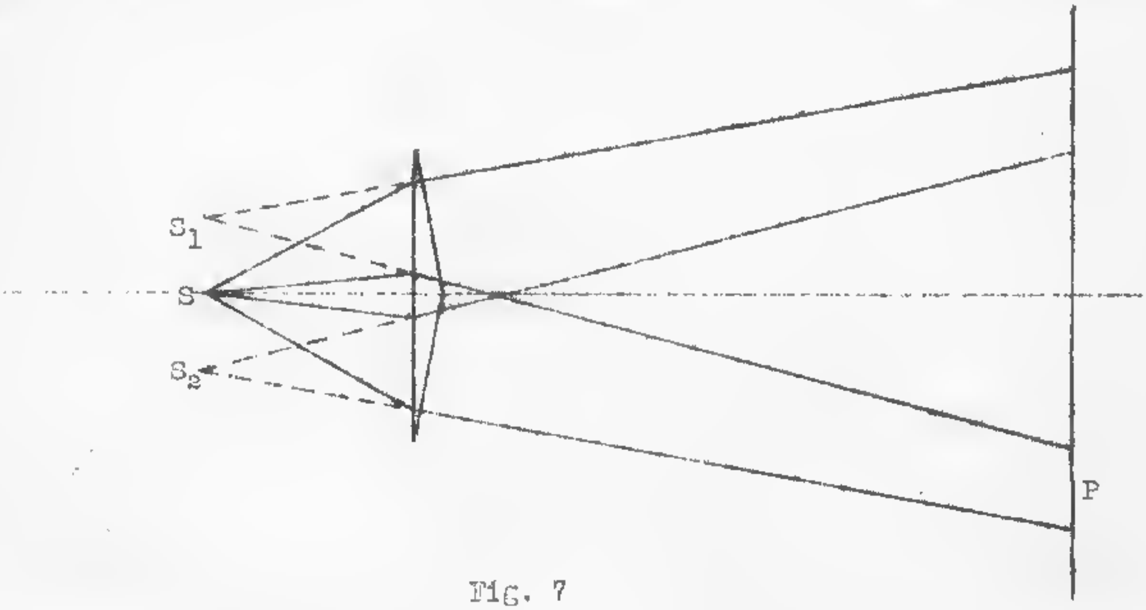


Fig. 7

virtual sources S_1 and S_2. Either of these methods, because of the relatively large solid angle from which the light is gathered, will produce fringes which are much brighter than those formed by the pin hole arrangement employed by Young.

<u>Interference Patterns Produced by Plane Parallel Surfaces and Thin Films</u>

In our discussion so far we have considered interference patterns produced only by the superposition of two disturbances which have traveled over two different paths but which have had their origin in the same point source. It was essential in those cases that the source of light be essentially a point, for otherwise several patterns displaced in position from one another would have been produced, and the overlapping of these would have resulted in uniform illumination. We shall now consider other ways in which interference patterns may be produced for the observation of which an extended source of light is required.

Let us first consider a plate of glass whose sides are accurately plane and parallel to one another. A pencil of parallel rays from a medium of refractive index μ_1 is incident at an angle i upon one surface of the glass plate of refractive index μ_2, as indicated in Fig. 8. The incident rays are in part reflected and in part refracted at the surface of the plate. The refracted rays after reflection from the second surface of the plate

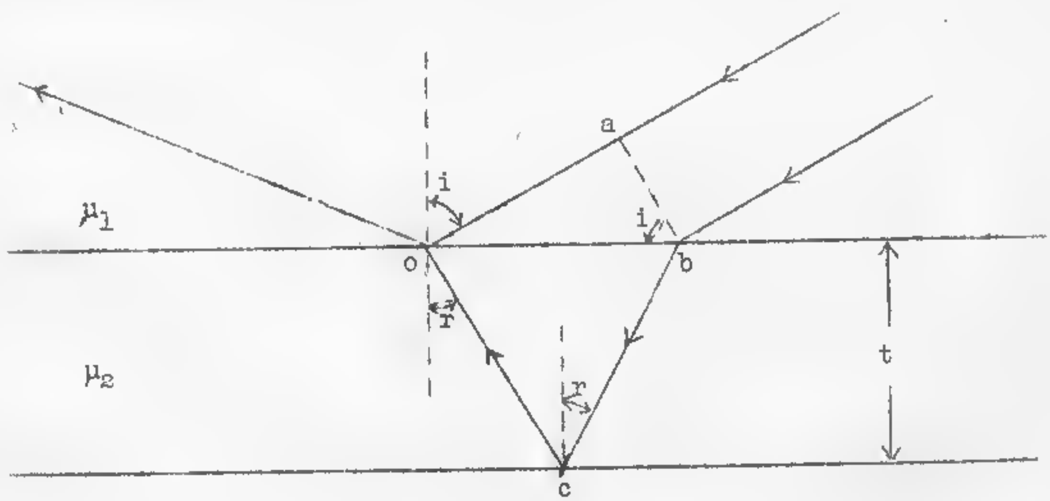


Fig. 8

emerge from the plate in a direction parallel to the reflected rays and hence is in a position to interfere with the reflected rays. Either constructive or destructive interference occurs depending upon the phase relationship of the two rays. A wave front of the incident rays is ab. From the position of this wave front the reflected ray travels a distance oa to reach o. The refracted ray travels a distance bc plus co to reach o. If L_1 and L_2 represent the optical paths of these two rays, respectively, then from the figure,

$$L_2 = 2\ell\mu_2$$

and $$L_1 = \mu_1 ob \sin i = 2\mu_1 \ell \sin r \sin i$$

where $\ell = bc$, and where for the optical path we take the product of the geometrical path and the refractive index (Sec. 5, Chap. I). The difference in optical path is thus given by

$$L_2 - L_1 = 2\ell(\mu_2 - \mu_1 \sin r \sin i)$$

But $$\mu_1 \sin i = \mu_2 \sin r$$

hence $$L_2 - L_1 = 2\ell\mu_2(1 - \sin^2 r) = 2\ell\mu_2 \cos^2 r$$

Since $\ell \cos r$ is equal to the thickness t of the film,

$$L_2 - L_1 = 2t\mu_2 \cos r \qquad (2)$$

It is necessary to add a second term to the right hand side of Eq. (1) since a change in phase of π radians occurs upon reflection if the waves impinge upon a medium of greater optical density.[1] This will be true for reflection at the point o since we shall take μ_2 to be greater than μ_1. A change in phase of π radians is equivalent to an increase in the path difference by an amount $\lambda/2$, where λ is the wave length of the light. Hence Eq. (2) becomes

$$L_2 - L_1 = 2t\mu_2 \cos r - \frac{\lambda}{2} \qquad (3)$$

Now it will be remembered from Sec. 3 that if the difference in path length is an even multiple of a wave length constructive interference occurs, and if it is an odd multiple of a half wave length the interference is destructive. Thus if we write Eq. (3) in the form

$$2t\mu_2 \cos r - \frac{\lambda}{2} = k\lambda$$

$$\text{or} \quad 2t\mu_2 \cos r = (k + \tfrac{1}{2})\lambda \qquad (4)$$

---

[1] If the waves strike a medium of lesser optical density no change in phase occurs. A corresponding relationship holds when a sound wave is reflected from a medium of different density (See MRW, p. 391).

the interference will be constructive for k = 0, 1, 2, 3 ··· and destructive for $k = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$.

Actually, however, each incident ray can suffer several partial reflections within the plate, and emerge as a series of parallel rays as indicated in Fig. 8. Furthermore, since the incident plane parallel wave

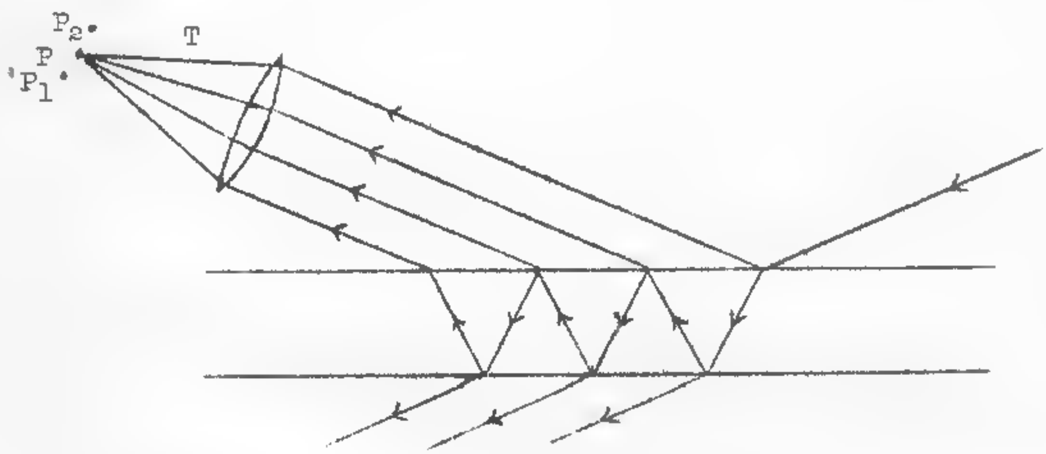


Fig. 9.

front will in general be broad, a broad beam of parallel light will be reflected from the plate which may then be observed either by eye or by means of a telescope T, either of which must be focused for infinity. By considerations similar to those above it is apparent that between adjacent pairs of rays of Fig. 8 there will be the same path difference, $2t\mu_2 \cos r - \frac{\lambda}{2}$, given by Eq. (3). It follows then that the intensity observed either by eye or by the telescope T will depend upon the angle r, if t and μ_2 are kept constant. Now if plane parallel light is incident on the plate at all angles, and if the plate is not too thin, i.e., if its thickness corresponds to a sufficiently large number of wave lengths of the light employed, then a set of dark and bright fringes will be observed corresponding to different angles of incidence of the incident light. This can readily be seen by Eq. (3) since if the plate is sufficiently thick only a small change in the angle r will be sufficient to change the difference in path by $\lambda/2$. For example, if $2t\mu_2 = 10{,}000\ \lambda$, then a change in cos r of one part in 20,000 will correspond to a change in path difference of $\lambda/2$, and the adjacent bright fringes such as P_1, P, and P_2 of Fig. 8 will be sufficiently close to one another so that several fringes are encompassed by the image of the eye or the telescope. Now if the thickness of the plate is considered to decrease then the angular separation between adjacent bright fringes will increase until for a plate thickness of only a few wave lengths the angle of separation between adjacent fringes will exceed the field of view of the eye or telescope and a set of fringes can no longer be observed. As indicated in Fig. 8 the same effects can also be observed in the light transmitted by the plates. Fringes of this type are known as <u>fringes of equal inclination</u>.

Interference effects, however, can also be observed in the case of a transparent medium whose thickness is of the order of only a wave length of light. Examples of these effects are the phenomena of color and the light and dark bands often seen in soap bubbles, thin films of oil, or in very thin flakes of glass. Interference effects of this type are in some respects quite different from those discussed above, and sets of interference fringes can be observed in thin films only if the bounding surfaces are not parallel to one another. Color effects, however, are observed in thin films bounded by parallel surfaces.

In a manner similar to our previous treatment we now consider light incident upon a medium of index of refraction μ_2 bounded by two nearly parallel surfaces which make a small angle δ with one another. In Fig. 10 let R_1 be a ray of light from the source S_1, which is reflected from the upper

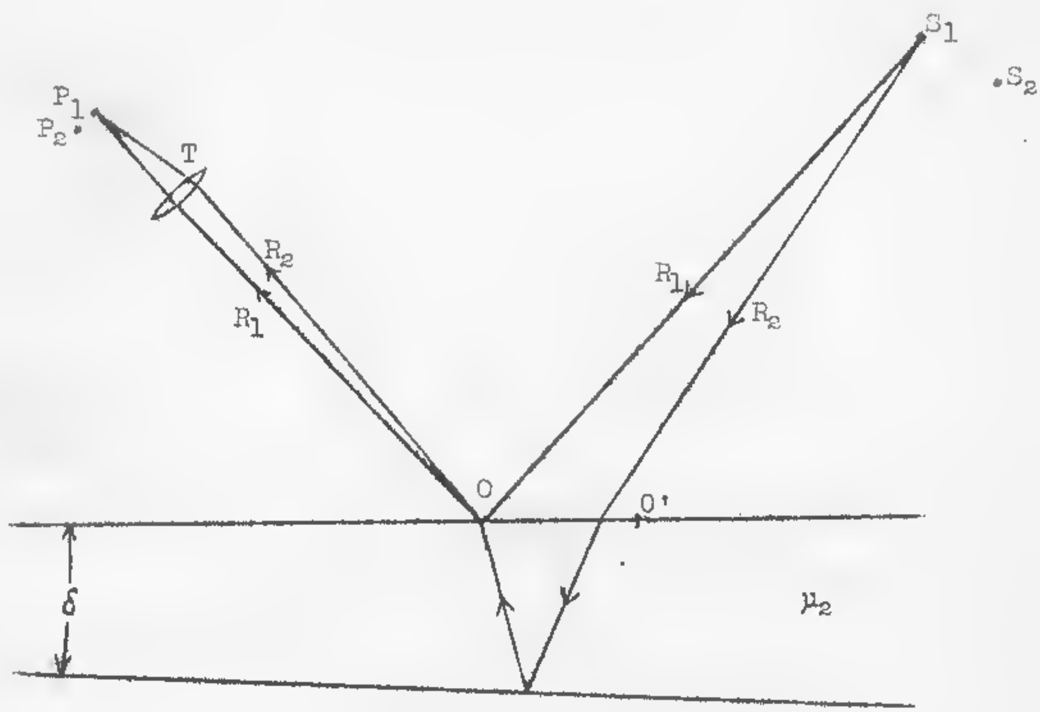


Fig. 10

surface at O, and enters an eye or the telescope T. The eye or telescope in this case is focused on the point O of the film. Let R_2 be a second ray from the source which is so chosen that it also passes through the point O, but only after passing through the film and being reflected at its lower surface. Since the angle δ is small this ray will be nearly parallel to R_1 and will also enter the telescope and since the telescope is focused on the point O, both R_1 and R_2 will be brought to focus at the same point P_1. Now at the point P_1 we may have either constructive or destructive interference depending upon the phase relationship of the two disturbances at that point due to the two rays R_1 and R_2. If the path difference between R_1 and R_2 is an even integral multiple of $\lambda/2$ constructive interference will occur, and if it is an odd integral multiple of $\lambda/2$ the interference will be destructive.

