LIGHT

A TEXTBOOK FOR STUDENTS WHO HAVE HAD ONE YEAR OF PHYSICS

By

H. M. REESE

Professor of Physics in the University of Missouri

Columbia, Missouri
MISSOURI BOOK COMPANY
1921

PREFACE.

The writing of this book was undertaken because no existing text on the subject quite filled the needs for my own classes. The first draft was mimeographed, and has been used in that form with some success for several years.

It is planned for students who have had no training in the calculus, because many of those who take second-year physics at the University of Missouri suffer from that handicap. The first part has purposely been made rather easy, the intention being to lead gradually from less to more difficult matter. A persistent attempt is made to lay stress upon the experimental basis for our theories, and to point out such reasons as exist for and against them; because my own experience has been that many students, though they may learn the *facts* of a science conscientiously and in a sense thoroughly, fail completely to realize the inductive processes on which the theoretical structure is founded, thus missing one of the chief educational values to be derived from the study of science. It is in line with the same idea that certain matters have been introduced, particularly in the last two chapters, whose purpose is to give the reader an idea, incomplete though it may be, of the present state of optical theory and allied branches of physical science.

Thanks are due to my colleague, Professor O. M. Stewart, for a number of valuable suggestions, and also to Professor Henry Gale, of the University of Chicago, who read the manuscript and suggested changes and additions which I have been glad to make.

H. M. R.
Columbia, Missouri,
October, 1920.

CONTENTS

CHAPTER I.

CHAPTER II.

CHAPTER III.

CHAPTER IV.

CHAPTER V.

CHAPTER VI.

CHAPTER VII.

CHAPTER VIII.

CHAPTER IX.

CHAPTER X.

CHAPTER XI.

CHAPTER XII.

CHAPTER XIII.

CHAPTER XIV.

CHAPTER XV.

CHAPTER XVI.

CHAPTER XVII.

LIGHT.

CHAPTER I.

1. Introduction.—2. Velocity.—3. Roemer's method.—4. Bradley's method.—5. Fizeau's method with the toothed wheel.—6. The rotating mirror method of Foucault (and Fizeau).

1. Introduction.—Simple and familiar observations teach us that the sensation of vision is caused by some agency that emanates from bodies external to us, and enters our eyes. For instance, we cannot see anything in a room with tightly closed blinds where there is no source of artificial illumination, such as a candle, fire, or electric lamp. Incidently, this experience also teaches that we may classify the objects we see into two groups: first, luminous objects, such as a candle, a fire, the sun, or the stars; second, objects such as a book, a tree, the walls of the room, etc., which can be seen only through the agency of a luminous body.

From the physicist's point of view, the study of light is the study of this activity, whatever its nature may be, which originates in luminous bodies and causes the sensation of vision when it enters the eye. His interest lies primarily in the way this agency starts into action in a luminous object, how it propels itself through space, how it behaves on striking objects of different kinds, such as glass, crystals, silver, water, etc., and its relations to all other physical phenomena, such as heat, electricity, and magnetism.

On the other hand, the physicist proper does not concern himself much with the parts that the eye and the nervous system play, in registering in our consciousness the *sensation* (vision), whose primary cause is the *physical agency* that we call light. This question is important and interesting enough, but it belongs primarily to the domains of the physiologist and the psychologist.

Let us begin our study by making a summary of such facts as common knowledge gives us about light.

In the first place, besides differences in brightness, which may be called a matter of *quantity*, there are also differences of *quality* to be considered, as shown in the phenomena of *color*.

Second, light travels approximately in straight lines, as is shown in the formation of shadows. Nevertheless, we shall see later that light does bend around corners to some slight extent, though not nearly so much as sound does.

Third, it differs from sound also in that it travels without hindrance through a vacuum. In coming to us from the stars, it travels through millions of miles of the most perfectly empty space obtainable.

Fourth, it is either itself a manifestation of energy, or else it carries energy with it, since any object which receives and absorbs it becomes heated.

Fifth, when it strikes a surface, more or less of it is generally reflected. (Those exceptional surfaces which reflect no light are said to be *black*). If the surface is highly polished, the light is reflected at a definite angle, in which case we say the reflection is *regular*. If the surface is rough, like that of a sheet of paper, the light is scattered in all conceivable directions, and the reflection is said to be *diffuse*.

Sixth, there are many substances, such that when light strikes their surfaces, although part is reflected, part enters the material and passes through it rather freely. Such materials are said to be *transparent*. Light traverses transparent materials approximately in straight lines, as it does the air or free space, but there is an abrupt bending of the rays at the place where they pass through the surface. This bending is called *refraction*.

Seventh, light travels either instantaneously, or else with enormous velocity. Here again the comparison with sound is very striking. The phenomenon of echoes shows that sound travels with a speed which, though great, cannot be called enormous, and indeed a fairly accurate measurement of this speed could be made by noting with a stop-watch the time required for an echo to be heard from a cliff or large building, whose distance from the observer is known. An exactly analogous experiment with light would be to note with a stop-watch the time that elapses between the flashing of a light and the perception of its reflection in a mirror, whose distance from the observer is known. Such an experiment would fail completely, because no stop-watch could record a short enough time-interval; and even without that objection, no human being has a "reaction-time" constant enough to manipulate a stop-watch with anything like the necessary precision.

2. Velocity.—Of the above mentioned seven points of common knowledge about light, the last (in regard to velocity) is of so much interest, and can be so easily discussed without a thorough knowledge of other optical phenomena, that we shall consider it here at some length.

It is interesting to note that Galileo actually tried to measure the velocity of light by the method outlined above, except that instead of using a mirror to send back the light (probably none then available were good enough to use over great distances) he stationed two observers with lanterns a great distance apart. Observer number one flashed his lantern, and number two answered by a flash of his own as quickly as possible. Number one then tried to measure the interval of time between his own signal and his perception of the answering signal. Of course no perceptible time-interval was found, and Galileo concluded from this that the velocity of light was too great to measure.

Since a velocity is always a distance divided by a time, Galileo's failure shows that in order to measure so great a velocity we may proceed in one of three possible ways. First, we may choose a distance so great that, in spite of the great velocity to be measured, the interval of time will be large enough to measure conveniently by ordinary methods. Second, it might be possible to get a direct comparison between the velocity of light and some known velocity (such as that of the earth in its orbit) which, although much smaller, is yet far greater than that of anything we can handle in the laboratory. Third, we may return to the principle of Galileo's method, with a relatively short distance (say a few miles) and correspondingly small time-interval, if we use a mirror to return the light and find some very refined method for measuring an exceedingly short time. The last method would have this advantage over the other two, that since the distance concerned is not excessive, it might be possible to measure the velocity, not only in air or in free space, but also in water and other transparent materials.

It is a matter of historical fact that each of the three possibilities suggested above has been successfully carried out, —the first by the Danish astronomer Roemer, in 1676, the second by the English astronomer Bradley, in 1728, and the third by two French physicists, Fizeau and Foucault, in 1849 and 1850 respectively.

3. **Roemer's Method.**—The planet Jupiter, like the earth, revolves about the sun in a nearly circular orbit, but its orbital radius is so much larger that it takes nearly twelve of our years to complete the circuit. It has several satellites, similar to our moon, one of which circles the planet in about 11 hours. Once in every revolution, it enters the shadow of Jupiter and, since it is not a luminous body, but can be seen only by virtue of the sun's light, it then disappears for a short time. The interval of time between two successive eclipses is called the *period.* We would naturally expect the period to be constant, but it was long known that it seems to vary, according to the relative positions of Jupiter and the earth. In figure 1, the larger circle represents the orbit of Jupiter, the smaller that of the earth, with the sun at the center of both. (Actually each orbit is an ellipse, with the sun at one focus.) Suppose that at a given time the earth is at E_1 and Jupiter at J_1, the two being in line with the sun. A little more than six months later, they will again be in line, with the sun between them, the earth at E_2, Jupiter at J_2, for Jupiter moves more slowly in its orbit than does the earth. Again, at a still later time, the earth will be at E_3 and Jupiter at J_3, the former having made something more than a complete circuit, while the latter has travelled only through the arc $J_1J_2J_3$. Evidently there are times, as at A, when we are receding from Jupiter, and other times, as at B, when we are approaching him. It was noticed that when the earth is receding from Jupiter the period of the satellite seems to be longer, when it is approaching him shorter, than the average. Thus if,

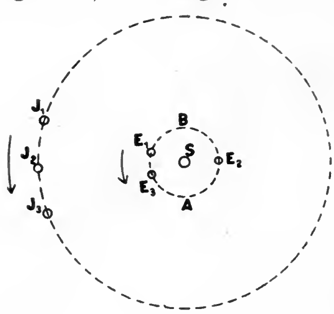

Figure 1

when the earth is in such a position as E_1, with Jupiter at J_1, a complete schedule of satellite eclipses be made out in advance, on the supposition that they occur with a regular period, it will be found that they appear more and more behind schedule time, till the earth and Jupiter are in the positions E_2 J_2, and then begin to pick up till they are again actually on schedule time, when the earth is at E_3 and Jupiter at J_3.

Roemer saw that this phenomenon could be explained perfectly by supposing that the eclipses occur at perfectly regular intervals, provided that a finite time is required for the light which brings us the news of an eclipse to travel the very great distances involved. For when we are moving away from Jupiter, as at the position A, each succeeding eclipse is announced to us by light that must travel a somewhat greater distance, and therefore the apparent period would be increased by the time required for the light to travel the additional distance. On the basis of a schedule of eclipses, such as was described in the previous paragraph, it is found that the eclipses observed when the earth and Jupiter are on opposite sides of the sun seem to be about 16.6 minutes behind the schedule. According to Roemer's views, this would indicate that it takes that much time for light to cross the earth's orbit. Since the mean radius of the orbit is about 92.8×10^6 miles, this gives for the velocity of light 18.6×10^4 miles/sec., which is the same as 2.99×10^{10} cm./sec.

It is worth noticing that the great distance of Jupiter from the earth does not enter into the problem, only the changes in that distance, and it would not be possible to determine the velocity of light by observations of the satellite if the earth were stationary with respect to the planet. It is true that each eclipse would be observed some time later than its actual occurrence, but we could not know *how much* later, unless we already knew the velocity as well as the distance from Jupiter, and calculated back from the time of the appearance to the time of actual occurrence. That is, one would have to know the very thing which it is his object to find.

4. Bradley's Method.—The so-called fixed stars are so far from us that the relatively small range of motion of the earth in its orbit hardly changes their apparent positions, that is, their directions from us. Still, as astronomical methods became more refined, it was observed that when the earth is on one side of its orbit the position of a star seems shifted slightly to the other side, as we should expect from ordinary geometry. Of course this effect, which we call "parallax," is most pronounced on the nearest stars, and its measurement enables us to estimate the distance of such stars.

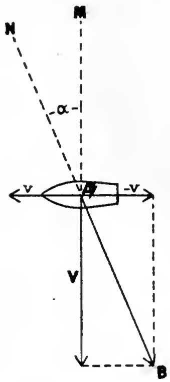

Figure 2

The phenomenon discovered by Bradley, known as "aberration," is an entirely different matter, though it too consists of an apparent change in the position of a star. It is a shift, not in a direction *opposite to that in which the earth stands from the sun*, but *towards the direction in which the earth is moving* at the time, that is, it depends not on the *position* of the earth, but on its *velocity*. Bradley found an explanation for this phenomenon, suggested by the way a flag acts when it is affected both by the wind and by the motion of the ship on which it is carried. For instance, in figure 2, let the vector v represent the velocity of the ship, V that of the wind, which is here supposed to blow directly across the ship. The flag will not stand out in the true direction of the wind, but in a direction such as AB. That is, the flag is affected not only by the true wind, but also by an apparent wind equal and opposite to the velocity of the ship. Evidently it will form an angle with the true direction of the wind whose tangent is v/V. This comes to the same thing as saying that to a person travelling with the ship the wind appears to come, not from the true direction M, but from the direction N, where tan. $a = v/V$.

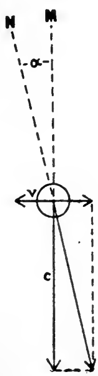

Figure 3

Now let us take another figure (3) in which we replace the ship by the earth moving through space, and the wind by light coming from a star whose position is broadside on to the earth's motion. Instead of the velocity V of the wind, we have the velocity c, of light. Then the star, whose real position is in the direction M, will appear in the direction N, making an angle a with the true direction, such that tan. $a = v/c$. The angle a cannot be measured directly, since we see only the apparent direction of the star, not its true position; and we could never hope to find the velocity by means of the aberration if the earth continued always moving in the same direction. But six months later the earth would be on the opposite side of its orbit, and moving on the opposite direction. Therefore, if we take the angle between the two apparent positions of a star at times six months apart, this will be twice the angle a. Then knowing the velocity of the earth in its orbit, we can at once calculate the velocity of light. This was done by Bradley, who thus got a value for c which was quite close to that obtained by Roemer's method.

5. Fizeau's Method with the Toothed Wheel.—The main idea in this method is to send a beam of light through a small hole toward a mirror from which it is reflected back toward the hole. The hole is opened and closed very rapidly, the rapidity of this action being gradually increased till the light that passed through the opening going out finds it closed when it returns. This rapid opening and closing of an aperture was accomplished by using a toothed wheel, as shown in profile in figure 4. The wheel was rotating rapidly at a controlled speed, and the light passed out through the gaps between the teeth. The actual arrangement of the optical parts of the apparatus was much more complicated than the simple diagram of figure 4, and is better shown in figure 5, which we shall consider in detail because it involves so many of the principles common to most optical experiments.

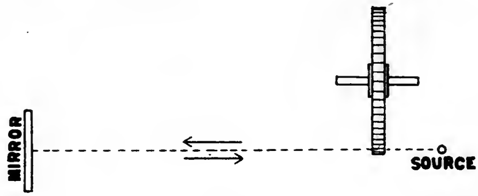

Figure 4

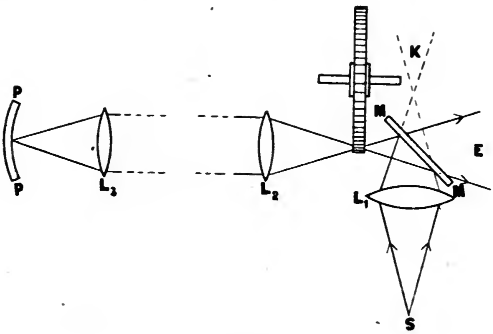

Figure 5

In the first place, since an observer must watch to see when the proper speed is reached to prevent the light from returning through the hole through which it passed out, some place must be left for the eye, which obviously cannot be placed either in front of the source of light or behind it. The source, for instance an arc-light, is therefore placed to one side, as at S, and reflected toward the wheel by a small mirror MM, placed at 45°. This is not a heavily silvered affair, like a household mirror, but either a simple unsilvered sheet of plane glass, or, better yet, one which has a coating of silver so thin that about half the light is reflected, half transmitted. It is obvious that by such a device half the available light is lost, by passing through in the direction of K, and half of what is left is again lost in the returning beam, for it also is half reflected, so that only ¼, or even less, of the original beam can reach the eye placed at E. Nevertheless, enough is left for the purpose, and it is possible with this arrangement to see the returning light without interfering with its passage outward.

Besides the mirror, a system of lenses is introduced, whose purpose is two-fold: first to prevent the light from spreading out indefinitely, and so becoming weakened, second to concentrate it where it meets the rim of the wheel, so that it can pass through only one gap at a time. The lens L_1, between the source and the inclined mirror, forms an image of the source just at the point where it passes through the wheel. If one of the openings happens to be at this point the light will pass through, spreading out from the hole just as if the latter were itself the source. A second lens L_2, whose principal focus is at the rim of the wheel, receives the rays and converts them into a parallel beam, which can travel to any required distance without suffering any further weakening except that due to the small and unavoidable absorption by the air through which it passes. In Fizeau's experiment it was carried about 4 miles. At the end of this distance, it could be made to fall upon a plane mirror, as shown in the diagram of figure 4, and so reflected back, but a plane mirror perfect enough over its whole surface for this purpose would be hard to construct. Consequently, a third lens, L_3, is inserted, at whose principal focus is a concave mirror PP, with its center of curvature at the lens. By this means only a small portion of the mirror is used, and it makes little difference if it is not perfect all over.

The light now returns through L_3 and L_2, is again focussed on the same part of the wheel, and if an opening is there it passes through, part of it getting through the mirror MM to the eye.