We have found in our previous treatment where parallel light was incident on a plate of plane parallel sides (Fig. 9) that the difference in path length between adjacent rays was given by Eq. (3) as

$$L_2 - L_1 = 2t\mu_2 \cos r - \frac{\lambda}{2} \qquad (5)$$

For the case we are here considering it can be shown that the same expression is a sufficiently valid approximation if $\underline{t}$ is taken to be the thickness of the plate at $\underline{Q}$, and $\underline{r}$ is the angle of refraction of either $\underline{R}_1$ or $\underline{R}_2$, provided that the thickness of the plate is small compared with the distance from $\underline{Q}$ to the source $\underline{S}_1$ and provided that $\underline{\delta}$ is a small angle. The exact expression differs from the above only in negligibly small terms of the second or higher order in δ. The trigonometric details of the derivation, however, will not be given here.

It is apparent that in order to obtain an extended image of an interference pattern in the telescope by this method an extended source of light must be employed, for if there were only a single point-source $\underline{S}_1$ then the light which is reflected from the film and enters the telescope can come only from a small region of the film in the immediate neighborhood of $\underline{Q}$. Thus, with reference to Fig. 10, let $\underline{S}_2$ be a second point of an extended source of light, and consider two rays from $\underline{S}_2$, corresponding to the rays $\underline{R}_1$ and $\underline{R}_2$ from $\underline{S}_1$, which are brought to focus in the telescope at $\underline{P}_2$ after having passed through a point Q' in the surface of the film. For these rays the values of both $\underline{t}$ and $\underline{r}$ in Eq. (5) will differ from the corresponding values for the rays from $\underline{S}_1$, and hence the phase relationship of the two disturbances focused at $\underline{P}_2$ will differ from that of the two disturbances focused at P_1. Let us assume first that the thickness of the film does not vary between $\underline{Q}$ and $\underline{Q}'$. Then the change in the path difference between the reflected and refracted rays at $\underline{Q}$ as compared with that at $\underline{Q}'$ will depend only upon the amount by which cos _r_ changes between _Q_ and _Q'_. As previously pointed out for the case of a plate of constant thickness, a relatively large change in cos _r_ is required to produce a change in the path difference by an amount $\lambda/2$, if the plate is thin. Hence one would not expect in this case to observe a set of interference fringes for then the separation between adjacent fringes in the image formed in the telescope would correspond to a distance greater than the size of the image itself.

This, however, is not the case if the film is of a variable thickness, for then a change in the thickness _t_ could be sufficient by itself to produce a change of $\lambda/2$ in the path difference between the reflected and refracted rays at _Q_ as compared with those at _Q'_. Thus the distance between a bright fringe and an adjacent dark one in the image of the telescope could, for example, be equal to a relatively small distance P_1P_2, and the observation of a set of fringes within the field of view of the telescope becomes possible.

Example 1. Plane waves of Na-D light are incident at all angles upon a thin sheet of glass ($\mu = 1.500$) of constant thickness 10^{-4} cm. Compute the angles with respect to the plane of the glass plate of those directions for which complete constructive and complete destructive interference occurs. Treat also the case where the glass plate is 1 cm thick, and compare the angular separation of the directions which correspond to constructive and destructive interference in the two cases.

Quite brilliant patterns of color are often seen when a thin film of a transparent medium is observed in white light. These effects are well known in the case of soap bubbles, oil films on water, etc. The appearance of color arises from the fact that, because of the different wave lengths corresponding to different colors of light, destructive interference may occur for one portion of the spectrum and constructive interference occur for another. From the discussion given above it will be clear why a thin film whose thickness is uniform over a given region will appear a uniform color over that region, whereas if the film is wedge-shaped a pattern of colored bands will appar. A thin film of an evaporating liquid will appear black just before it breaks. Why?

Newton's Rings

An experiment performed by Newton affords a beautiful example of the interference of light. A plano-convex lens of large radius of curvature is pressed against a plane glass plate. The space between the lens and the plate then consists of a thin layer of air of varying thickness. Interference is produced in this layer of air which gives rise to an interference pattern of concentric circular dark and light bands. The pattern is known as Newton's Rings.

By an application of the principles outlined above we can obtain a relationship among the radii of the interference rings r, the radius of curvature of the surface of the lens R, and the wave length of the light, (Fig. 11).

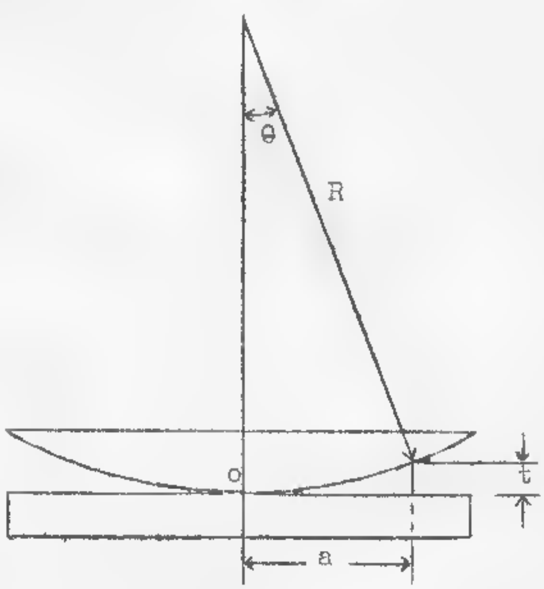


Fig. 11

From the figure we have

$$t = r(1 - \cos\theta) = 2R \sin^2 \frac{\theta}{2}$$

If the angle θ is small then $\sin^2 \frac{\theta}{2}$ may be set equal to $\left(\frac{\theta}{2}\right)^2$ or $\frac{\theta^2}{4}$, and θ may be set equal to a/R. Hence

$$t = 2R \cdot \frac{a^2}{4R^2} = \frac{a^2}{2R} \qquad (6)$$

If light is incident from above the lens and if the interference pattern is viewed from above, then the conditions are analagous to those which obtained in the thin film of Fig. 10, and we can apply Eq. (4). In the present case since the layer of air is optically less dense than the layers of glass which bound it, a change of phase will occur in the reflection from the lower surface of the layer of air but not in the reflection from its upper surface. This is opposite to the former case, but the student can easily demonstrate the validity of Eq. (4) for both cases. We may take the refractive index of air equal to unity, and if we substitute for t from Eq. (6) into Eq. (4), we obtain ($\mu_2 = 1$ for air film)

$$\frac{a^2}{R}\cos r = (k + \frac{1}{2})\lambda \qquad (7)$$

If the interference pattern is viewed in a direction which is nearly normal to the surface cos r may be set equal to unity, thus,

$$\frac{a^2}{R} = (k + \frac{1}{2})\lambda \qquad (8)$$

where from Eq. (4) the condition for bright rings is obtained when $k = 0, 1, 2, 3 \cdots$ and for dark rings when $k = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots$

If the rings are viewed in white light then one obtains a superposition of many patterns where the radii of the corresponding rings of each pattern are proportional to $\lambda^{1/2}$, hence giving rise to color effects.

5. The Propagation of Waves through Apertures. As previously mentioned there are certain conditions under which light is observed to bend around corners. This phenomenon is usually concerned with interference effects produced when an opaque object is placed so as to intercept a portion of the cross section of a wave front. Such effects are called diffraction effects.

Diffraction Pattern Due to a Single Slit

We shall treat first the diffraction pattern produced by a single slit in an opaque screen. Consider the case of a train of short waves, which, proceeding from a distant source, passes through an opening ac (Fig.12) and falls upon a screen mn. A distant source is chosen so that the

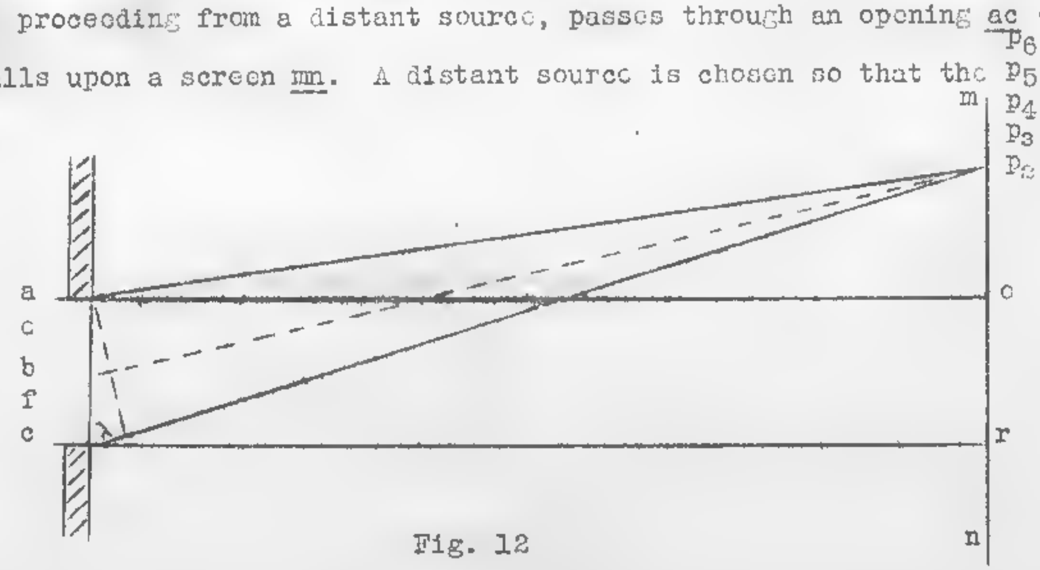


Fig. 12

wave front of the disturbance which reaches the aperture ac may be practically a plane, and thus admit of the consideration of all the particles lying in the plane of the aperture as being in the same phase of vibration.

The lines ao and cr are drawn from the source, assumed to be a point, past the edges of the opening ac to the screen; i.e. they are the lines which mark the limits of the geometrical beam. Suppose that the wave length and the opening ac are so related that the point p_2 on the screen, for which the distance cp_2 is exactly one wave length greater than the distance ap_2, falls outside the limits of the geometrical beam, i.e. above the point o. Then the particles a and b will differ in distance to p_2 by a half wave length. Hence the vibrations produced at p_2 by these two particles mutually neutralize each other. Similarly the disturbance originating in the first particle below a will at p_2 be just one half wave length ahead of the disturbance coming from the first particle below b. Thus every particle between a and b may be paired off with a corresponding particle between b and c such that the effects of the two particles neutralize each other at p_2. Hence the total effect at p_2 of the disturbances coming from the portion ab of the opening is completely neutralized by the effect of the disturbances coming from the portion bc of the opening.

Consider next a point p_4 which is so situated that the distance cp_4 is two wave lengths more than the distance ap_4. The opening ac may now be divided into four parts, ae, eb, bf, fc, such that eb neutralizes at p_4 the effect of ae, since ep_4 is one half wave length more than ap_4, and fc neutralizes the effect of bf, since fp_4 is one half wave length more than bp_4. There is therefore no disturbance at all at p_4.

At some point p_3, between p_2 and p_4, the distance cp_3 will be one and a half wave lengths more than ap_3. If we now divide ac into three equal parts, the effect of the upper third will be completely neutralized at p_3 by that of the next lower third, but the effect of the lowest third has nothing to neutralize it at p_3; hence there is a disturbance at p_3 which is due simply to one third of the particles between a and c, and even the effects of the particles in this third partially neutralize one another at p_3, since they differ somewhat in phase. It is obvious that between p_2 and p_4 the disturbance increases from zero at p_2 to a maximum at p_3, and then falls gradually to zero at p_4; that, further, there are other points of zero disturbance, p_6, etc., so situated that the distance from c to the point in question is any even number of half wave lengths more than the distance from a to this point; and that between these points of zero disturbance are points of maximum disturbance, p_5, etc., so situated that the distance from c to the point in question is any odd number of half wave lengths more than the distance from a to this point. But it will also be noticed that the successive maxima, p_3, p_5, etc., diminish rapidly in intensity, since, while but two thirds of the particles between a and c completely neutralize one another's effects at p_3, four fifths of these particles neutralize one another's effects at p_5, six sevenths at p_7, etc. Hence it is not necessary to go a great distance above o in order to reach a region in which there are no points at which there is any appreciable disturbance. Further, if we consider wave lengths which are shorter and shorter in comparison with ac, the points of maximum disturbance p_3, p_5, etc., draw closer and closer together, and soon some of them begin to fall inside the limits of the geometrical beam, i.e. below the point o. Hence those that are left above o are weaker and weaker members of the series. It follows, therefore, that when the wave length becomes very short in comparison with ac, the disturbance will have become practically zero at a very short distance above the point o. In other words, a wave motion should be propagated in straight lines through an opening, or past an obstacle, and should not bend around appreciably into the region of the geometrical shadow, when and only when the wave length is very minute in comparison with the size of the opening; for in this case the disturbances from the various elements of the opening must interfere in such a way as completely to destroy one another at practically all points outside the limits of the geometrical beam. The analysis of the conditions which exist inside the limits of the geometrical beam when p_2, p_4, etc., fall below o will not here be taken up, since we are not concerned at this point with showing what happens inside of or so much as with proving that practically nothing happens outside of or. Suffice it to say that experiment and theory both show that under the conditions assumed there is practically uniform disturbance within the region or.

Now since ordinary sound waves have a wave length of from 1 to 8 feet, it will be seen from the above analysis that in passing through a window or any ordinary opening they may be expected to spread out in all directions beyond the opening, as in fact we know that they do. Indeed, if the aperture is less than one wave length in width, it should be impossible to find any point of quiescence whatever on the side of the screen which is away from the source. It is clear, then, that we must use extremely short sound waves, if we are to hope to observe with any ordinary openings the diffraction phenomena presented in the above theory.

In light, however, since the average wave length is only about .00005 cm, we should expect that with ordinary openings the maxima p_3, p_5, etc., would lie so near to the edges o and r of the geometrical beam as not to be easily discernible, or indeed inside the geometrical beam. In order, then, to bring them into evidence at all, we should expect to be obliged to work with exceedingly small openings.

Diffraction phenomena of the type discussed in this section, in which a diffracting screen is placed between the source and the observing screen, but in which no lenses or mirrors are used is known as Fresnel diffraction. On the other hand, if a diffracting screen is placed in the path of a plane wave of incident light and then the diffracted waves observed by means of a telescope focused at infinity the diffraction pattern formed is known as a Fraunhofer diffraction pattern. Both types of diffraction, however, arise from the same basic phenomenon and can both be treated by means of Huygen's principle.

The Nature of the Real Image of a Point Source

Let us consider as in Fig. 13 a long narrow slit of width ac upon which light is incident from a point source a great distance away so that we may consider the incident light plane parallel light. A converging lens L

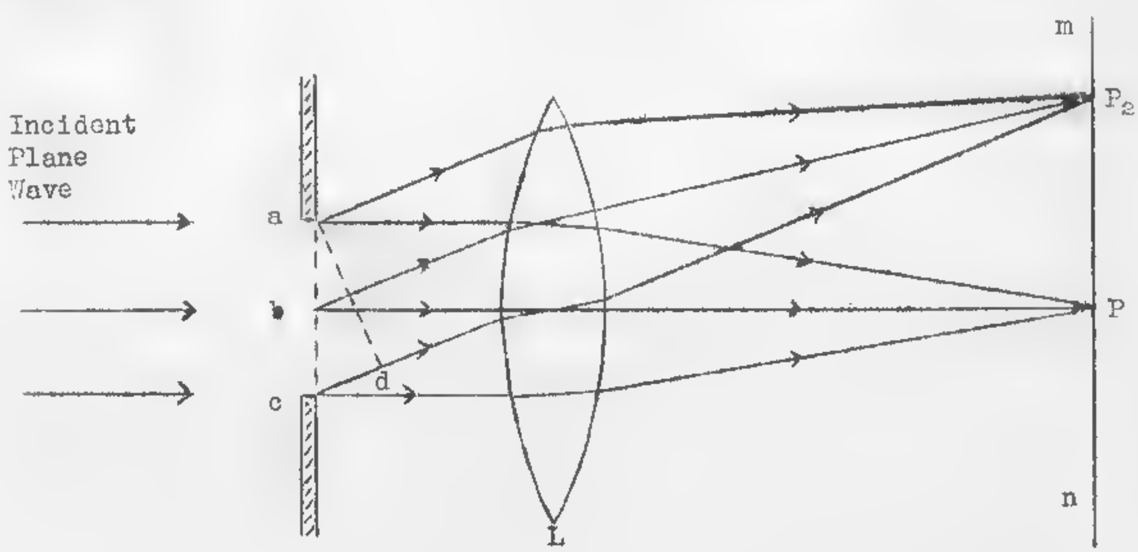


Fig. 13

is placed as shown so that plane parallel light is brought to focus at the point P on the screen. We will therefore direct our attention to that light which emerges from the slit ac as plane parallel light and which is propagated in a given direction. Let us consider first the light emerging from the slit in a direction parallel to the incident light. All the particles along the line ac are in the same phase of vibration, and since the optical path lengths from each point on ac, through the lens L to the point P on the screen, are equal, all the rays emerging from the slit parallel to the axis of the lens will reinforce one another at P, and P will be a point of constructive interference and appear bright on the screen.

But we expect light to emerge from the slit in other directions as well, for the line ac is a wave front of the incident light, and therefore by Huygens' principle each point on ac represents a new source of disturbance from which light is radiated spherically in all directions. Consider that direction perpendicular to the line ad in the figure where the distance cd is made equal to one wave length λ. Light having this direction would be brought to focus at some point P_2. By similar reasoning to that employed in the previous section it is apparent that the disturbance originating at the point mid-point b has a path to travel which is just one-half wave length less than that of the disturbance originating at C, before reaching P_2. Hence the disturbance from b and c will cancel one another at P_2. And by reasoning similar to that employed in the previous section it is clear that for every point between a and b there is a point between b and c such that the two disturbances from these points cancel one another at P_2. The interference at P_2 therefore will be completely destructive and no light will appear at P_2. This will be true also for other points P_3, P_4, P_5, ... etc. on the screen corresponding to a distance cd equal to any integral number of wave lengths. For points on the screen in between these points complete destructive interference will not occur and a series of bright and dark bands will appear, with the bright bands weaker in intensity the farther they are from the central point P.

The pattern on the screen, however, is in reality an image of the remote point source in which the light has its origin. We see then that the image of a point is not in reality a point, but consists of a series of maxima and minima. Similar reasoning shows that this is true even if the point source is not far removed from the slit, i.e., even if the incident light is not plane parallel light.

In the optical system we are here considering, a slit has been placed over the lens whose width is small compared to the diameter of the lens. The diffraction pattern therefore will closely approximate a series of straight and parallel bright and dark bands.

In Fig. 13, if we designate the two equal angles dac and PbP_2 by θ, and the width of the slit by d, we see that

$$\sin \theta = \lambda/d$$

where θ is the angle measured from the central direction out to the first minimum or the first dark band. Likewise the angle out to the nth dark band is given by

$$\sin \theta = n\lambda/d \qquad (9)$$

where n is any integer.

For simplicity the lens in Fig. 13 was taken to be fixed. This is permissable if θ is a small angle, i.e., if the width of the slit is considerably greater than one wave length. From the reasoning employed in this section it is clear that Eq. (9) is not restricted to small angles. A more adequate arrangement of the apparatus to demonstrate the effect at larger angles would be to rotate the lens about a line through the midpoint of the slit so that it would be perpendicular to the direction of the particular rays under investigation. Also the screen should then be in the form of a spherical surface whose center is at b in order that a fixed distance from the lens to the screen is maintained.

If now we imagine the slit to be widened, i.e., the distance ac is increased, the angles out to any of the diffraction bands, say the nth band, would decrease, and for slits whose width is large compared to λ, we can use the approximation $\sin\theta \approx \theta$, and Eq. (9) becomes

$$\theta = \frac{n\lambda}{d} \qquad (10)$$

If in place of a long narrow slit an aperture which has an opening of circular cross-section is placed in front of the lens it is clear from symmetry that the diffraction pattern will consist of a circular bright disc surrounded by a set of concentric dark and bright rings. A calculation of the position of the successive maxima and minima for a circular aperture is considerably more difficult than that for a long and narrow slit and will not be attempted here. The result, however, is similar, and the angle subtended by the first dark ring at the center of the aperature is given by

$$\theta = 1.22\,\frac{\lambda}{d} \qquad (11)$$

if the diameter d of the aperture is large compared with λ.

Even if no aperture were placed in front of the lens, the boundaries of the lens itself correspond to an aperture of finite size, and therefore the image of a point source produced by any lens will be a diffraction pattern and not a simple point.

This phenomenon places a limitation on the clearness of the detail which can be brought out in an image produced by any optical system. For example, consider two point sources $\underline{S}_1$ and $\underline{S}_2$ which subtend a small angle α at the center of a lens, or at the center of a circular aperture if one is placed in front of the lens, as shown in Fig. 14. Let $\underline{P}_1$ and $\underline{P}_2$ represent the centers of the diffraction patterns which are, respectively, the images of the two point sources $\underline{S}_1$ and $\underline{S}_2$. Obviously if the two diffraction patterns $\underline{P}_1$ and $\underline{P}_2$ are sufficiently close to one another there may be overlapping to such an extent that they would appear as the image of a single point instead of two. Experience has shown that if two diffraction patterns are superposed and spaced such that the central bright discs of one falls at the center of the first dark band of the other, then they are just distinguishable as two separate patterns, and they are then said to be resolved. An angle α which corresponds to a spacing between images of this amount is known as the limit of resolution of a lens, and by Eq. 11 is equal to $1.22\,\frac{\lambda}{d}$, where $\underline{d}$ is the diameter of the aperture of the lens. In some applications, as for example, of microscopes where a very high magnification is desired, and therefore making necessary a high resolving power, ultraviolet light is used because of the smaller value of λ for ultraviolet as compared with visible light.

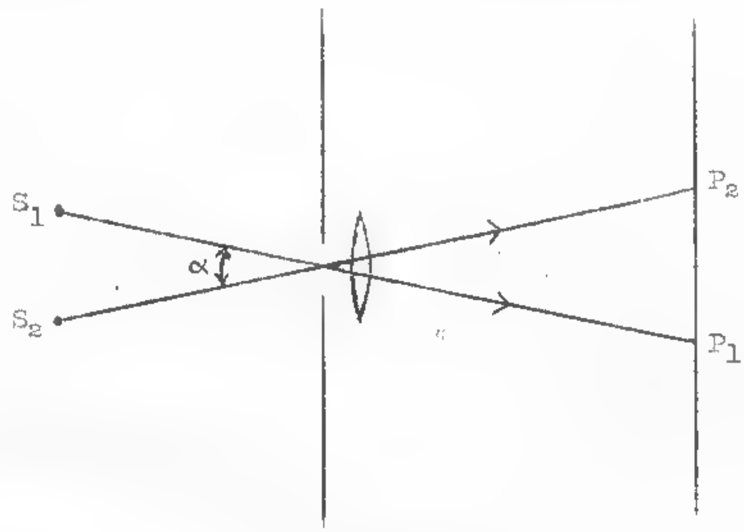


Fig. 14

6. <u>The Diffraction Grating</u>. The phenomena of diffraction are most strikingly exhibited with the aid of an instrument devised in 1821 by the celebrated German optician Fraunhofer, and known as the <u>diffraction grating</u>. Such a grating consists essentially of an opaque screen in which are placed at regular intervals small parallel slits for the transmission or reflection of light. Thus in Fig. 15 <u>qq'</u> represents a cross section of the grating, and the openings <u>m</u>, <u>p</u>, <u>s</u>, etc, represent the slits, which are thought of as extending at right angles to the plane of the page. For the sake of convenience in analysis these slits will at first be regarded as exceedingly narrow in comparison with their distance apart, that is, each opening will be thought of as a mere line. Let the

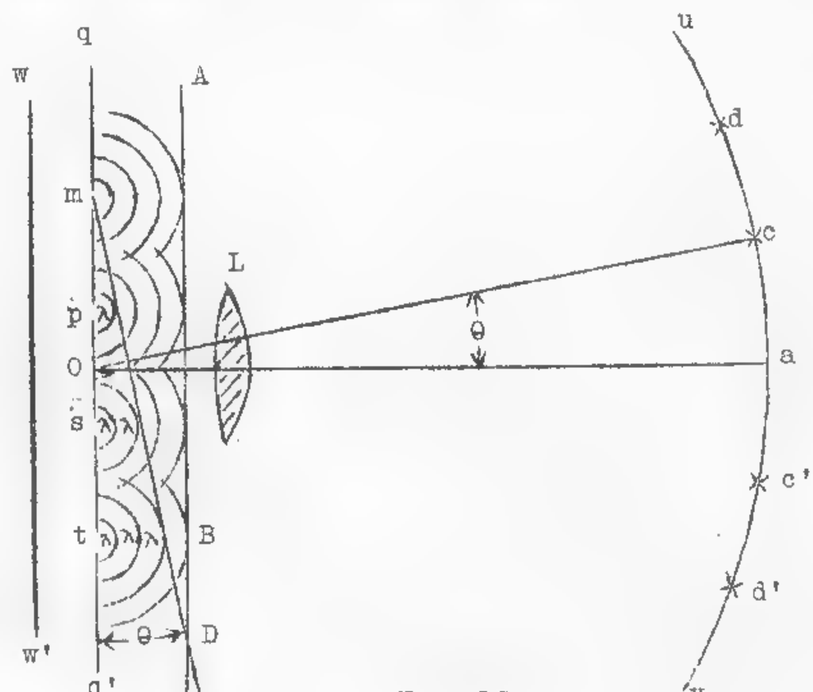


Fig. 15

source of light be so far distant that the wave surface <u>ww'</u> which falls upon the grating is practically plane. If no grating were present, a lens <u>L</u> interposed in the path of the wave <u>ww'</u> would form an image of the distant source <u>S</u> at some point <u>a</u>, while at all other points on the screen <u>uv</u> there would be destructive interference and therefore total darkness. Let us see how this conclusion would be modified if we take out certain portions of the wave front <u>ww'</u> by means of the grating. When the wave <u>ww'</u> reaches the grating <u>qq'</u> the points <u>m</u>, <u>p</u>, <u>s</u> become new sources of spherical waves, and if we draw the envelope to all these waves after the disturbance has traveled a small distance forward, we shall obtain precisely the same surface AB which we should have had if the grating had not been present, the only difference being that the intensity of disturbance in the plane AB is much less than before, since now but a few points, namely m, p, s, etc., are sending out spherical waves to AB, while before all the points in qq' were so doing. The lens will take this plane wave AB, consisting of vibrations all of which are in the same phase, and bring it to a focus at a, so that an enfeebled image of the distant source S will be formed at this point. Thus far, then, the only effect of the grating has been to diminish the intensity of the image at a.

But AB is not now the only surface which can be drawn to the right of the grating so as to touch points all of which are in the same phase of vibration, for a surface mD, so taken that the distance from p to it is one wave length, that from s two wave lengths, and so on, satisfies this condition quite as well as does the surface AB. Hence mD may be regarded as another plane wave, which, after passage through the lens, will be brought to a focus at some point c in the line drawn perpendicular to mD through the center of the lens, in precisely the same way in which the plane wave AB was brought to a focus at a. Here, as before, we shall rotate the lens about an axis in the grating indicated in the figure by O so that the plane of the lens is parallel to the wave front mD. It follows, then, that an image of the source should be formed at c as well as at a. Precisely the same line of reasoning will show that another image of the distant source should be formed at c' as far below a as c is above it. But a, c and c' are not the only points at which images of the source will be formed, for it is possible to pass a plane through m such that the distance from p to this plane is 2λ instead of λ, that from s, 3λ, etc. It is obvious that all points in this plane will be in the same phase of vibration, and hence that the resulting plane wave will be brought to a focus at some point d on the perpendicular drawn from the plane through O, and at the same distance from O as are a and c.

Similarly, there will be other images whose direction from O is determined by the simple condition that the successive distances from the slits to the wave front differ by a whole multiple of a wave length; thus po = λ, 2λ, 3λ, 4λ, 5λ, etc. The first image, namely that at c, where this difference is one wave length, is called the image of the first order. Similarly, that at d is the image of the second order, and so on. In a word, then, a lens and grating disposed as in Fig. 15 should produce a whole series of images of any distant source of light. This means, of course, that under these conditions light waves will bend far around into the region of the geometrical shadow and be discernible at a large number of different points instead of simply at a.