Now consider what happens as the wheel is started in revolution, slowly at first, but with increasing speed. At first a flash will be seen whenever a gap in the wheel leaves the path clear. But the eye cannot detect flashes coming more frequently than about 20 per second, therefore the succession of flashes will change gradually to an apparently steady light as the speed of the wheel increases, for much more than 20 teeth per second must pass across the field of vision before the speed is sufficient for a tooth to completely block the path of the returning ray which went out through the adjacent gap. But if the speed of rotation is increased still more, this apparently steady light will gradually become fainter, as each tooth encroaches more and more on the returning beam. It vanishes completely when the speed is such that a tooth moves into the place formerly occupied by a gap while the light is passing from the wheel to the concave mirror and back again. It is obvious that a further increase of speed will cause a reappearance of the light, which indeed can go through a number of maxima of intensity, separated by total darkness, if the speed of rotation increases indefinitely.

In this experiment, the toothed wheel, whose speed of rotation can be determined by suitable mechanical devices, serves as a means of measuring very short time-intervals, thus enabling us to measure the velocity of light over relatively short distances.

6. The Rotating Mirror Method of Foucault (and Fizeau). —This method, first proposed by Arago, another Frenchman, was worked out by Foucault and Fizeau, jointly at first, but afterwards independently. Foucault finished the task first, Fizeau having been delayed by an accident.

In this case, as in that of the toothed-wheel method, we shall consider first a diagram showing only the crude principle of the method, figure 6. M_1 is a flat mirror, which can be rotated rapidly about an axis in its own plane, perpendicular to the plane of the paper. It receives a beam of light from a source S, and reflects it in a direction depending upon the position of the mirror at the instant. Once in every revolution, and only during a very small part of the revolution, the reflected light falls upon a second mirror, M_2, which reflects it back to M_1, and M_1 in turn reflects it back toward the source. If the mirror were stationary, or if the velocity of light were infinite, so that it travelled from M_1 to M_2 and back before M_1 had turned at all, then the returning beam would come exactly back to the source S. But since it requires a finite time to traverse this distance, the mirror will have turned a little, and the beam will not return exactly to the source, but to a new point S′.

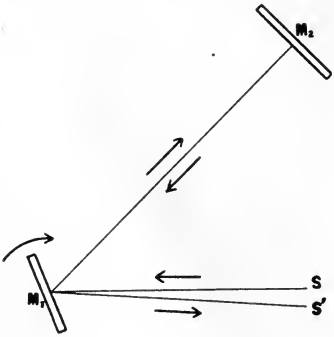

Figure 6

Here, as in the tooth-wheel method, some other details must be introduced to make the experiment really possible. The most important thing to remember in all optical experiments is that a beam of light never consists of a single ray, but always of a great many, generally having different directions. For this reason it is impossible to make accurate measurements of position unless, by the use of lenses or other means, the rays are concentrated to a definite focus.

Figure 7 is a complete diagram of the apparatus. Light from the source S passes through a narrow slit, introduced to give sharp edges to the illuminated area, and falls, after reflection from the half-silvered 45-degree mirror mm, upon the lens L, which would, if the mirror M_1 were not present, bring it to a focus at the point I_1. The mirror M_2 is concave, forming a section of a sphere whose center is at the axis of M_1, and which passes through the point I_1. Let R be the radius of this sphere, that is the distance from M_1 to M_2. As the mirror M_1 rotates, the direction of the rays that it reflects changes, but always in such a way that an image of the illuminated slit is formed somewhere on the circle $I_1A_1M_2B$. We may regard this image as sweeping round the circle with an angular velocity twice that of the rotating mirror. During a short interval of time, it will fall upon the fixed mirror M_2, which reflects the light directly back over its original path as

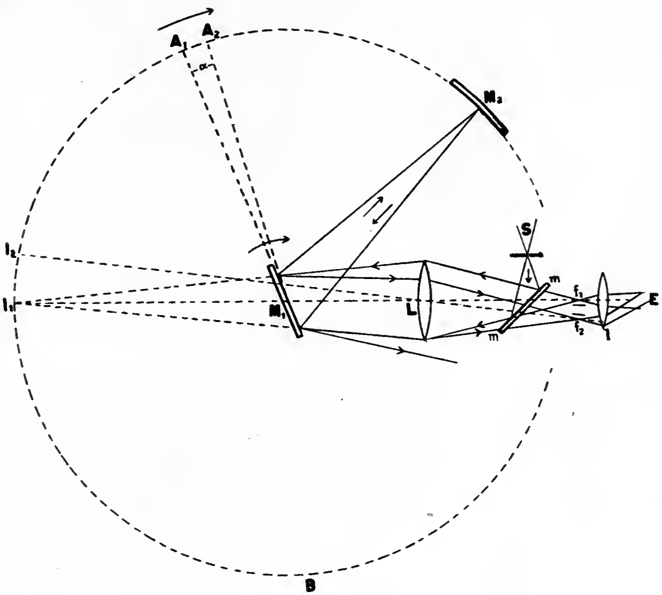

Figure 7

far as M_1. If the latter had been at rest, the light would have continued to retrace its path as far as the half-silvered mirror mm. Part of it would then penetrate mm and come to a focus at f_1, forming there an image that could be seen by the eye at E, aided perhaps by an eye-lens l. It will be noticed that this returning light would act exactly as if it came from the point I_1 instead of from M_2. In fact, I_1 is the "image by reflection" of the point M_2, formed by the mirror M_1. According to the laws of reflection in a plane mirror, which will be taken up in detail later, the angles $A_1M_1M_2$ and $A_1M_1I_1$ are equal, A_1 being the point where the plane of the mirror M_1 cuts the circle. Since the rays striking the lens would appear to come from I_1, the image f_1 would be the "conjugate focus" of I_1. According to the law of lenses, the line I_1f_1 passes through the center of the lens, and f_1 is at such a distance that

$$\frac{1}{Lf_1}+\frac{1}{LI_1}=\frac{1}{F}$$

where F is what we call the "focal length" of the lens L. (section 31)

In point of fact, however, M_1, instead of being at rest, is rotating in the direction shown by the arrow. Consider the light that starts from it toward M_2 when M_1 is in the position indicated. When this light returns to M_1 the latter will have turned through a small angle a, so that its plane will now intersect the large circle in the point A_2. Therefore it will reflect the light returned from M_2 as if it came, not from I_1, but from a new point I_2, such that $A_2M_1I_2 = A_2M_1M_2$. After reflection then, the lens L will bring it to focus, not at f_1, but at the new point f_2, such that the straight line I_2f_2 passes through the center of L.

The eye would see an image of the slit at f_1 if the mirror were at rest, at f_2 if it were rotating. In the latter case, the light would not be really steady, but consist of a series of flashes; but, since the flashes come much more rapidly than 20 per second, it would to all appearance be steady. If the mirror were started from rest and gradually picked up speed, the image would first appear at f_1 and gradually move away, but would seem perfectly still as long as the speed of the mirror were steady.

Evidently, the distance f_1f_2 depends upon the velocity of light, together with the distance between the two mirrors M_1 and M_2, the speed of rotation of the mirror M_1, and the focal length F; consequently we should be able to get the velocity of light if these other quantities are known. The small distance f_1f_2 is measured with a *micrometer* (see section 37), the large distance M_1M_2 by steel tapes, or by surveyors' methods, and the speed of rotation of the mirror by a special revolution-counter and stop-watch, or some equivalent mechanical device. The focal length F is supposed to be known, or it can be found by methods to be described later.

In order to derive the formula for finding the velocity of light, c, we shall let R = the distance M_1M_2, d = the distance

f_1f_2, $t =$ the time required for light to travel the distance 2R from M_1 to M_2 and back, which is also the time required for the mirror to turn through the angle α, and $n =$ the number of revolutions per second of the turning mirror. Then the velocity of light is

$$c = \frac{2R}{t} \qquad (1)$$

and the angular velocity of the mirror, in radians per second, is

$$2\pi n = \frac{\alpha}{t} \qquad (2)$$

Since, by the laws of reflection, the angle $A_2M_1M_2 = A_2M_1I_2$, and $A_1M_1M_2 = A_1M_1I_1$, therefore $I_2M_1I_1 = 2\alpha$. In the actual experiment, the lens L is placed very close to M_1, not more than a few feet away, while the distance $I_1M_1 = I_2M_1 = R$ is quite large, say several hundred meters. Therefore, very nearly, the angle $I_2LI_1 = I_2M_1I_1 = 2\alpha$, and the opposite angle f_1Lf_2 is also approximately equal to 2α. With I_1 and I_2 so far from the lens, the images f_1 and f_2 come practically at the focal distance from the lens, that is $f_1L = f_2L = F$. Therefore, in radian measure,

$$f_1f_2 = d = 2\alpha F \qquad (3)$$

Now, from equations (1), (2), and (3) we can eliminate t and α and we get

$$c = 8\pi nRF/d$$

from which the value of c can be computed, as soon as n, R, F, and d are measured.

Professor A. A. Michelson, of the University of Chicago, has made a number of improvements in the details of Foucault's method, but has not altered the principles involved.

Measurements of the velocity of light, obtained by experimental methods, vary from 298,000 to 301,382 kilometers per second. It is usually regarded as sufficiently accurate for all purposes to take the round figure 300,000, or, when expressed in centimeters per second, 30,000,000,000 $= 3 \times 10^{10}$.

Problems.

1. The star Sirius is 5×10^{13} miles away from us. How many years are required for its light to reach us?

2. Calculate the time required for light to travel four miles and return after being reflected by a mirror.

3. What is the maximum angle by which, owing to aberration, a star may seem to be displaced from its true position?

4. Derive the formula applying to Fizeau's toothed-wheel method for finding the velocity of light.

5. Referring to figure 1, at what positions of the earth does the observed period of the satellite eclipses seem longest and shortest respectively?

6. It frequently happens that the moon passes between the earth and a star (occultation of the star). What would be the effect upon this phenomenon if red light travelled faster than blue, in the space between moon and earth?

7. The "parallax" of a star is the angle which the radius of the earth's orbit, 92.8×10^{6} miles, subtends as seen from the star. A "parsec" is the distance of a star whose parallax is one second of arc. Find its value in miles, and in "light-years," the distance light travels in a year.

CHAPTER II.

7. Refraction through a prism.—8. Newton's conception of color.—9. Impure colors.—10. Color due to absorption.—11. Color due to other causes.—12. Black and white.—13. Complementary colors and color mixture.—14. The eye.—15. Color vision theories.

7. Refraction through a prism.—Of the earlier physicists, the one who made greatest progress in the study of light was Sir Isaac Newton. It must be admitted that he was led to believe in certain hypotheses which have since been discarded, but in spite of that fact he accumulated, by experimental methods, a large amount of needed definite information; and his philosophical discussion helped greatly in the development of the theory that later supplanted his own faulty one.

Newton was the first to get a clear idea of color, which he attained through a study of glass prisms. Everyone knows that a prism of any transparent substance not only bends rays of light, but also makes a beam of white light to show color on the edges. Thus, let W in figure 8 represent a window, through which white light enters a room, passing through the prism and entering the eye placed at E. The arrows indicate in a general way the course of the rays. Because we judge the position of an object by the direction of the rays as they enter our eyes, the window appears to be displaced from its true position to some such place as W′. But, more than this, the window appears white only in the middle. That edge of it which, as seen through the prism, is nearest to its proper position, is red, the other edge violet. Newton saw that this experiment indicates white light to be a composite of many colors, the color effect at the edges being a result of some property of the prism which causes it to bend, or *refract*, some of these component colors more than others; for instance, the

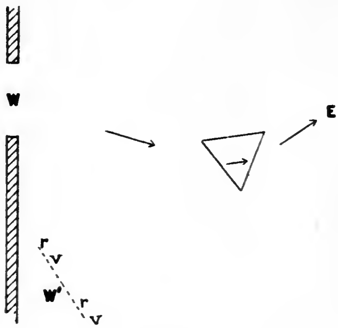

Figure 8

violet more than the red, and other colors to an intermediate degree. According to this hypothesis, the eye would see a *red* image of the window, as indicated by the rectangle rrrr in figure 9, or as shown in the plan of figure 8 by rr; while, slightly displaced from it, would be seen a violet image (vvvv in figure 9, vv in figure 8). Any other color, such as green, would also form an image of the window, displaced less than the violet but more than the red. (Note that, owing to a certain distorting action of the prism, which we shall not here attempt to explain, the vertical edges of the images appear not straight, but curved, as shown by the dotted lines of figure 9). Now, if we bear in mind that there exist in the white light, not three colors, nor only seven, but an infinite number of gradations shading into one another, each of which produces its own image of the window, it is easy to see that all of them will overlap in the middle, so that this part will be white, just like the light as it enters the window. But on passing from the middle toward one edge, we find first the violet missing, then the colors nearest to violet (blue-violet, blue, etc.), until finally, at the extreme edge, only the red is present. On the other hand, passing from the middle toward the other edge, first the red is missing, then the intermediate colors, and at the extreme edge only the violet remains. It is clear that only the extreme colors, red and violet, are seen *pure*, that is, unmixed with other colors, because all the intermediate ones overlap. But it is also evident that the overlapping would be very much reduced if, instead of a wide window, a very narrow *slit* were used for the admission of the light. Since it is impossible to use an *infinitely* narrow slit, there will still be a small amount of overlapping of the images produced by shades of color very close to one another, but none at all in the case of distinctly different colors. This can be understood clearly if the reader will imagine each of the rectangles of figure 9, rrrr, gggg, vvvv, etc., to be made much narrower, without changing the distance between their centers.

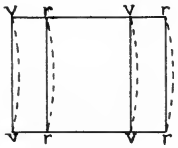

Figure 9

Any person possessing a prism can try this experiment for himself, by allowing light to stream through the crack in a door left slightly ajar, and viewing the crack through the prism held before the eye as in figure 8, with the refracting edge vertical. A band of color will be seen, shading from violet at one edge, through blue, green, yellow, and orange, to red, at the other. The crack in the door acts as a slit, and if this be narrow enough very little overlapping will occur and white will nowhere be seen.

Newton's procedure was really somewhat different from the experiment outlined above. He allowed a beam of light direct from the sun to pass through a small hole O in a shutter (figure 10) and then through a prism P, which deflected it toward the white screen S. He could have placed his eye at the point E, and by looking into the prism, seen the colored band in the apparent position r′v′, since the red light would then have entered his eye as if it came from r′, the violet as if it came from v′, and the intermediate colors as if they originated at points intermediate between r′ and v′. Instead of doing this, he allowed the light to proceed to the white screen, forming a red spot at r, a violet spot at v, etc. Since the pencil of light coming from the sun through a small hole is rather narrow, including an angle of only about one-half degree, there was not much overlapping of the colors, and the whole colored band showed the intermediate, as well as the extreme colors, fairly pure. Newton called this band of color a *spectrum*, and in technical language it is further defined as a *real* spectrum because the light actually passes through it, or at least *to* it, as distinguished from the so-called *virtual* spectrum, seen in the apparent position r′v′ when one looks into the prism. The light does not actually pass through r′v′, but merely enters the eye *as if* it came from there. There is much less overlapping of colors in the virtual than in the real spectrum, in Newton's experiment, that is, the former is more

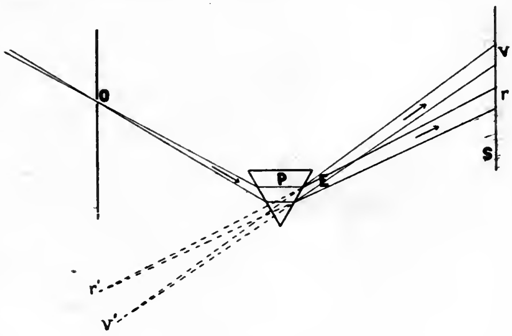

Figure 10

pure. The overlapping in the virtual spectrum can be almost entirely eliminated by making the hole through which the light is admitted very small.