These theoretical deductions from the wave theory of light are completely confirmed by experiment. Furthermore, the experiments illustrating them are so simple and so much a part of everyday experience that the wonder is that they escaped detection and explanation for so long a time. Thus if one looks through a handkerchief held close to the eye at a distant arc light, gas flame, or bright star, one can always see nine and sometimes as many as eighteen or more images of the light. These are due to the two sets of gratings formed by the two sets of threads which run at right angles to each other. It is usually possible to see as many as three distinct images by simply squinting at a distant light through the eyelashes which act in this case as a very imperfect grating. In these experiments the retina of the eye takes the place of the screen uv, and the lens of the eye the place of L.

In its simplest practical form the grating consists of a plane piece of glass upon which are ruled with a diamond point, say, a thousand lines to the centimeter. The grooves cut by the diamond point constitute the opaque spaces in the grating, for the light which falls upon these grooves is scattered in all directions, so that a negligible part of it passes through in the direction in which the light is traveling. The clear glass between the rulings corresponds to the openings m, p, s in the screen of Fig. 15. If such a grating is held immediately before the eye and a source of monochromatic light viewed through it, the series of images formed at a, c, d, etc. on the retina are apprehended by the observer as a series of images of the source lying in the prolongations of the lines ao, co, do, etc. The images may be thrown upon a screen, if screen, lens, grating, and source are given the relative positions shown in Fig. 15.

In Fig. 15 if the angle between the grating and the direction of the wave front mD which forms the image of the first order is denoted by θ, and if d is the distance between successive openings -- the grating space or grating constant, as it is called, then the obvious relation exists,

$$\sin \theta = \lambda/d$$

In general, if n represents the order of any image and θ the angle between the grating and the wave front forming that image, then

$$\sin \theta = n\lambda/d \qquad (12)$$

Eq.(12) is known as the grating equation.

The Grating Spectrum

If the source sends out, not monochromatic light, but, instead, white light, the series of sharply defined images of the source is found to be replaced by a single central image of the source in white light at s, bordered on either side by broad bands of colored light. In the first band the end farther from the source is red. From the red the color grades into orange, yellow, green, blue-green, blue, and finally into violet at the end nearer to the source. The band of light thus produced is called a spectrum, and the phenomenon of its production is known as dispersion.

The action of the grating in producing dispersion is then easily seen; for since the position of every image except the central one is determined by the condition $\sin\theta = n\lambda/d$, it is evident that there is a different value of θ corresponding to each value of λ. Now the wave lengths which compose white light vary from about .000076 in the red to about .000039 in the violet. Hence when the source is white light the image of each order as it appears with monochromatic light is replaced by a series of adjacent images in different colors, each image corresponding to a particular wave length or color. This series of adjacent images constitutes the colored band or spectrum of each particular order. The central image is white and sharply defined because the wave front AB (Fig. 15) which gives rise to this image is at the same distance from each of the openings, and in consequence this wave front is the same for all wave lengths.

The spectrum of the first order is the only pure spectrum which a grating can produce, for it can be shown that the spectra of higher orders overlap. Thus, since for the red of the second order $\sin\theta = 2 \times .00007/d$, approximately, and since for the violet of the third order $\sin\theta' = 3 \times .00004/d$, it will be seen that $\sin\theta$ is greater than $\sin\theta'$, and hence that a part of the third violet overlaps a part of the second red. It is on account of this overlapping that one never sees more than two or three spectra on a side, for in the higher orders the overlapping is so complete as to reproduce white light.

The Dispersive Power of a Grating

It will be evident at once from Fig. 15 that the smaller the distance between openings, the farther apart will be the successive images a, c, d; in other words, that the angular separation of different orders produced by a grating, and hence, also, the angular separation of different colors in the same order, increases as the distance between the lines of the grating decreases. The ratio of the angular separation of two beams of different wave lengths to the difference in their wave lengths is known as the dispersive power of a grating. Thus the dispersive power is equal to $\frac{d\theta}{d\lambda}$, and by a simple differentiation of Eq. (12), we obtain

$$\cos\theta \, d\theta = \frac{n\,d\lambda}{d}$$

or

$$\frac{d\theta}{d\lambda} = \frac{n}{d\cos\theta} \tag{13}$$

The dispersive power of a grating, therefore, increases with the order of the spectrum, and is greater the smaller the grating constant. A grating which contains a large number of lines per centimeter has a greater dispersive power than one with a lesser number of lines per centimeter.

The Resolving Power of a Grating

A second important characteristic of a grating is its ability to produce images which are sufficiently sharp so that the images formed for two spectral lines of almost equal wave length are clearly distinct from one another. From the theory of Sec. 5 we must expect that the image of each spectral line will be a diffraction pattern with a bright central band and a series of alternating dark and bright bands on either side. This, of course, would be true even though the light represented by the spectral line were itself strictly monochromatic. We shall here take as the limit of resolution of the images of two spectral lines a criterion similar to that employed in Sec. 5; i.e., two images are said just to be resolved if the central bright maximum of one coincides with the first dark band of the other.

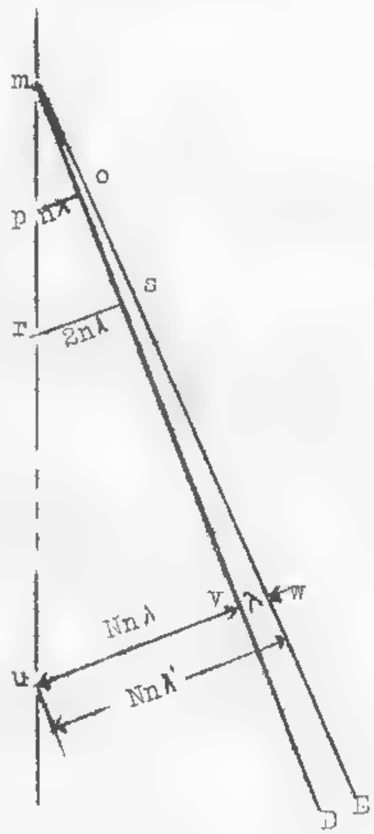


Fig. 16

In Fig. 16 are represented a grating and the wave fronts in the nth order spectrum corresponding to two spectral lines of wave length λ and λ' respectively. The first opening is represented by m, the second by p, etc. The distance po will therefore be $n\lambda$ for one spectral line and $n\lambda'$ for the second line. If the grating contains a total number of lines N, and if u represents the last opening, then the distance from the opening u to the two wave fronts will be $Nn\lambda$ and $Nn\lambda'$ respectively. The first dark band of one image will coincide with the central bright band of the other if the wave fronts have the relative positions shown in the figure, i.e., if vw is equal to λ. For with reference to the wave front mE and to the light of wave length λ, the disturbance from u will be just one-half wave length out of phase with the disturbance from that opening just midway between the first and last openings, and therefore the two disturbances will cancel one another's effects. Similarly the disturbance from each opening in the upper half of the grating will be cancelled by the disturbance for a corresponding opening in the lower half. Thus the wave front <u>mE</u> corresponds both to the first dark band for the light of wave length λ, and to the central bright band for the light of wave length λ'.

From Fig. 16,

$$Nn\lambda + \lambda = Nn\lambda'$$

and setting $\lambda' = \lambda + \Delta\lambda$ we find

$$\lambda = Nn\Delta\lambda$$

$$\text{or} \quad \frac{\lambda}{\Delta\lambda} = Nn \qquad (14)$$

The magnitude $\lambda/\Delta\lambda$ is taken as the measure of the resolving power of a grating, and from Eq. (14) we see that the resolving power is proportional both to the order of the spectrum and the total number of lines.

<u>Effects of finite width of slit on series of images produced in monochromatic light.</u> In the use of a grating one often observes that the image of the third order, for example, will be brighter than that of the second, that of the fifth brighter than that of the fourth, etc. The cause of this lies in the finite width of the open spaces which have heretofore been considered to be mere lines.

In order to understand the effect of a finite width in the openings upon the relative brightness of the successive images, consider Fig. 17, in which is shown on a large scale a section of a portion of a practical grating, <u>mo</u>, <u>pr</u>, and <u>su</u> representing the finite openings, and <u>op</u>, <u>rs</u>, etc., the opaque spaces. The points <u>m</u>, <u>p</u>, and <u>s</u> correspond to the line openings of the preceding discussion (Fig. 15). The line mD represents the wave front which gives the image of the first order. It is drawn as the envelope to the spherical waves due to the particles m, p, and s. The disturbances produced by these particles at the points m, h, and k are all in the same phase of vibration, for each point differs from the next in its distance from its respective slit by a whole wave length. The disturbances due to these points will then reenforce each other at the image c formed by the lens (Fig. 15). Similarly, particles such as n, q, and t, which bear the same relation in position to m, p, and s respectively, will also produce disturbances on the line mD at points which differ successively in their distances from these particles by a whole wave length. The disturbances due to these points will therefore reenforce each other at c. And similarly, for all the other points in the opening mo there will be corresponding points in the other openings which will reenforce these vibrations. But it is now to be noticed that the disturbances which start from the different parts m, n, and o of the same opening are not in quite the same phase of vibration when they reach the plane mD, and further that the wider the openings the greater becomes this phase difference. Suppose, then, that the open spaces mo, etc., are just equal to the opaque spaces op, etc., and that we are considering the image of the second order. We have seen that the condition which must hold

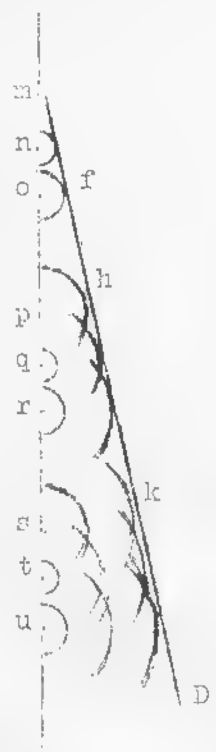


Fig. 17.

for this image is that $\underline{ph} = 2\lambda$, $\underline{sk} - \underline{ph} = 2\lambda$, etc. Since, then, $\underline{mo} = \underline{op}$, we have $\underline{of} = \lambda$. But when this condition exists the disturbance from $\underline{n}$ is one half wave length behind that from $\underline{m}$, and therefore completely destroys it at $\underline{c}$. Similarly, the disturbances from all the points between $\underline{n}$ and $\underline{o}$ destroy at $\underline{c}$ the disturbances from the points between $\underline{m}$ and $\underline{n}$. Similarly for all the other openings. Thus the image of the second order will be entirely missing, and also, for exactly the same reason, the images of the fourth, sixth, etc., orders. If the opening is one third of the grating space instead of one half, the missing images will be those of the third, sixth, ninth, etc., orders. If conditions of this sort are only approximately fulfilled, as is usually the case, the images considered will be simply weakened but not entirely cut out.

Reflection gratings. By far the greater part of spectroscopic work is now done with the aid of gratings, but in actual work reflection gratings are much more common than transmission gratings like that which has been studied. These reflection gratings are made by ruling very fine lines on a reflecting metal surface, rather than on a transmitting glass surface. The grooves destroy the light, while the spaces between them reflect it regularly. The light from any white source which is reflected from such a grating and then brought to a focus by means of a lens shows a central white image at a position such that the angle of incidence equals the angle of reflection. On either side of this central image are found spectra of the first, second, and third orders, precisely as in the transmission grating. The theory of the two gratings is in all respects identical, for it obviously makes no difference how the lines $\underline{m}$, $\underline{p}$, $\underline{s}$, etc., of Fig. 15 become sources of disturbance, whether by reflecting or by transmitting a disturbance from some other source.

A form of grating which has rendered possible some of the most important of advances in spectroscopy is the concave grating invented by the late Professor Henry A. Rowland of Johns Hopkins University. The essential difference between this and other gratings is that the lines are ruled upon the surface of a concave spherical mirror of large radius of curvature, for example 20 feet. Under such conditions the mirror itself forms the series of images corresponding to a, c, d, etc. (see Fig. 14), so that it is not necessary to interpose a lens. This eliminates all difficulties arising from the absorption of the waves by the lens, difficulties which are especially pronounced in the ultra-violet and infra-red regions of the spectrum.

Problems

1. In Fig. 4 assume that S_1 and S_2 represent long narrow slits perpendicular to the plane of the figure. Take the separation of the slits to be 0.05 mm and the distance from the slits to the screen P to be 60 cm. Compute the distance apart of the central two dark bands on the screen for light of wave length 4500 A.

2. What thickness water film will produce a strong first order reflection of Na-D light which is incident normally on the film? What is the wave length of the light inside the water film?

3. A vertical soap film, observed by reflected light, becomes black at the top just before it breaks. Explain.

4. A thin wedge of air is formed between two narrow pieces of plate glass which touch at one end and are separated by a thin piece of paper at the other end. When the wedge is viewed normally by monochromatic reflected light an interference pattern of alternate light and dark bands is seen. If the distance apart of the dark bands is 2 mm when Na-D light is used, calculate the angle of the wedge.

5. If the air space in Problem 4 is now replaced by water and the wedge viewed in a similar manner, what will be the separation of adjacent dark bands in the interference pattern?

6. Newton's Rings are formed by a convex lens resting on a plane surface. The 20th dark ring is 1.5 cm from the center when Na-D light is used. What is the radius of curvature of the lens, and what is the thickness of the air film at that point?

7. If in Problem 6 the space beneath the lens is filled with water what will be the distance from the 20th dark ring to the center? What is the distance from the 15th dark ring?

8. A convex lens of radius of curvature 60 cm resting on a plane surface is illuminated with light consisting of two wave lengths $\lambda_1 = 5.4 \times 10^{-5}$ cm and $\lambda_2 = 4.5 \times 10^{-5}$ cm. The nth dark ring for λ_1 is seen to coincide with the $(n + 1)^{th}$ dark ring for λ_2. What is the diameter of the nth dark ring for λ_1, and what is the value of n?

9. A long slit 0.01 mm wide is illuminated normally by a plane wave of $\lambda_1 = 7.0 \times 10^{-5}$ cm and $\lambda_2 = 4.5 \times 10^{-5}$ cm. A lens of focal length 90 cm is placed immediately behind the slit and focused on a screen. Find the distance between the two innermost dark bands for λ_1 and λ_2 on the screen.

10. Mizar, the larger of the two stars at the bend of the handle of the Great Dipper, is a double star. Its two components are separated by 14.5" of arc. What is the smallest aperture of telescope that can resolve this doublet?

11. The objective lens of an astronomical telescope has a diameter of 40 cms and a focal length of 800 cm. What should be the focal length of the eye-piece if by visual observation it is desired to distinguish as two separate images the images of two stars which are just within the limit of resolution of the telescope objective? Assume an unaided eye is able just to distinguish two images when they subtend an angle of 2' at the eye.

12. Na - D light is incident normally on a grating ruled with 4000 lines per cm. What is the diffraction angle in the third order? How many orders are present?

13. In Problem 12 what is the dispersive power in each order?

14. If a grating is ruled with 2000 lines per cm how far must the rulings extend in order that the D-lines of Na (5890 and 5896 A) will be just resolved in the 2nd order? In the 3rd order?

15. The new Palomar telescope has a diameter of 200 inches. Estimate the smallest angular separation two stars can have and still be resolved, for λ = 4000 A and 6000 A.

16. Design a grating which will just resolve the two magnesium lines of 5184 and 5173 Angstroms, and which will produce a separation of one minute of arc in the diffracted beams, when used in the third order.

17. A grating has slits one half as wide as the grating space. How many orders will be present if the wave length is 5000 Angstroms and the grating is ruled with 2000 lines per cm?

18. Parallel white light is incident normally on a slit .05 cm wide placed directly in front of a lens of -1 meter focal length. In a screen placed at the focal plane there is a pin hole 0.3 cm from the axis of the system. What wave lengths between 3000 and 8000 Angstroms are missing in the spectrum of the light observed through the hole?

19. Deduce the grating formula which is applicable for light which strikes the grating at other than normal incidence.

20. A bi-prism (Fig. 18) is a piece of glass in the form of an isosceles triangular cylinder. A bi-prism with a 3' angle is placed in a parallel beam of light originating from a point source placed at the focus of a collimating lens. A screen S is placed at a distance of one meter from the plane of the prism. Describe the interference pattern produced at S assuming the light is Na D light and the prism is made of #123 crown glass. Calculate the separation between the interference bands.

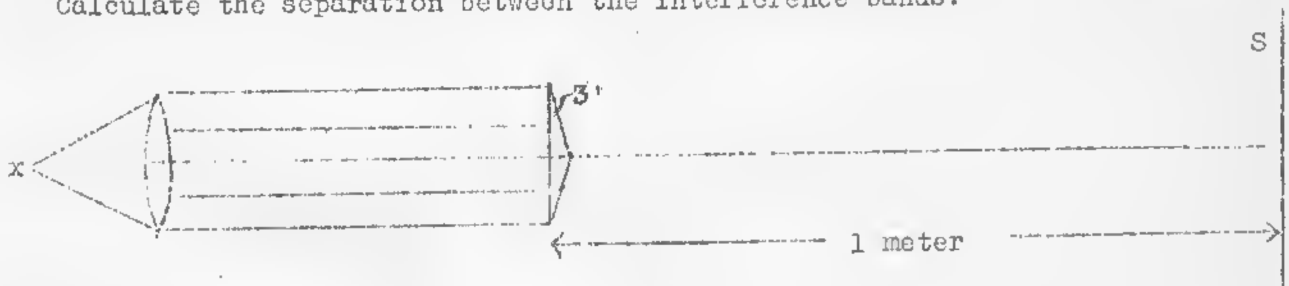


Fig. 18

21. Monochromatic light (λ = 5000 A) from a point source strikes a mirror at nearly grazing incidence. The reflected light then interferes with the direct beam from the source. The perpendicular distance from the source to the plane of the mirror is d = .1 mm. A screen S is placed at a distance D = 2 meters from the source. Let $d << D$. Calculate the separation of the fringes. Such an arrangement is known as a Fresnel mirror.

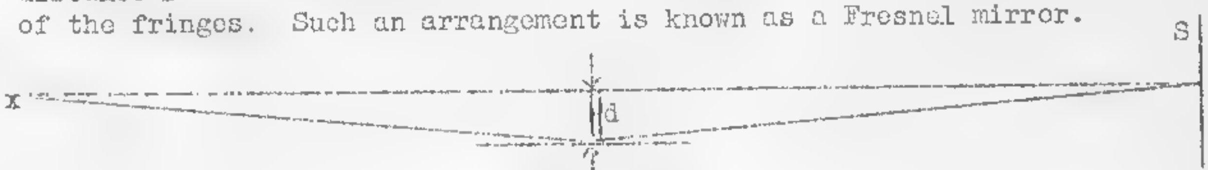


Fig. 19

22. A converging lens of -1 m focal length is sawed into two halves and the halves are separated by 1 mm. A monochromatic source A (λ = 5000 A) is placed at a distance of 2.5 m from the plane of the lens, forming two images, B_1 and B_2. A screen S is placed at a distance of 3 m from the lens. Describe the interference pattern and calculate the separation of the fringes. Assume that $\overline{B_1B_2} << \overline{SB_1}$.

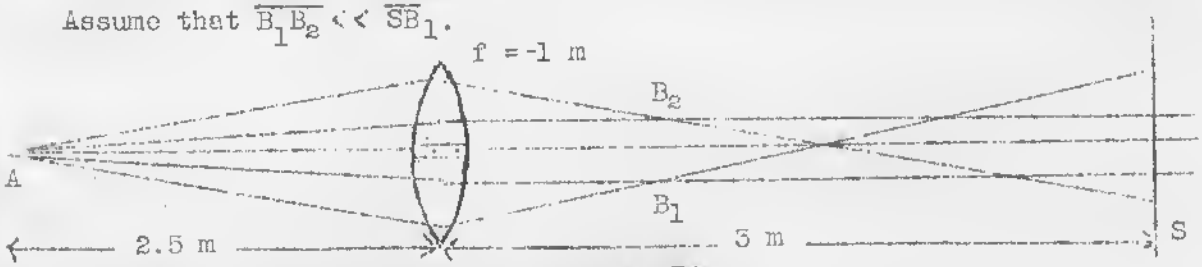


Fig. 20

CHAPTER FIVE

POLARIZED LIGHT

1. Polarization by Reflection. All of the phenomena of light which have been thus far studied have been found to be explicable upon the basis of the same wave theory which applies to the phenomena of sound. In other words, so far as the fundamental facts of reflection, refraction, diffraction, emission, and absorption are concerned, sound and light are identical in all respects except in the lengths of their waves and in the nature of the media which act as their carriers.

There is, however, a class of phenomena, known as the phenomena of polarization, which differentiate light completely from sound, and show that light waves are not compressional waves at all as are sound waves, but are instead transverse waves similar to those which elastic solids are able to propagate by virtue of their rigidity. These phenomena are so far removed from ordinary observation that they will be here presented in connection with a series of experiments. The facts presented in the first experiment were discovered in 1810 by the French physicist Malus (1775-1812).

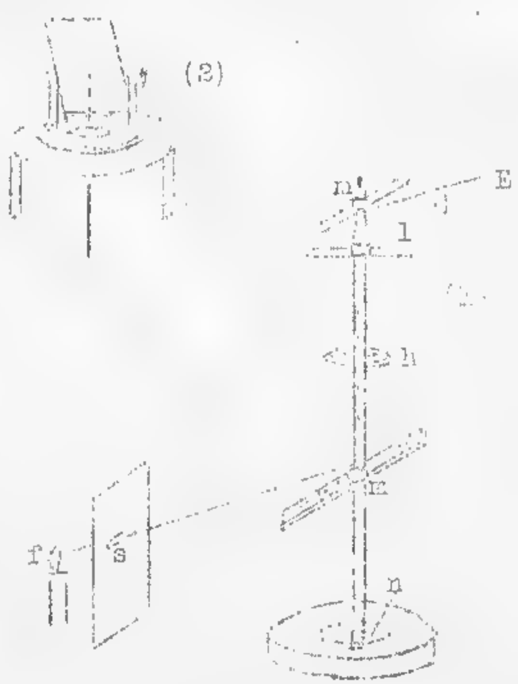


Fig. 1.

Experiment 1. Set the plane glass reflector m of the so-called Nörrenberg polariscope of Fig. 1

so that its plane makes an angle of about 33° with the vertical. Adjust the position of a horizontal slit s (about 5 mm wide) and a sodium flame f so that when you remove the black glass mirror m' and look vertically down upon the middle of m you see a portion of the flame. In this experiment the mirror n may be covered with a piece of black paper. Place m' in position and turn it so that it is exactly parallel to m, that is, so that its plane also makes an angle of 33° with the vertical. Place the eye at E in such a position that when you look at the middle of m you see the twice-reflected image of the sodium flame. Then rotate m' in its frame about a vertical axis and observe the image of the flame as you do so. When you have turned m' through 90°, that is, into the position shown in Fig. 1, 2, the image of the flame will have completely disappeared.

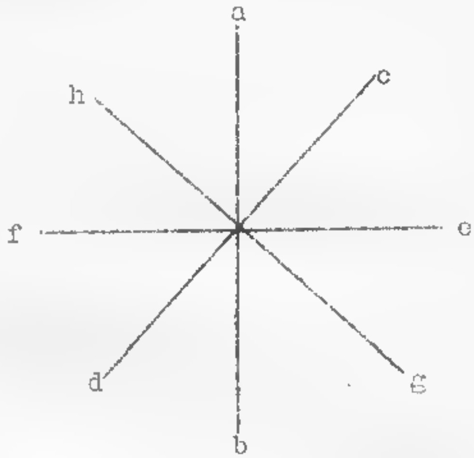


Fig. 2.

The experiment shows that light waves cannot be longitudinal, for if the particles of the medium which transmits the light from m to m' vibrated in the direction of propagation of the light, then the conditions of symmetry would demand that the wave be reflected in precisely the same way after m' has been rotated through 90° as before. But we can understand the experiment if we assume that light consists of waves in which the direction of vibration of the particles of the medium is always transverse to the direction of propagation of the waves. We shall consider again that a beam of light is made up of a large number of elementary wave trains, and that a specified direction of vibration can be associated with each elementary wave train. We assume further that the directions of vibration associated with the elementary wave trains are oriented at random with respect to one another. Hence in Fig. 2 let ab, cd, ef, gh, etc. each represent the direction of vibration to be associated with a particular elementary wave train. If then the beam of light consists of a large number of elementary wave trains the diagram corresponding to Fig. 2 would consist of a large number of directions distributed symmetrically about the center point of the diagram. In other words a beam of natural or unpolarized light has a constant amplitude of vibration when measured in any direction perpendicular to the direction of propagation of the light.

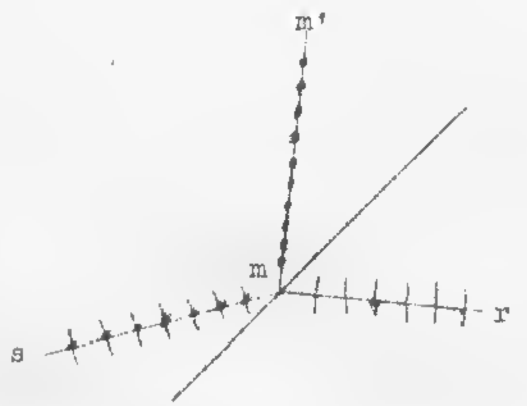


Fig. 3

All the vibrations of unpolarized light can be resolved into two component vibrations, the one perpendicular to the plane of incidence smm' (Fig. 3), that is, at right angles to the plane of the page, and represented by the dots in the line sm, and the other in the plane of incidence and represented by the straight lines drawn across the path of the ray sm.

When the ray sm strikes the mirror there is one angle of incidence such that the refracted ray mr and the reflected ray mm' are at right angles to each other. Now when this condition is fulfilled experiment shows that then the reflected ray consists only of vibrations which are perpendicular to the plane of incidence. The ray mm' is said to be a ray of plane polarized light, and the angle of incidence at which the ray sm must fall upon the mirror in order that the reflected ray mm' may consist only of vibrations in this one plane is called the polarizing angle. That this angle is always the angle for which the reflected and refracted are at right angles was discovered in 1815 by Sir David Brewster (1781-1868), and is known as Brewster's law. It may easily be shown that another form of statement of the same law is the following: <u>the angle of complete polarization is the angle the tangent of which is the index of refraction of the reflecting substance</u>. This is the form in which Brewster announced his law. That the two forms of statement represent one and the same physical relation may be seen from the following:

If the angle <u>cod</u> (Fig. 4) is 90°, then we have $90^\circ - i + 90^\circ - r = 90^\circ$ or $i + r = 90$; hence $\sin r = \cos i$; hence the index μ (= sin i/sin r) may be written in the form

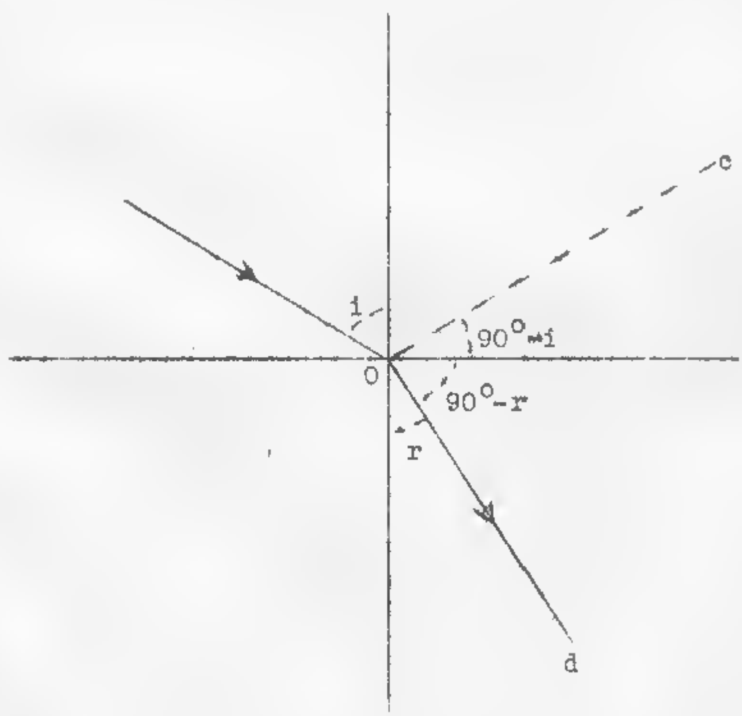


Fig. 4

$$\mu = \frac{\sin i}{\cos i} = \tan i,$$

or

$$i = \tan^{-1} \mu. \qquad (1)$$

The reason that we originally set the mirror <u>m</u> so as to make an angle of 33° with the vertical was that the index of refraction of crown glass is about 1.55, and the angle the tangent of which is 1.55 is 57°. In order that the angle of incidence might be 57° it was necessary to make the angle between the plane of the mirror and the vertical 33°.