The width of the spectrum can be increased by replacing the round hole with a narrow slit, and the overlapping of colors in the real spectrum can be much reduced by the insertion, either before or behind the prism, of a lens of suitable focal length, so placed, that each of the colors is brought to a focus on the screen. Under these circumstances, it is possible to regard the real spectrum as made up of an infinite number of images of the slit, side by side, the color of each image being slightly, though imperceptibly, different from that of the image next to it. If the light coming through the slit contains every conceivable gradation of color, as is the case with light coming from an ordinary electric lamp, there will be no gaps in the spectrum. In the case of sunlight there are certain missing shades, and these defects are made evident, if the slit is very narrow and the focussing very good, by certain gaps, or "black lines" across the spectrum, in the positions of those images of the slit which would be supplied by the missing colors if they were only present. (see section 56)

8. Newton's conception of color.—Newton's conception of the formation of the spectrum, then, was that the different colors are already present in the white light, and the prism serves only as a separator. This view was directly opposed to that held by some others, that the prism in some way modifies the light so as to change it from white to colored. According to Newton's ideas, if a second prism be introduced behind the first one, but with its refracting edge turned in the opposite direction, as in figure 11, this should reunite the colors into a white beam again. A trial shows that this actually occurs, provided the second prism is of the same kind of glass, and has the same angle. Another test of Newton's theory is this:

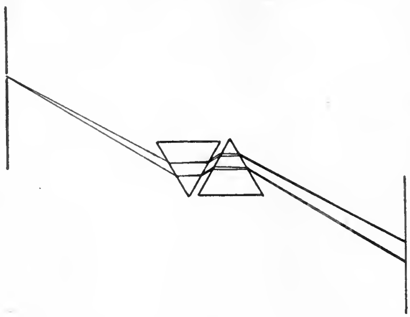

Figure 11

If we could pass through a prism only light of a single spectral tint, for instance deep red, or a definite shade of any other color, the prism should simply bend it, and not separate it into more colors. This can be tried by the arrangement shown in figure 12. The white screen, S of figure 10, is replaced by a screen in which there is a narrow slit, through which any single portion of the spectrum can be passed to a second prism. It is found that the second prism does bend the light, but does not spread it out into any more colors; that is, if only red light enters it, only red light leaves it, and that without being spread out to any appreciable extent. For these experimental reasons, we shall accept Newton's view of color as being the correct one.*

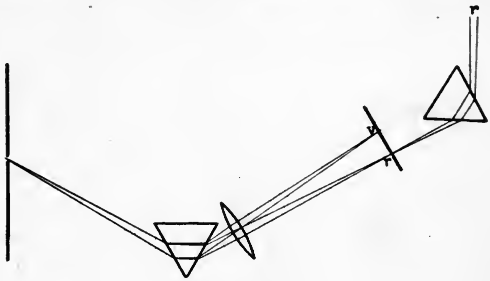

Figure 12

9. Impure colors—There is one fact about the spectrum which can hardly fail to strike one who examines it closely, viz., that we fail to find in it certain colors which are more or less common in nature. For instance, there is no purple, and though there are several shades of red, there is none that could be called pink. These two cases are good illustrations of the general fact that most of the colors of nature are not "pure colors," in the sense of spectral colors that cannot be further separated by a prism, but are mixtures of two or more of these latter. Purple, for example, is a mixture of red with blue or violet or both. This fact can be shown by placing two small

*Certain modern investigations show that the contention of Newton's opponents, that a prism actually manufactures the different colors from white light instead of merely separating out constituents that are already present, is in a certain sense true. This is embodied in what is called the *pulse-theory* of white light. But, after all is said, this differs from Newton's theory only in the point of view, and the latter not only explains all the phenomena in a satisfactory manner, but is much easier to deal with. Therefore we may accept it as true in the pragmatic sense.

mirrors in the spectrum, one in the deep red part, the other in the blue or violet part, turning them so that each reflects to the same spot on a white screen. This spot, illuminated by both red and blue or violet light, appears purple. Purple, then is a sensation produced when both red and blue or violet light fall upon the retina of the eye.

We can get a pink spot by illuminating a white screen simultaneously with white light and red light. Consequently, pink is produced by a combination of all the spectral colors, with a considerable excess of red. Similarly, pale blue is a combination of all the colors with an excess of blue, pale green a similar combination with excess of green, etc. In technical language, any such combination of one or more definite spectral colors with white is said to be *unsaturated*. Thus, pink is unsaturated red, pale blue is unsaturated blue, etc.

That the pigments used ordinarily in painting are very impure colors, can be demonstrated by a very simple experiment. Take a strip of paper about 1/32 inch wide, and color it in quarter-inch lengths with the following artists' pigments, each pair of colors being separated by a short length of white: alizarin crimson, alizarin crimson mixed with gamboge, gamboge alone, gamboge mixed with prussian blue, new blue, and alizarin crimson mixed with new blue. The strip will then appear as a very narrow ribbon showing the following succession of colors: red, white, orange, white, yellow, white, green, white, blue, white, violet. Now lay it against a dull black background, such as a piece of black felt, illuminate it with sunlight, and look at it through a prism whose edges are parallel to the strip. Each of the white portions forms a complete spectrum, with which the spectra of the painted portions can be compared. It will be noted that each of them shows through the prism not only the color which it appears to have when viewed directly, but certain other parts of the spectrum. Not one of them is a pure color, but at best each shows only a strong *excess* of the color it is meant to have. The most nearly pure of all is the red part, but even it shows quite a little green, with traces of the other colors.

A narrow blade of grass, when observed through a prism in such a manner, shows, beside strong green, a great deal of red and yellow, and even some blue and violet.

10. Color due to absorption.—Naturally we are led to enquire: why is a blade of grass green, or the petal of a rose red? Since grass is not itself luminous, but is seen only because it reflects diffusely the sunlight that falls upon it, its green color, or rather its very mixed color with green predominating, must arise from the fact that some of the chemical substances in the grass have the property of "absorbing," to a greater or less extent, certain colors, or as we often express it, certain *parts of the spectrum.* For instance, these constituents of the grass either do not absorb green at all, or more likely simply absorb it less than they absorb red and yellow, and still less than they absorb blue and violet. But absorption, in the proper sense of the word, can occur only while the light is actually passing through a material, not in the mere act of reflection at the surface. Consequently, it must be that the light penetrates the surface to some appreciable depth, that is, the substance of the grass is to some extent transparent. That this is true, can be readily proved by holding a blade of grass between the eye and a bright source of light. A considerable fraction of the incident light passes entirely through the grass to the eye. Indeed, it may truly be said that any material is to some extent transparent, and if made into a thin enough sheet will allow an appreciable amount of light to pass through it. But the farther the light is caused to pass through a material the more of its energy is absorbed.

Evidently, what happens in the case of the blade of grass is something like this: Of the white light that strikes the surface, part is reflected without penetrating, as would be the case with glass or water; and this part, if it could be seen alone, would be white, like the incident light. The rest of the incident light penetrates the surface; but since the material is not completely transparent, but only what we call translucent, the rays do not pass straight through to the back surface, but are diffused, or scattered, within the material. In this way part of the light eventually gets back into the air through the front surface, and we call this part "diffusely reflected" light, although it has been within the body of the material. Part also gets out through the back surface, and we call this "transmitted" light. Both the diffusely reflected, and the transmitted, light, during its passage through the material, suffers losses by absorption in the chemical substances of the leaf. Part of the green is absorbed, more of the yellow, orange, and red, and most of the blue and violet. In this way, both the reflected and the transmitted light become colored, and indeed both parts show about the same color.

Most natural objects owe their colors to the same cause. Part of the light penetrates the surface, and part of this emerges again, after suffering absorption. This explanation is satisfactory so far as it goes, but it must not be forgotten that we do not know why, for instance, the leaf-substance absorbs more red than green, while just the reverse is true of the coloring-material known as alizarin-crimson. This question cannot be answered without a far greater knowledge of atomic structure than we now have.

11. Color due to other causes.—There are some objects whose color is produced in a different way. The yellow color of gold, as an example, is mostly due to the fact that a certain shade of yellow light seems unable to enter the gold at all. It is completely reflected at the surface. Consequently, a very thin sheet of gold-leaf transmits light which is completely lacking in this color. The transmitted light is dull green, while the reflected light is yellowish. A similar phenomenon occurs in the case of some dyes, which in concentrated form show quite a different color according as they are seen by transmitted or by reflected light. Common red ink is an example; it reflects green very strongly when concentrated, but transmits red.

The brilliant colors of rainbows are due, not at all to absorption, but to a separation of the colors something like that which occurs in a prism. The theory of rainbows will be taken up later, (section 83).

The blue of the sky is caused by a sort of scattering of the light by the particles of the atmosphere, similar to the scattering produced when a beam of light is sent through milky water. If there were no atmosphere the sky would appear black, and we would receive light only from the sun, moon, and stars directly. In the scattered light from the sky, all colors of the spectrum are represented, with blue in excess. In order to explain this preponderance of blue, we must anticipate to some extent facts that properly come later in these pages. It will be shown in the next chapter that light consists of waves, the shortest of which are the violet, the longest the red, the intermediate colors having waves of intermediate length. Although all these waves are very short, the length of even the shortest is much greater than any of the dimensions of a molecule. Since the molecules are so small, they are much more efficient in reflecting, or scattering, short waves than longer ones, just as small pieces of wood floating on the surface of water will reflect short ripples, but simply ride on the very long waves. Therefore the molecules in the air reflect—scattering in all directions—those colors that lie near the violet end of the spectrum to a considerably greater degree than those near the red end, thus giving a bluish color to this scattered light. That it appears blue rather than violet is because the violet is at best very weak.