It will now be obvious why we obtained no reflected light at all from m' when we had rotated it from position 1 to position 2 (Fig. 1). For in this latter position m' bore precisely the same relation to the vibration of the ray mm' as did the mirror m to the component of sm which was vibrating in the plane of incidence smm'.

Light is said to be plane polarized if the vibrations associated with it all lie in one plane, and this plane is known as the plane of polarization.

Experiment 2. Now set m and m' again so that there is no light from f reflected at m' and then rotate m' about a horizontal axis, observing the middle of m' all the while. You will find that there is always some of the light ray mm' reflected from m' except when m' is set exactly at the polarizing angle. The amount of the light thus reflected will be found to increase rapidly as the position of the mirror departs in either direction from the polarizing angle.

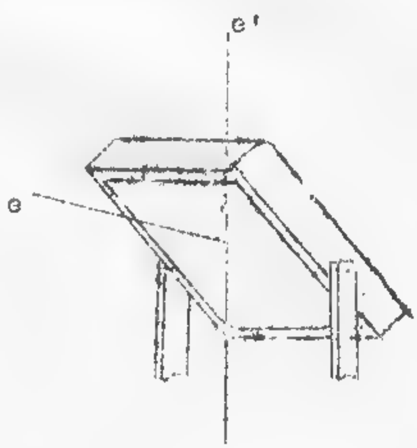


Fig. 5

Experiment 3. Replace the glass mirror m' by a pile of about fifteen thin glass plates set at the polarizing angle (see Fig. 5), and then observe not only, as above, the reflected ray e, but also the ray e', transmitted by the plates as the pile is turned about a vertical axis. You will find that when the pile of plates is in position 2 (Fig. 1), that is, in the position such that the reflected ray disappears, the transmitted ray is of maximum brightness, and when the plates are rotated into position 1 (Fig. 1), that is, into a position such that the reflected ray is of maximum brightness, the transmitted ray has almost entirely disappeared.

In explanation of these effects consider that an incident beam sm (Fig. 6) of ordinary light is resolved into two components, one vibrating in, and one normal to, the plane of incidence. Let the intensity of each of these components be represented by 50 (see Fig. 6). At the polarizing angle none of the 50 parts which vibrate in the plane of incidence are reflected, while photometric measurements show that about 16 per cent of the light which is vibrating perpendicular to the plane of incidence is reflected; that is, 8 per cent of the incident beam is reflected at the polarizing angle. Hence, after the first refraction, the transmitted light consists of 50 parts vibrating in the plane of incidence and 42 parts vibrating in the plane perpendicular to the plane of incidence. After the second refraction these numbers have become 50 and 35.3; after the third refraction, 50 and 29.7; after the fourth, 50 and 25, and so on. After passage through twelve or thirteen plates the transmitted light has become nearly plane polarized by this process, the plane of its vibrations obviously being at right angles to the plane of vibration of the reflected light. A pile of plates of this sort furnishes a very inexpensive means of obtaining plane polarized light, but it suffers from the disadvantage that the polarization is not quite

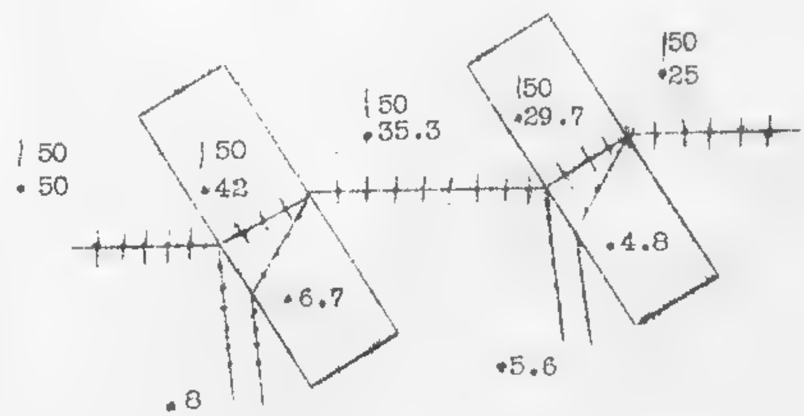


Fig. 6

complete. If no light whatever were absorbed or scattered by the glass, the transmitted ray would become more and more nearly plane polarized the larger the number of plates, but in practice there is found to be no advantage in increasing the number of plates beyond thirteen or fourteen.

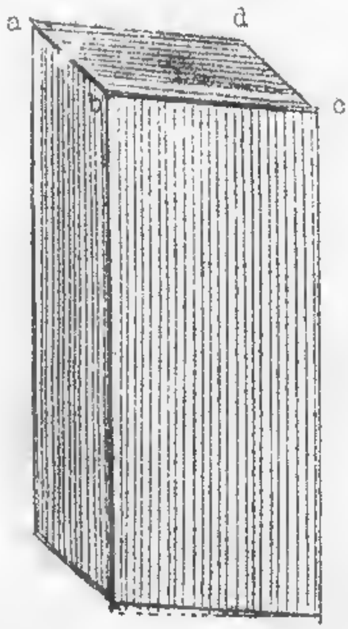


Fig. 7

Experiment 4. Replace the upper mirror m' by a Nicol prism (Fig. 7), the construction of which will be considered later, and looking down through this Nicol at the image of f reflected in m, rotate the Nicol about a vertical axis. You will find that the ray mm' is cut off completely by the Nicol when the latter is in a certain position, but that the light from the flame is transmitted with maximum brightness when the Nicol has been rotated through an angle of 90° from this position. From a knowledge of the plane of vibration of the ray mm' (Fig. 1) decide what must be the plane of vibration of a ray with respect to the face abcd of the Nicol, in order that it may be wholly transmitted by the latter, and mark the direction of this transmitting plane of the Nicol by an arrow drawn on the face of mounting containing the Nicol. Henceforth you can use the Nicol as a detector of the plane of vibration of any polarized light which you may observe.

2. Polarization by Double Refraction. The phenomena which will be presented in the following experiments were discovered in 1670 by the Danish physicist Erasmus Bartholinus (1625-1698), who first noticed the fact of double refraction in Iceland spar, and by Huygens (1629-1695) in 1690, who first noticed the polarization of the doubly refracted beams produced by the Iceland spar, and first offered an explanation of double refraction from the standpoint of the wave theory.

Experiment 5. Make a pinhole in a piece of black cardboard, and lay the cardboard on a piece of plane glass on the frame h (Fig. 1). Some inches beneath this, for example on the plate m turned into the horizontal position, lay a piece of white paper and illuminate it well. Then lay a crystal of

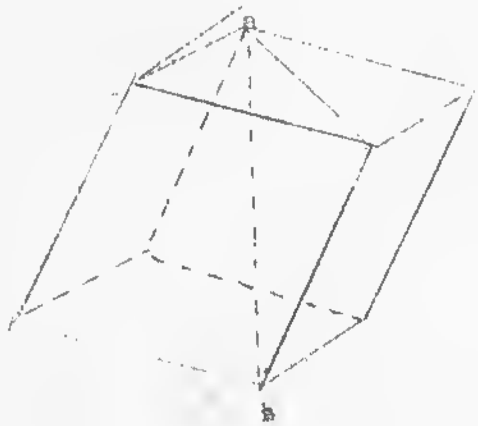


Fig. 8

Iceland spar (Fig. 8) over the hole in the cardboard. Remove m' and look vertically down upon the crystal. You will see two holes instead of one. Rotate the crystal about a vertical axis. One image will remain stationary, while the other will rotate about it.

That the image which remains stationary is produced by light which has followed the usual laws of refraction is evident from the fact that it behaves in all respects as it would if viewed through a glass plate. The image which rotates, however, must be produced by light which has followed some extraordinary law of refraction; for although it has passed into the crystal in a direction normal to the bottom face, and out of it in a direction normal to the top face, it must have suffered bending inside the crystal, since it emerges from the crystal at a point different from that at which the other ray emerges. We must conclude, then, that a ray of light which is incident upon the lower face of such a crystal of Iceland spar is split into two rays by the spar, and that these two rays travel in different directions through the crystal. The ray which follows the ordinary law of refraction is called the <u>ordinary</u> ray, the other the <u>extraordinary</u> ray.

<u>Experiment 6</u>. To find the direction in which the extraordinary ray travels, rotate the crystal about a vertical axis above the pin hole and note that the extraordinary image always lies in the line connecting the ordinary image and the solid obtuse angle of the face which is being viewed, and, further, that the extraordinary image is always on that side of the ordinary which is away from this solid obtuse angle.

It will be evident, then, from these experiments that if Fig. 9 represents a section of the crystal made by passing a plane normal to the top

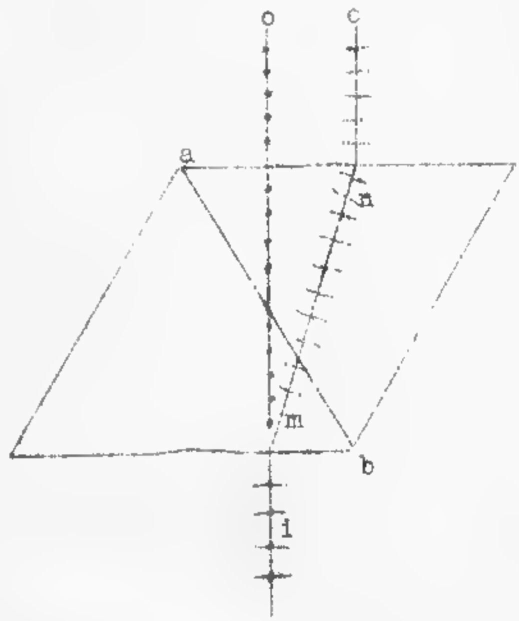


Fig. 9

and bottom faces and through the vertices of the two solid obtuse angles of a crystal all of whose sides are equal, the line io will represent the path of the ordinary ray through the rhomb, while the broken line imno will represent the path of the extraordinary ray.

Experiment 7. In order to determine whether the ordinary or the extraordiny ray travels the faster through the rhomb, observe again the two pin holes, or, better, observe at close range a dot on a piece of white paper upon which the crystal lies, and note which image, the ordinary or the extraordinary, appears to be the nearer to the upper face.

This will evidently correspond to the ray which has suffered the largest change of velocity in emerging into the air; that is, it will correspond to the ray which travels more slowly in the crystal. This will be found to be the ordinary ray.

Experiment 8. If you can obtain a crystal which has been cut so that its top and bottom faces are planes which are at right angles to the line ab (Fig. 8) which connects the two obtuse angles of a perfect rhomb, that is, a rhombohedron having all of its faces equal, view the pin hole normally through this crystal. You will observe that there is now but one ray, and that this ray does not change position upon rotation; that is, that it behaves in the ordinary way.

The direction of the line connecting the two obtuse solid angles of a crystal all of whose sides are equal is the optic axis of the crystal. This axis is not a line, but rather a direction. Any ray of light which passes through the crystal in a direction parallel to the line ab (Fig. 8), that is, parallel to the optic axis, does not suffer double refraction.

Experiment 9. Place the Iceland spar again over the pin hole in the manner indicated in Experiment 5, and view the two images through the Nicol prism as the latter is rotated about a vertical axis. You will find that both the ordinary and the extraordinary images consist of plane polarized light, but that the planes of vibration of the waves which produce the two images are at right angles to one another.

Hence we may conclude that the Iceland spar has in some way separated the incident light into two sets of vibrations, one of which consists of all the components of the initial vibrations which were parallel to a particular plane in the crystal, while the other consists of all of the components of the initial vibrations which were perpendicular to this plane.

Experiment 10. With the aid of the Nicol, the transmitting plane of which you determined in Experiment 4, find whether the ordinary or the extraordinary ray consists of vibrations which are parallel to the plane which includes the optic axis of the crystal.

You will find that it is the extraordinary ray the vibrations of which are in this plane, while the vibrations of the ordinary ray are perpendicular to this plane (see Fig. 9).

3. Theory of Double Refraction. The foregoing experiments can be understood in terms of the optical anisotropy of Iceland spar. In crystalline substances many physical properties may be different in different directions. Therefore, since the velocity of propagation of a light wave in a crystal depends upon certain physical properties of the crystal, one would expect the velocity to have different values depending upon the orientation of the plane of vibration of the light wave with respect to directions fixed in the crystals. This is actually the case and gives rise to the many important and beautiful optical phenomena which occur when light passes through crystals.

We mentioned above in connection with Exp. 8 that a crystal may have a certain preferred direction which is known as the optic axis. In many crystals there is more than one such preferred direction, but we shall restrict ourselves to a discussion of the relatively simple case of uniaxial crystals, i.e., crystals which contain only one preferred direction. Iceland spar, sometimes called calcite ($CaCO_3$), is such a uniaxial crystal.

The effects described in Exps. 5 - 10 can be understood if it is assumed that waves whose vibrations are perpendicular to the optic axis travel with a velocity corresponding to an index of refraction μ_o, the so-called ordinary index of refraction. Hence the velocity in the crystal for waves of this type is c/μ_o, where c is the velocity of light in free space. On the other hand, waves whose vibrations are parallel to the optic axis travel with a velocity corresponding to an index of refraction μ_e, or the extraordinary index of refraction. For such waves the velocity in the crystal is equal to c/μ_e. Values of the refractive indexes for Iceland spar and quartz are given in Table I.

Table I

Refractive Indices for Different Colors

		Red(C)	Yellow(D)	Blue(F)
Iceland spar	Ordinary ray	1.6545	1.6585	1.6679
	Extraordinary ray	1.4846	1.4864	1.4908
Quartz	Ordinary ray	1.5418	1.5442	1.5496
	Extraordinary ray	1.5509	1.5533	1.5589

Consider a crystal cut with four of its sides parallel to the optic axis, and the other two sides perpendicular to the optic axis as indicated in Fig. 10. A coordinate system is chosen with its axes parallel to the edges of the crystal and with the y-axis parallel to the optic axis.

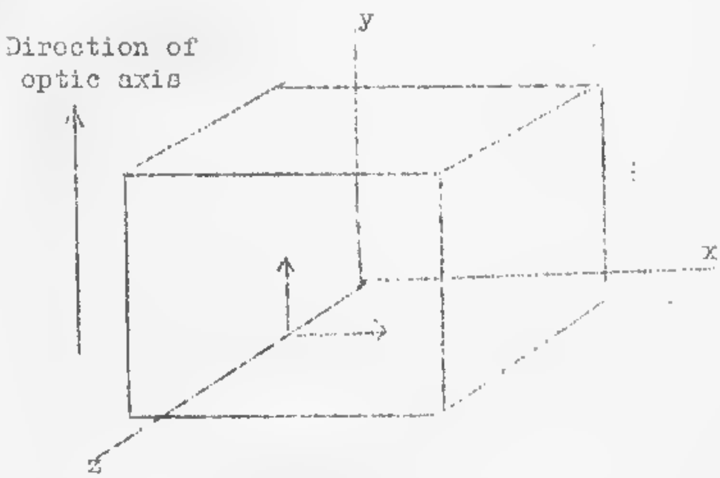


Fig. 10

Consider now a plane polarized wave traveling in the positive direction of the z-axis. If its plane of polarization is the xz plane the vibrations are perpendicular to the optic axis and μ_o is the index of refraction. If the plane of polarization of the wave is the yz plane the vibrations are parallel to the optic axis and the refractive index is μ_e. A plane polarized wave whose plane of polarization makes an angle ϕ with the xz plane may be resolved into two component waves, the one, whose plane of polarization is the xz plane, and the other whose plane of polarization is the yz plane. In Fig. 11 if the amplitude of the incident wave is A its displacement r at any instant of time t is given by

Fig. 11

$$r = A \sin \omega t \qquad (2)$$

and the displacements of the two component waves are given by

$$\begin{aligned} x &= A \cos \phi \sin \omega t \\ y &= A \sin \phi \sin \omega t \end{aligned} \qquad (3)$$

where $A \cos \phi$ and $A \sin \phi$ are their respective amplitudes. These two component waves which represent vibrations in phase as they enter the crystal will travel at different speeds corresponding to the two indexes of refraction and therefore will not in general be in phase when they emerge from the other side of the crystal. Let us consider three special cases.

In order to compute the phase difference between the two emerging waves it is necessary to calculate the distance by which one wave lags behind the other due to its slower speed in the crystal. Let h in Fig. 12 represent the distance the faster wave has traveled in the air after emerging from the crystal at the instant of time that the slower wave has just passed through the crystal. If d is the thickness of the crystal then the time t which has elapsed since the waves entered the crystal is given by,

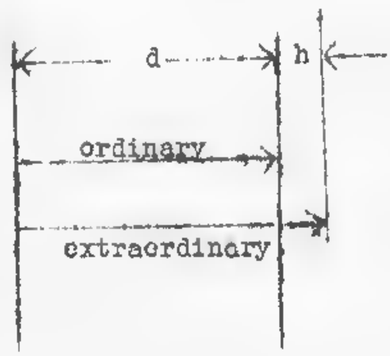


Fig. 12

$$t = \frac{d}{c/\mu_o} = \frac{d}{c/\mu_e} + \frac{h}{c} \qquad (4)$$

In Iceland spar μ_o is greater than μ_e and the ordinary wave is the slower of the two. Crystals in which $\mu_o > \mu_e$ are called negative unixial, and those in which $\mu_o < \mu_e$ are called positive uniaxial.

From Eq. (4) the distance h is given by

$$h = d(\mu_o - \mu_e) \qquad (5)$$

The two waves upon emerging will in general differ in phase. The phase difference* δ is equal to $2\pi h/\lambda$ where λ is the wave length in air. Hence the phase difference δ is given by

$$\delta = \frac{2\pi d}{\lambda}(\mu_o - \mu_e) \qquad (6)$$

If the Eqs. (3) represent the components of the wave incident on the crystal, the components of the emergent wave are represented by

* See MRW, p. 376, for a discussion of phase difference.

$$x = A \cos \phi \sin \omega t$$
$$y = A \sin \phi \sin (\omega t + \delta) \qquad (7)$$

The wave emerging from the crystal is then given by the combination of these two component waves, and the kind of polarization exhibited by the emergent light will depend upon the value of δ. In general the emergent light is elliptically polarized. We shall treat here only three special cases, i.e., where $h = \lambda/4$, $\lambda/2$ and λ, and the corresponding phase differences δ are $\pi/2$, π and 2π respectively. Crystals which produce phase differences of these amounts are known as quarter-wave, half-wave and full-wave plates, respectively.

Quarter-wave Plate

For light emerging from a quarter-wave plate the component waves are,

$$x = A \cos \phi \sin \omega t$$
$$y = A \sin \phi \sin (\omega t + \frac{\pi}{2}) = A \sin \phi \cos \omega t \qquad (8)$$

The form of polarization represented by Eqs. (8) is known as elliptical polarization. To obtain the path of the vibration in the medium we have merely to eliminate the parameter t from Eqs. (8). This is readily done by squaring and adding. Thus

$$\frac{x^2}{A^2 \cos^2 \phi} + \frac{y^2}{A^2 \sin^2 \phi} = 1 \qquad (9)$$

which is a common form for the equation of an ellipse. In the special case where $\phi = \pi/4$, $\cos \phi$ and $\sin \phi$ are both equal to $\sqrt{2}/2$ and Eq. (9) takes the form,

$$x^2 + y^2 = A^2/2 \qquad (10)$$

which represents circularly polarized light.

Example 1. How would you distinguish between circularly polarized light and unpolarized light?

Half-wave Plate

If the crystal has a thickness to correspond to a half-wave plate the emerging component waves are denoted by,

$$x = A \cos \phi \sin \omega t$$
$$y = A \sin \phi \sin (\omega t + \pi) = -A \sin \phi \sin \omega t \qquad (11)$$

Again to obtain the path of the vibration in the medium we eliminate t, and obtain

$$y/x = - \tan \phi \qquad (12)$$

which is the equation of a straight line whose direction is shown in Fig. 13. The emergent wave is then plane polarized, but its plane of polarization has been rotated through an angle 2ϕ. The amplitude of vibration of the emergent wave is given from Eqs. (11), by

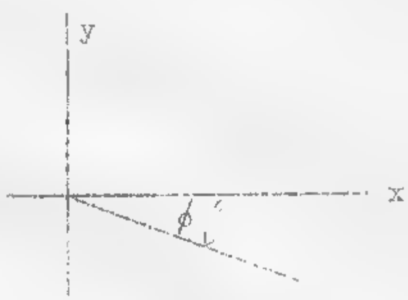


Fig. 13

$$r = \sqrt{x^2 + y^2} = A \sin \omega t$$

which is the amplitude of the incident wave, Eq. (2). In the special case where $\phi = \pi/4$, the plane of polarization is rotated through an angle of $\pi/2$.

Full-wave Plate

In the case of a full-wave plate the phase difference between the component waves upon emergence is 2π, and the equations representing the emergent light are identical with those for the incident light. Thus the emergent light is plane polarized, and the plane of polarization is the same as that of the incident light.

Plate of Any Arbitrary Thickness

In the general case where the phase difference of the emergent waves has the arbitrary value δ, the emergent light is elliptically polarized. The emergent waves are then represented by the following equations,

$$x = A \cos \phi \sin \omega t$$
$$y = A \sin \phi \sin (\omega t + \delta) \tag{13}$$

These may be written in the form

$$x = A \cos \phi \sin \omega t$$
$$y = A \sin \phi (\sin \omega t \cos \delta + \cos \omega t \sin \delta) \tag{14}$$

Eliminating t, we obtain,

$$y^2 - 2xy \tan \phi \cos \delta + x^2 \tan^2 \phi = A^2 \sin^2 \phi \sin^2 \delta \tag{15}$$

A general second degree equation such as Eq. (15) represents an ellipse if the coefficients satisfy the following relationship:

$$\tan^2 \phi \cos^2 \delta < \tan^2 \phi$$

This condition is here satisfied since cos δ is less than unity.

Example 2. Show that Eq. (15) leads to the correct results when it is applied to the three special cases considered above, i.e., when applied to a quarter-wave, half-wave and full wave plate.

In Fig. 10 if the direction of propagation of the light is parallel to the optic axis, that is along the y-axis, the index of refraction has only a single value μ_0 for all planes of polarization, and double refraction does not occur.

Light Intensity

For a plane polarized wave of a given frequency the average energy flux (ergs/cm^2/sec) or <u>intensity</u> is proportional to the square of the amplitude A*. Hence the relative intensity of the plane polarized wave, Eq. (2)

---

* See as an example the computation of the energy of a vibrating string, MHW, p. 378.

incident on the crystal in Fig. 10 is

$$I = kA^2 \tag{16}$$

The two component waves which emerge from the crystal (Eqs. 13) have the relative intensities I_x and I_y as follows,

$$\begin{aligned} I_x &= kA_x^2 = kA^2 \cos^2 \phi \\ I_y &= kA_y^2 = kA^2 \sin^2 \phi \end{aligned} \tag{17}$$

or the total intensity I of the emergent wave is

$$I = I_x + I_y = kA^2 \tag{18}$$

which is equal to that of the incident wave.

Oblique Incidence

If, as in Experiments 5, 6, 7, 9 and 10, the light is incident on a surface which is neither parallel to nor perpendicular to the optic axis, then two refracted beams are formed inside the crystal. These beams travel in different directions and hence may be completely separated from one another if they are sufficiently narrow. This phenomenon may also be understood in terms of the two different speeds of propagation of waves polarized respectively parallel to and perpendicular to the optic axis.

For the sake of simplicity we shall confine attention to the wave form in a single plane in the crystal, namely the plane which is perpendicular to the upper and lower natural cleavage faces of the crystal and includes the optic axis. This is called the principal plane. It is the plane of the paper in Figs. 14 and 15. As we have already seen, any incident beam of light which passes normally into the crystal through the hole in the cardboard (see Figs. 14 and 15) may be thought of as consisting of equal

Fig. 14

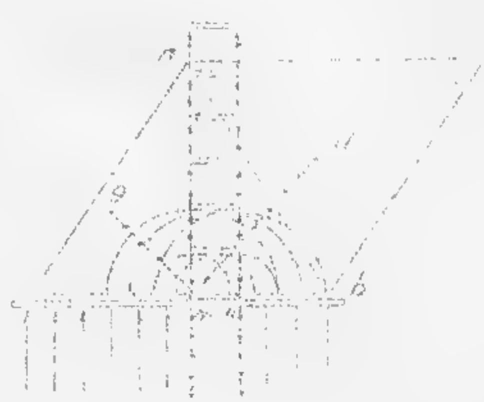


Fig. 15

vibrations in two planes, one perpendicular to the plane of the paper and the other parallel to this plane. Let us consider these two vibrations as separated, so that we may treat of one in Fig. 14 and the other in Fig. 15. Any vibrations which are parallel to the direction of the optic axis ab pass through the crystal with greater speed, than do vibrations which are perpendicular to this direction. The component in the plane of the paper (see Fig. 14) of the incident vibrations will give rise at the boundary mn of the crystal to transverse disturbances which will travel outward in all directions through the crystal. The portion of the wave front, however, which travels at right angles to the axis ab, that is, in the direction md (Fig. 14), will have its vibrations parallel to the optic axis, while the portion of the wave front which travels in the direction mg will have its vibrations perpendicular to this axis. If, then, vibrations parallel to ab travel faster than do such as are perpendicular to ab, the wave which originates at any point on mn will travel faster in the direction md than in the direction mg, and will consequently have an elliptical rather than a spherical form, the longer axis of the ellipse being in the direction at right angles to the optic axis ab. The envelope of all the ellipses which originate in the points on mn will be the line m'n': The beam will therefore travel through the crystal in a direction other than that of the normal to its wave front; that is, in the direction mm'. For the reason given in Sec. 5, Chap. 4, there will be destructive interference at all points outside of the parallels mm', nn'.

On the other hand, the waves which start out from each point on mn because of the propagation into the crystal of the vibrations which were perpendicular to the plane of the paper (see Fig. 15) will be everywhere perpendicular to the optic axis, and hence will travel with equal speeds in all directions. The beam will therefore follow the usual law of refraction and will travel in a direction at right angles to its wave front, the waves from each point being now spheres instead of ellipses.

4. Construction of the Nicol Prism. In order that the light which is transmitted by a crystal of Iceland spar may consist of vibrations in one plane only, it is necessary to dispose in some way either of the ordinary or extraordinary beam so as to prevent it from passing through the crystal. This was first accomplished in 1828 by the German physicist Nicol in the following way. If the beam bc (Fig. 16) is made to enter the face of the

crystal at a certain oblique angle, the ordinary ray, being refracted more than the extraordinary (see Exp. 7), will travel in the crystal in the direction co, for example, while the extraordinary ray will take the direction ce.

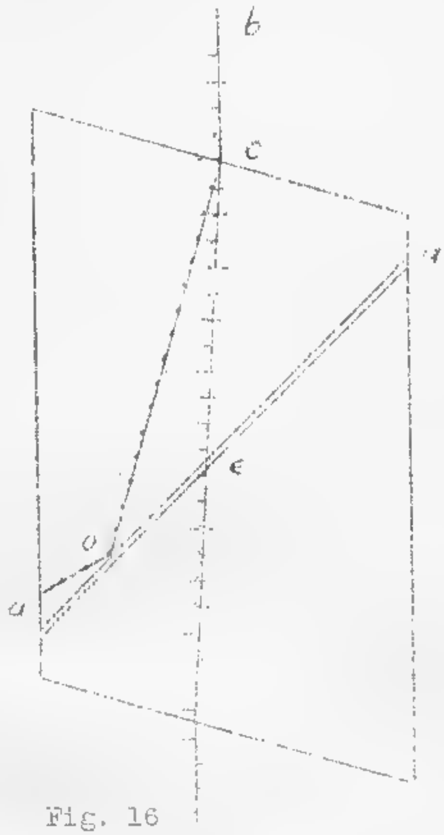


Fig. 16

Now Nicol cut the crystal into two parts along the plane aa, and then cemented the parts together again with Canada balsam. This balsam has an index of refraction which is smaller than that of the ordinary ray, but larger than that of the extraordinary ray; hence it was possible, by using a long crystal like that shown in the figure, to choose the plane aa so that the ordinary ray would be totally reflected and absorbed in the blackened walls of the crystal, while the extraordinary ray would pass through.

5. Dichroic Crystals; Polaroid. Certain uniaxial crystals of which tourmaline is an example have the remarkable property of transmitting light polarized along the optic axis, but at the same time absorbing light whose plane of polarization is perpendicular to the optic axis. Such crystals are called dichroic. A dichroic crystal, if used as in Fig. 10, with the direction of propagation of the light along the z-axis, forms a very simple polarizer since the waves polarized along the x-axis are absorbed by the crystal itself, but the wave polarized along the y-axis are transmitted through the crystal.