Of course, since a beam of direct light from the sun is thus robbed of a greater percentage of its blue and violet than of its red, that part of it which passes on through must be abnormally rich in red—relatively speaking—and therefore must appear more reddish in color than it was when it emerged from the sun. This is particularly marked when the light has passed through a long distance in air before reaching the eye, as is the case near sunrise or sunset. That the light at such times is exceptionally impoverished in blue and violet, is well known to every photographer, for most of the photographic action of light is produced by these colors, and a plate must be exposed several times as long when the sun is low in the sky as at midday. However, the reddish color of the sun when it is near the horizon is familiar to all.

12. Black and white.—A black object, strictly speaking, is one which absorbs completely all colors, and reflects none. But an object may appear black simply because the light which illuminates it contains no constituent save those which it absorbs completely. For example, a deep red rose will appear black when placed in the blue or violet part of the spectrum, because it absorbs these colors completely, and the only color which it can reflect freely, red, is not present.

A *white* object is one which reflects diffusely, that is in all directions, all colors to the same extent. The whiteness of the snow is an interesting case. Snow is really composed of numerous little ice crystals, and ice in bulk is not a white body, but a transparent one. That is, a large chunk of ice allows most of the light which falls upon it to pass through, and that part which it reflects is reflected, not diffusely, but in a definite direction. But when we have a great mass of very small ice crystals, arranged in an irregular manner, although each crystal surface reflects in a particular direction, the whole mass reflects about as much in one direction as in another. Moreover, the light that passes through the crystals on top of the layer of snow will strike other surfaces below, which again cause reflection, part of the reflected light finding its way out of the mass again. Thus the whole mass reflects irregularly a very large amount of the light, and, since no color is absorbed by the material, this reflected light is white in color. Therefore, the whiteness of snow is due to the great number of reflecting surfaces, irregularly arranged. The same effect can be produced by crushing a piece of glass with a hammer. The many cracks in the glass cause a multitude of reflecting surfaces, arranged irregularly, and the mass immediately becomes white. The whiteness of clouds is also due to numberless little reflecting surfaces, the surfaces of millions of small water-drops.

Of course, a white object will not *appear* white unless the light which falls upon it contains all the colors of the spectrum, i. e., is white light. If it be illuminated by red light it will appear red, if by blue light, blue, etc. Thus, the white screens of figures 10 and 11 show whatever color falls upon them, and it is just this property that makes a white screen suitable for such experiments.

13. Complementary colors and color mixture.—Any two colors which together produce the sensation of white are called complementary colors. A convenient way of showing these is illustrated in figure 13. White light passes through the slit S to the lens L_1 which makes the rays parallel. It then passes through the prism P and the second lens L_2, which focusses the spectrum in the plane vr. Instead of having a screen at this place, another lens, L_3, is placed just behind it. The two lenses L_2 and L_3 together form an image of the face of the prism on a properly placed white screen A. Since light of all colors comes through the whole face of the prism, and all the light passes through the two lenses, this image will be uniformly white. Now, if an opaque obstacle be placed just in front of L_3, i. e., just in the plane in which the spectrum is formed, the obstacle being of sufficient width to cut off a certain spectral region, say the green, then only the remaining colors will reach the screen A, and the image of the prism-face will therefore show the color complementary to the color cut out.

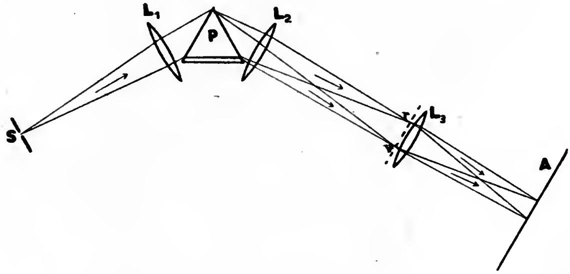

Figure 13

By this means, we find that the color complementary to the average spectral green is a peculiar shade of red, that complementary to spectral blue a sort of golden yellow, etc. Generally, one or both of a pair of complementary colors are impure in the spectral sense.

It is found that a combination of the red, the green, and the blue of the spectrum, in suitable proportions, will produce the sensation of white, without the presence of the other colors, such as violet, orange, and yellow. This can be tried with the arrangement of figure 13 by placing before the lens L_3 a card with holes cut through it so as to let pass only some of these three colors. The best exact location of the holes, and their proper sizes, can be found only by trial. Furthermore, by suitably altering the relative intensities of these three colors, as by stopping down one or two of the holes, any other color, either a pure spectral hue, or such a color as purple, can be closely imitated.

The mixing of pigments shows some results which at first sight are very surprising. For instance, since a combination of all the different colors, of proper proportions, produces white, we should naturally expect that when many different paints are mixed, the mixture would tend to become white.

On the contrary, it tends to become black; and in general, the more different pigments are put into a mixture, the darker it becomes. The reason is that by mixing we combine the *absorbing* powers rather than the *reflecting* powers of the constituents. We have seen that the common pigments, alizarin crimson, gamboge, and new blue, have each a distinctive absorption, and in a mixture of the three any part of the spectrum would be strongly absorbed by one or another, so that little if any of the incident light would escape absorption. An example less extreme than this is seen in the mixture of the crimson and the blue to produce violet. If it were not for the fact that certain parts of the spectrum escape complete absorption in either of these pigments, the mixture would be black instead of violet. It is, as a matter of fact, extremely dark, much darker than the violet seen in the spectrum from direct sunlight, relative to the brightness of the other colors.

It is quite plain then, that the mixing of two or more paints produces quite a different result from throwing simultaneously upon a white screen lights of the corresponding colors; and this fact seriously limits the ability of artists to produce desired effects by mixing paints. The school of painters known as impressionists introduced a new method. Instead of mixing their paints, they lay them on the canvas in little blotches side by side. Thus, where a painter of the older schools would mix crimson and yellow to paint a surface of orange color, the impressionist covers the surface with dots of crimson and dots of yellow close together but arranged in irregular order. Such a painted surface looks very confusing when viewed at close range, but at a greater distance the blotches of red and yellow seem to blend together to produce the effect of a uniform orange, so that the impressionist secures in this way the same effect that we could get in the laboratory by simultaneously illuminating a white surface with red and yellow light. As a result, paintings by impressionists are usually far more brilliant than those of the old masters, though it is true that the latter have a sombre richness which is itself a great charm.

14. The eye.—The organ of vision, the eye, is an optical instrument more analogous to the photographic camera than to anything else. It is shown diagrammatically in figure 14. It consists of a shell roughly spherical in shape, of which the front wall, C, the cornea, is transparent. Behind this is the iris I, a screen or diaphragm containing a circular hole,—the pupil, P,—whose diameter contracts in brilliant illumination or expands in dim light, by involuntary muscular action. The lens L is capable of a slight forward and backward motion, like that of a camera lens in focussing, but most of the focussing in the eye is accomplished by altering the radii of the lens surfaces. The material of the lens is of course somewhat plastic, and in structure it resembles an onion in being built up in layers. The lens forms an image of any object looked at upon the retina R, which is spread over the rear and side surfaces of the shell. The space between the lens and the retina is filled with a jelly-like material called the *vitreous humor*. The material between lens and cornea is watery, and is called the *aqueous humor*. By means of the above mentioned muscular distortion of the lens, together with its slight axial movement, any object toward which the eye is directly turned can be sharply focussed upon the retina, provided it is not closer than a few inches, in the case of normal eyes. This adjustment of focus is called *accommodation*. With the eye relaxed, very distant objects should be in sharp focus. In viewing a very small object, however, it is advantageous to bring it close, so that its image upon the retina may be larger; but if it is brought closer than about 10 inches the muscular strain of accommodation becomes unpleasant and is in fact harmful. Consequently, about 10 inches (25 cm.) is the most favorable distance, with people of normal vision, for reading or for careful scrutiny of small objects. This is known as the "distance of distinct vision."