A commercial product known as Polaroid takes advantage of this property. It consists of a film of cellulose acetate which has imbedded in it a very large number of minute synthetically formed dichroic crystals. These crystals are all given the proper orientation, by stretching the film in one direction during the manufacturing process. An outstanding advantage of polaroid over other polarizers is that polaroid films can readily be produced which are several square feet in area. They have already found wide commercial applications.

6. Effects produced by the passage of polarized light through thin crystals.

Experiment 11. Arrange the polarizing apparatus precisely as in Fig. 1, save that the Nicol prism replaces the mirror m'. Rotate the Nicol until the flame is completely extinguished. Then obtain from the instructor a half-wave plate (for sodium light) of mica or selenite and place it on the slide holder h. You will find that, in general, the insertion of the mica causes the light to reappear. Rotate the mica about a vertical axis and note that in one revolution there are four positions, just 90° apart, at which there is extinction. These are the positions in which the plane of vibration of the light which is incident upon the mica is either parallel to or perpendicular to the plane containing the optic axis of the mica. Rotate the mica in a horizontal plane until it is just 45° from one of these positions of extinction. Then rotate the Nicol. The image of the flame will be found to disappear when the Nicol has been rotated through 90°.

This experiment can be understood in terms of the properties of the half-wave plate discussed in Sec. 3. The emergent light is plane polarized with its plane of polarization perpendicular to that of the incident light.

Experiment 12. Replace the half-wave plate by one half as thick, that is, by a quarter-wave plate. Set it at first so that it is 45° from the point of extinction, the Nicol being set in the position corresponding to extinction when no plate is interposed; then rotate the Nicol and note that no change takes place in the intensity of the transmitted light because of this rotation.

No change in intensity results from rotating the Nicol since in this case the light is circularly polarized. If now the quarter-wave plate is turned to a different position elliptically polarized light will be formed.

The major and minor axes of the ellipse may easily be found by observing in what direction the analyzing Nicol must be turned in order to obtain a maximum or a minimum of transmitted light.

7. Colors Produced by Thin Crystals in Polarized Light.

Fig. 17

Experiment 13. Set the polariscope in a window in the position shown in Fig. 17. the black paper being removed from the mirror n. If m has precisely the same inclination which was given it in Exp. 1 (Fig. 1), then, when the polariscope is so turned that the prolongation of the line ma meets the clear sky, the white light from the sky will strike the lower side of m at the polarizing angle, be reflected to the mercury mirror n, and return with little loss as a plane polarized beam to the Nicol N. Set N so that this beam is extinguished. Place a sheet of mica about twice as thick as a half-wave plate for sodium light upon h and turn it until it is just 45° from a position of extinction. When viewed through the Nicol it will be found to be brilliantly colored. Rotate the Nicol slowly and notice that a rotation of 45° causes the color to disappear, but that a rotation of 90° causes a color which is the complement of the first color to appear. Further rotation through 90° will cause the first color to return, and so on. When sheets of mica of different thicknesses are used different colors will be produced, but a rotation of the Nicol through 90° will always cause the color to change to that of the complement of the original color.

In order to understand the cause of this phenomenon suppose, for simplicity, that the mica is just thick enough to produce a retardation of one-half wave length of the longest red wave. Since the shortest violet waves have about one half the wave length of the longest red, this same plate will produce a retardation in the violet of one whole wave length. The violet wave will therefore emerge from the crystal with both of its components in the same phase, and these components will recombine into a plane vibration precisely like that which entered the crystal. The red ray, however, will emerge from the crystal with one of its components one-half wave length behind the other, and these two components will recombine into a vibration at right angles to that of the entering ray. If, then, the Nicol is in the position for extinction when no crystal is interposed, it will cut out all of the violet in the incident white light and transmit all of the red, so that if these red and violet waves fell alone upon the crystal, a rotation of the Nicol would cause red and violet to appear alternately. As a matter of fact, however, if the incident light is white, all of the colors between the red and the violet will be present, and the vibrations of the transmitted light which correspond to them will be ellipses of some form. However, the wave lengths which are close to the red, namely orange and yellow, will be largely transmitted by the Nicol along with the red, and will have but small components to be transmitted with the violet, while the wave lengths near the violet, namely the blues and the shorter greens, will be mainly transmitted with the violet. Hence the light which passes through the Nicol in its first position will be some shade of red, because it will have most of the shorter wave lengths subtracted from it; while, when the Nicol is turned through 90°, all of the wave lengths which were before cut out will be transmitted. The color will therefore be exactly the complement of the first color, that is, it will be some shade of blue.

A crystal which is too thin to produce one-half wave length retardation of the shortest visible rays, namely the violet, cannot show any marked color effects in polarized light, since no wave length can be entirely cut out for any position of the Nicol. On the other hand, a crystal which is so thick as to produce retardation of very many wave lengths of any one color will produce also a retardation of an exact number of wave lengths for each of many other colors scattered throughout the spectrum. These colors will all be cut out by the Nicol, and the transmitted light will likewise consist of wave lengths which are taken from all parts of the spectrum, and will therefore reproduce the effect of white light. Hence these color phenomena in polarized light can be observed only with crystals which produce a small number of wave lengths of retardation. By scraping crystals down to proper thicknesses in different parts, color patterns of much beauty are often produced when the crystals so treated are viewed in the polarized light. All the colors, of course, change to the complements upon rotation of the Nicol through 90°.

Experiment 14. Place a number of these designs in selenite or mica upon the slide holder h, and observe the appearance of the complementary colors in different portions of the design as the Nicol is rotated.

Experiment 15. Observe in convergent polarized light a crystal of Iceland spar, say 1 mm thick, the upper and lower faces of which are planes perpendicular to the optic axes. The beam of convergent light is most easily obtained by placing the crystal very close to the Nicol in the arrangement of Fig. 17, so that the observer looks down through the crystal upon a field of considerable width, from all parts of which polarized light is approaching the eye. If the Nicol was originally set for extinction, you will observe a dark center surrounded by a series of brilliantly colored rings upon which is superposed a black cross (see Fig. 18). Rotating the Nicol will cause the black cross to change to a white one, and all of the colors to change to their complements (see Fig. 19).

Fig. 18

Fig. 19

These effects may be explained as follows. The central rays pass through the crystal in a direction parallel to the optic axis. They therefore suffer no resolution into ordinary and extraordinary components, and hence no change in the character of their vibration. They are cut out by the analyzing Nicol, hence the black center. The rays, however, which converge upon the eye after passing through the outer edges of the crystal have traveled in directions slightly oblique to the axis, and have therefore suffered decomposition into ordinary and extraordinary rays, which have undergone different retardations. A given retardation of one ray with respect to the other corresponds to a given color precisely as explained above. A given color must, of course, be symmetrically distributed about the axis of the converging beam, since the thickness of the crystal is so distributed; hence the concentric rings of color. The black cross is superposed upon these rings because in two particular planes, namely those for which the incident vibration is respectively in and perpendicular to the plane containing the axis and the ray, even these oblique rays are not split up into components, but pass through vibrating in their original direction, and are therefore cut out by the Nicol. Upon rotating the Nicol through 90° all of those extinguished rays are, of course, transmitted; hence the white cross.

8. Rotary Polarization.

Experiment 16. Arrange the polarizing apparatus as in Fig. 1, save that m' is replaced by a Nicol, and place upon h a crystal of quartz, say 5 mm thick, the upper and lower faces of which are made by planes which are at right angles to the optic axis of the quartz. When the Nicol is set for extinction the introduction of the quartz into the path of the beam will be found, in general, to cause the extinguished image of the flame to reappear. Rotate the Nicol, and measure the amount of rotation required to cause the yellow flame to disappear again. According to accepted results this rotation for sodium light should be 21.7° per millimeter of thickness of the quartz. Replace the sodium flame by the ordinary violet flame of the Bunsen burner and repeat, rotating this time until all trace of the violet color in the flame has disappeared. The rotation will be found to be nearly double that found for sodium light. The rotation in the case of light filtered through red glass will be found to be less than that in the case of yellow light.

These experiments show, first, that plane polarized light which passes through quartz in the direction of its optic axis remains plane polarized after transmission, and, second, that the plane of polarization of the light is rotated by the quartz, the amount of the rotation being greater for the short wave lengths than for the long. The discovery that quartz is able to produce these effects was made by Arago in 1811.

From the difference in the amount of rotation of different colors it follows that if plane polarized white light is incident upon the lower face of the crystal of quartz, the light which is transmitted by the analyzing Nicol will be colored, since this Nicol will extinguish completely only those vibrations which are perpendicular to its transmitting plane. It follows, further, that if the Nicol is rotated through 90°, all of the components of the white light which were before extinguished will be now transmitted and vice versa, and hence that the rotation through 90° will cause the color to change to the complement of the first color.

Experiment 17. Verify the above predictions by setting the polariscope so that light from the sky falls upon it in the manner indicated in Fig. 17. To show that the lights transmitted by the analyzer in planes 90° apart are complementary, it is best to replace the Nicol by a thick crystal of Iceland spar, or some other form of double-image prism, so that both of the lights to be compared may be transmitted at once. As the analyzer is rotated, the overlapping portions of the images will be found to maintain the color of white light, while the opposite non-overlapping portions will be of complementary colors.

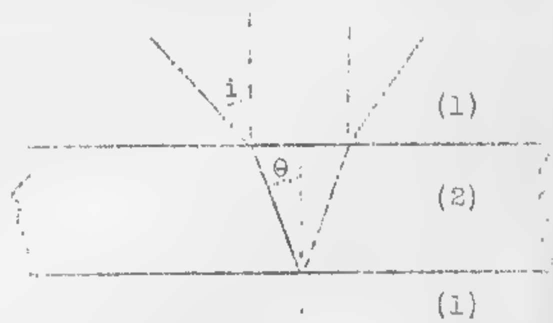
It has been found that there are two kinds of quartz crystals, one of which produces rotation to the right, the other to the left. This difference in optical behavior corresponds also to a difference in crystalline structure which makes it easy to distinguish the so-called right-handed from the left-handed quartz crystals without actually making the optical test.

Furthermore, it was discovered by Biot, in 1815, that there are certain liquids which possess the same property shown in the above experiments by quartz. Of these, solutions of cane sugar have received most attention, for the reason that the amount of rotation produced by a column of sugar solution of fixed length is taken as the commercial test of the strength of the solution. As in the case of quartz, there are found to be two kinds of cane sugar of precisely the same chemical constitution, but of slightly different crystalline form, which rotate the plane of polarization in opposite directions. The form which rotates to the right is called dextrose, the other levulose. It is possible to convert dextrose to levulose sugar by acting upon it with hydrochloric acid. This conversion is actually made in sugar testing, the rotation due to the conversion being the quantity directly measured.

Problems

1. At what angle must a beam be incident upon a smooth water surface in order that the reflected beam shall be plane polarized?

2. Monochromatic light is incident at Brewster's angle upon medium (2) from the less dense medium (1). Discuss and conclude whether or not (a) angle θ is Brewster's angle of incidence upon medium (1) from medium (2), and (b) whether or not angle θ is greater or less than the critical angle for total reflection at the lower surface.



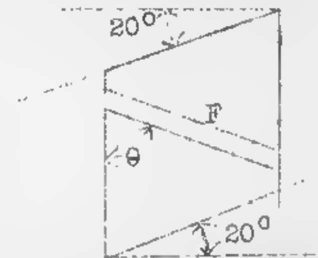
3. A thin piece of quartz is cut so its faces are parallel to the optic axis. If plane polarized light is incident normally on the quartz and if its plane of polarization makes an angle of 45° with the optic axis, what is the minimum thickness of quartz required to produce (a) a rotation of the plane of polarization through 90°, (b) emergent light which is plane polarized in the same plane as the incident light, and (c) circularly polarized light?

4. Circularly polarized light is incident normally on (a) a half-wave plate and (b) a quarter-wave plate? What is the state of polarization of the emergent light in each case?

5. Light passes in succession through a $\lambda/4$ and a $\lambda/2$ plate. The optic axes are parallel to one another. The incident light is plane polarized with its plane of polarization at 45° with the optic axes. (a) What type of polarization does the light emerging from the $\lambda/2$ plate exhibit? (b) Rotate the $\lambda/4$ plate through 45° about the light ray as an axis so that its optic axis becomes parallel to the plane of polarization of the incident ray. Answer questions as in (a).

6. Elliptically polarized light $\begin{matrix} x = A \cos \omega t \\ y = -B \sin \omega t \end{matrix}$ of intensity I_0 is incident upon a $\lambda/4$ plate with its optic axis in the y direction. Describe the nature of the emergent light with respect to its polarization, intensity, and orientation.

7. It is desired to construct a Nicol prism which will operate satisfactorily for Na-D light. Two pieces of Iceland spar are joined by Canada Balsam cement ($\mu = 1.53$). The end faces are cut at 20°. What angle θ would be proper for the interface F?



8. Two Nicol prisms are placed in line. The intensity of the emergent light from one source of light is equal to that from a second source when the angles between the principal planes of the Nicols are 30° and 45° respectively in the two cases. What is the ratio of the intensities of the original sources?

9. Three Nicol prisms are placed in line. The principal plane of the last prism is at right angles to the principal plane of the first and the principal plane of the second makes an angle θ with the principal plane of the first. Ordinary light of intensity I_0 is incident on the first Nicol. What is the intensity of the light emerging from the last Nicol?

10. Two Nicol prisms are placed in line with their principal planes making an angle θ with one another. What is the change in intensity of the emergent light if the angle θ is made ϕ? The incident light is unpolarized.

11. Na-D light, plane polarized at an angle of 30° with the optic axis of a quartz plate 0.6 mm thick, is incident normally on a surface which is cut parallel to the optic axis. Describe quantitatively the state of polarization of the emergent light.

12. Two Nicol prisms are placed with their principal planes at an angle of 45° with one another. A $\lambda/2$ plate cut with its optic axis in its plane is placed between the two Nicols normal to the beam. (a) Find the angular positions of the $\lambda/2$ plate for which the intensity of the emergent light is a maximum. (b) Find the ratio of the intensity of the emergent light when the $\lambda/2$ plate is in place and oriented for maximum intensity, to the intensity when the $\lambda/2$ plate is removed.

13. A thin plate of quartz, cut with its optic axis perpendicular to the plane of its large faces, is placed between two Nichols. The plane of its large faces is normal to the axis of the Nichols. For what thickness of quartz plate will no Na-D light emerge from the second Nicol? The incident light is unpolarized and the principal planes of the two Nicols are parallel to one another.