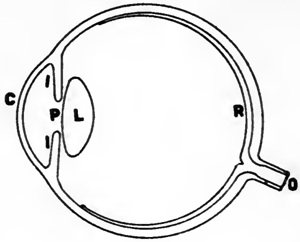

Figure 14

A *myopic*, or short-sighted, eye is one whose focal length in the relaxed condition is abnormally short, so that it cannot focus sharply upon distant objects. Just the reverse is a *hypermetropic eye*, which requires a certain amount of accommodation even to focus upon an infinitely distant object. *Presbyopia,* a trouble particularly common among older people, is an impairment of the ability to accommodate, due to a progressive stiffening of the muscles which change the radii of the lens. *Astigmatism* is caused by lack of axial symmetry in the lens or the cornea or both. It shows itself in an inability to see clearly, with the same accommodation, lines inclined at different degrees with the vertical, though equally distant.

The focus upon the retina is sharp only for a limited region near the axis of the lens, but unsharp vision is possible over a very wide angle without moving the eye.

The retina is the sensitive part of the eye, which in some manner is stimulated by light falling upon it so that the mind experiences the sensation of vision. It is a network of delicate nerve-fibers which are connected with the brain through the optic nerve O.

15. Color vision theories.—The student should understand that the actual mechanism of vision, the connecting link between the light-stimulus upon the retina and the consciousness of light and color, is a thing about which little is known. Even if we knew the physical and chemical processes that go on in a nerve, there would still be a gap or hiatus in our knowledge between that and the actual sensation. Consequently, our notions of light and color perception do not extend very far, and are somewhat uncertain at that. On the face of things, it seems very unlikely that there is a separate type of nerve for every gradation of color. Such an experiment as the production of the orange sensation by mixing red and yellow light, or the production of any other spectral sensation by a mixture in suitable proportions of red, green and blue, suggests that there are only a few distinct color sensations, perhaps three, and that the other sensations are the result of a simultaneous stimulation of these few. Experiments with color-blind people also support this view. There are in fact two principal theories of color-sensation, the Young-Helmholtz theory and the Hering theory. The former alone is often given in physics texts, but the latter is favored by at least a great many experimental psychologists, and the matter is one of psychology more than of physics.

According to the Young-Helmholtz theory, the retina contains three distinct sets of nerve-fibers, each giving only a single sensation, no matter what particular part of the spectrum corresponds to the light that does the stimulating. One set gives a red sensation, the second a green sensation, the third a violet or blue sensation. The three curves of figure 15, which are due to Koenig, show to what extent each of these

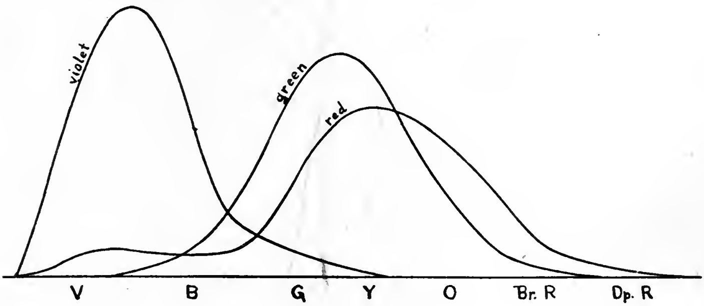

Figure 15

sensations is stimulated by light from different parts of the spectrum. In order to clearly understand these curves, consider how the spectrum would appear to a person whose eyes were provided with the red-sensitive nerves, but not with the other two sets. He would be able to see the spectrum throughout its entire extent, with the possible exception of the extreme violet end, but all of it would appear of the same color, red. The only differences between different parts would be differences of brightness, as indicated by the varying ordinates of the red curve. In fact, his retina would act in a way quite analogous to the action of a photographic plate, which responds to the influence of light of many different colors, but with a response which is the same in kind for all, differing only in degree.

Now consider an eye with all three sets of nerve-fibers, and suppose the retina to be stimulated by light from the yellow-green portion of the spectrum. A comparison of figure 15 shows that this light stimulates all three of the sensations, the green sensation most strongly, the red sensation to a less degree, and the violet sensation least of all. The complex of these three sensations acting together is what we are accustomed to call the yellow-green sensation. About one man in thirty is "red-color-blind," which, on the Young-Helmholtz theory, means that his eyes lack the nerve-fibers which give the sensation of red.

The Hering theory is quite different. Instead of three primary *sensations,* it postulates certain *contrasts,* caused by chemical changes, under the influence of light, in three hypothetical fluids present in the retina, which we shall designate as A, B, and C. Fluid A undergoes a certain decomposition when any sort of light, irrespective of color, falls upon, it, but recombines, or recovers, in darkness. It reacts upon the nerve fibers differently in its two states, causing a sensation of brightness in the one case and of darkness in the other. Fluid B is different. It undergoes a decomposition under the action of light of longer wavelengths, giving a red sensation, and a recombination under the action of light of shorter wavelengths, giving a green sensation, being entirely neutral for light of wavelength corresponding to some part of the yellow. Fluid C acts in a similar way, but the sensation produced by longer wavelengths is yellow, that by shorter wavelengths blue or violet, and the neutral condition would be for wavelengths in the green. Thus we have three contrasts, bright and dark, red and green, yellow and blue. According to this theory, the usual type of color-blindness is due to lack of the B fluid, resulting in an inability to distinguish reds from greens.

Color-blindness is not a disease, but a heritable defect, and though a handicap it is not a thing of which one need be particularly ashamed. Many people who have it are not conscious of it. Recent biological researches have shown the following interesting peculiarities about its inheritance. A color-blind man transmits the defect neither to his sons nor to his daughters, but to the sons of his daughters; that is, it passes from the male of the first generation to the male of the third, through the female of the second, but without showing actively in the second generation at all. A woman is never herself color-blind unless she inherits it both from her father and her mother's father. Consequently, cases of color-blindness among women are very rare.

Problems.

1. Suppose that a flower whose color is a *pure* blue is passed slowly through a spectrum, from one end to the other. What would be its appearance in the different parts? Suppose a blade of grass were treated in the same manner?

2. A story by Ambrose Bierce, entitled "The Damned Thing," has for its subject a supposedly invisible animal. The author argues that such a thing would be possible if the animal's fur reflected only *ultraviolet light*. What would be the actual appearance of such an animal?

3. Explain the whiteness of soapsuds and other froth.

4. Suppose blue glass were crushed to a powder. What would be the effect upon its color?

5. Explain why tobacco smoke appears blue against a dark, but brown against a bright, background.

6. Why is it that colored cloth can be changed by dyeing to a darker, but not to a brighter color?

7. Birds, animals, and fishes usually have a much lighter color on the lower than on the upper sides of their bodies. Is this fact of any importance in the economy of nature? Explain.

CHAPTER III.

16. The corpuscular theory of light.—17. The wave theory.—18. Bending of light into a shadow.—19. Nature of the ether.—20. Waves in general. Plane waves.—21. Mathematical formula for a wave.—22. Interference. Fresnel's mirrors.—23. Interference in white light.

16. The corpuscular theory of light.—We have now learned enough about some of the general properties of light to enquire with some degree of intelligence as to its nature. The two theories that have had any support may be called the *corpuscular theory* and the *wave theory.* According to the first, light consists of very small weightless material particles; according to the second, it consists of waves. Either theory strains the imagination greatly.

It is hard to think of material corpuscles flying with enormous speed through a solid substance like glass, with so little hindrance as glass seems to offer to the passage of light, though color might well be accounted for by differences in size, in shape, or in some other characteristic among the corpuscles. It is also extremely difficult to explain how, when these corpuscles strike such a substance as glass or water, some of them should be reflected while others pass into the material, being refracted as they do so. It is true that one might suppose that there are two kinds of such particles, a kind that is reflected and a kind that is refracted. But if this were the case, one reflection would completely separate these two kinds, so that if the *reflected* light struck another such surface all of it would be reflected, none refracted; while if the *refracted* light struck another surface, all would be refracted, none reflected. Then if one should observe in a plate glass window the reflected image of his own body, this image would become invisible to him if he held a small piece of glass in front of his eyes. A simple trial shows that this is not true. Furthermore, the light that got in through the first surface of the plate glass could not be reflected at all by the second surface. It can easily be proved that this conclusion also is false, for if one stands close to such a window he can distinctly see two images of himself, the brighter image being formed by reflection at the first surface, the fainter one at the second. (In reality there is a large number of such images, produced by multiple reflections, but all except the first two are very faint). In order to get around this difficulty, Newton, the chief advocate of the corpuscular theory, suggested that, although all the corpuscles are fundamentally alike so far as reflection and refraction are concerned, each one is at times in a state suitable for reflection, at other times in a state suitable for refraction; so that whenever the light strikes a reflecting surface there will be a certain proportion of them ready for reflection, even though some had already been reflected before. But such an hypothesis, though not absolutely absurd, seems clumsy and improbable, and Newton himself was far from being satisfied with it.

17. The wave theory.—In the wave theory, there is no difficulty in explaining reflection and refraction. Indeed it is characteristic of all kinds of wave motion that, whenever a wave strikes a surface separating two media in which the velocity of wave propagation is different (such as the surface between air and glass) part of the energy enters the second medium as a refracted wave, while part is sent back into the first as a reflected wave. Neither is it at all hard to think of waves passing through glass and other transparent media with high velocity and little resistance, for we know that mechanical waves, such as sound waves, do traverse such bodies very easily. Color, also, may be accounted for very simply on the wave theory, by the supposition that differences in color correspond to differences in the length of the waves, just as we know that differences in the pitch of musical notes correspond to differences in the lengths of the sound waves.

But it is difficult to understand how waves can pass, as we know that light does, through perfectly empty space, for the use of the name "wave" implies the existence of some medium in which the waves exist. On account of this difficulty with the wave theory, physicists have been led to assume the existence of a medium filling all space, even a so-called vacuum, to which the name *ether* has been given. The necessity for this assumption is to this day a very serious load on the shoulders of the wave theory of light, though it becomes less objectionable when we find that there are other phenomena, such as electric and magnetic attractions and repulsions, which also operate through a vacuum, and also seem to indicate the existence of some all-pervading medium.

Newton's chief objection to the wave theory, however, was not the necessity for an ether, but the fact that light apparently travels in straight lines, while other waves, such as sound waves, or waves on the surface of water, will bend freely round an obstacle placed in their path. In this connection, it should be noted that the degree to which waves bend round an obstacle, thus departing from the straight-line path, depends to a great extent upon the length of the waves. Long waves do so much more freely than short waves. Therefore, if it can be shown that light really does to some slight extent bend round a corner, this objection will be overcome, and at the same time evidence will be acquired that if light does consist of waves these waves are very short.

18. Bending of light into a shadow.—In fact, the bending of light about an obstacle can be shown, by a very simple experiment illustrated in figure 16. O represents any opaque obstacle with a straight edge perpendicular to the plane of the paper. The source of light, S, must be of very small dimensions; otherwise the bending of the light into the shadow, which the experiment is designed to show, will be masked by the penumbral effect always shown when a shadow is cast by a fairly large source of light. It should be either a very fine hole, or better still a very narrow slit, illuminated by an arc light or other brilliant illuminant. A is a white screen, to receive the shadow. Let a straight line SQ be drawn from the slit through the edge of the obstacle to the screen. Then if light were absolutely rectilinear, all parts of the screen above Q would be fully illuminated, all points below in complete shadow. There would be a sharp and abrupt division between the illuminated part and the shadow.

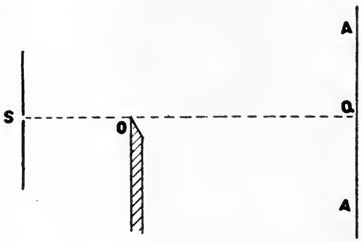

Figure 16

But, as a matter of fact, the light is found to shade off continuously, though rather rapidly, into the shadow; and above the point Q there are a number of bright and dark bands, parallel to the slit and to the edge of the obstacle. The gradual shading off of the light, merging into the shadow, disproves the rectilinear propagation of light, except as an approximation to the truth. As to the bands, it will merely be noted here that they are readily explainable on the wave theory, the actual explanation being deferred to later pages. (See section 70)

Probably the crucial reason for the discarding of the corpuscular theory and the definite adoption of the wave theory is the following. When a beam of light strikes the surface of glass or water at an angle, it is bent *toward the normal,* that is, it makes a more acute angle with the perpendicular within the glass than in the air. In order to explain this, the advocates of the corpuscular theory were obliged to assume that the corpuscles are attracted by the glass when they get very close to it, leading to the conclusion that they travel faster in the glass than in the air. On the other hand, the wave theory explains this bending of the rays at the surface very easily with the assumption that the waves travel *slower* in glass or water than in air. Thus the phenomenon of refraction furnishes the occasion for a definite clash between the two theories, one demanding that light travels *faster* in the refracting medium than in air, the other that it travels *slower.* During Newton's life, no means of actually measuring the velocity of light in such a medium as glass or water was known; but after Foucault had devised the rotating mirror method, the velocity in water was measured by filling a long tube, fitted with glass ends, with water, and inserting this in the path of the light. The experiment showed without any possibility of doubt that light travels *slower* in water than in air.

Since this experiment definitely discredits the corpuscular theory, and since the only outstanding objection to the wave theory is the hypothesis of the ether, for whose existence we have additional evidence from electric and magnetic phenomena, it is now accepted by physicists that light consists of very short waves in a hypothetical medium extending through all space. Even when light passes through a material like glass or water, we regard the ether as the carrier of the waves. We may think of the molecules of the glass as existing in the ether. Their presence modifies the transmission of the waves, partly by changing the wave-velocity, partly by absorbing part of the wave-energy, which they change into heat-energy, for heat is always produced when light is absorbed. A rough image of the state of affairs may be gotten by considering waves on the surface of water on which are floating many pieces of wood. The water represents the ether, the pieces of wood the atoms or molecules of the material substance.

19. Nature of the ether.—It is difficult to speculate about the nature of the ether. There are no reasons for believing that it is atomic or molecular in structure, and it seems to offer absolutely no resistance to the passage of bodies through it. A philosophically minded person might ask the question: Just what do we mean when we say there is an ether? Such a question is worth while because it leads us to take stock of our knowledge; but probably the only answer that can be given is this; if the wave theory of light is true, (and it is supported by too many facts now for us to doubt it) then the statement that there is an ether means only that empty space has properties other than mere extension, properties that enable disturbances carrying energy to pass through it, the passage requiring finite time. Whether we say there is an ether, or that empty space has properties other than those of pure geometry, matters little, but the name "ether" is a convenient one to symbolize these properties, and we shall hereafter use it in this sense.

We shall see that it is comparatively easy to devise experiments for measuring the wavelength of light, and that the measurements can be carried out to a very high degree of precision. (Chapter VII and IX) It will be easy to explain the laws of refraction from the fact mentioned above, that the velocity of light is less through such a substance as glass than through air, somewhat less through air than in the free ether. (Chapter IV) We can explain the formation of a spectrum by a prism, by showing that through glass the velocity of shorter waves is less than that of longer ones.

But when all this done, it must not be forgotten that we shall still be much in the dark as to many important questions. In the first place, are these waves longitudinal, like sound waves, or transverse, like those of a plucked string? Are they purely mechanical waves, or do they consist of rapidly alternating electrical disturbances, or are they disturbances of a kind unrelated to any other physical phenomena? What is the nature of the phenomena going on within the atoms of a substance which is emitting light, or in those of a substance which is absorbing it? How and why does the presence of material molecules, in such a substance as glass or water, change the velocity of the waves in the ether surrounding and permeating them?

Some of these questions have been solved satisfactorily, some have been only partially solved, others are still open. It is clear that if we are ever able to answer all of them we shall know a great deal about the inner structure of the atom and the molecule; and since the acquisition of such information is one of the principal aims of physical research, the study of light becomes one of the most important branches of science.

20. Waves in general. Plane waves.—Before going on with the study of light as a wave-motion, we shall devote some space to the consideration of waves in general. We find cases of waves which advance (1) in only a single dimension, like those that travel along a stretched rope if it is struck or moved in any way, (2) in two dimensions (i. e., spreading over a surface) like water waves, and (3) in three dimensions, like sound waves, light waves, or the waves that would spread through a block of jelly if some point in the interior were set vibrating. In the last case, it is clear that the waves would spread out in spheres with the point of origin as center, the direction of advance being along the radii. Under certain circumstances, however, we could have cases where they advance in planes, the direction of advance being perpendicular to these planes. There would be an approximation to this condition, for example, at a very great distance from the point of origin, for a small section of a sphere of very great radius is nearly plane; but such waves would be feeble because the energy initially given to them would be spread over a great surface. In the case of light, as we shall see in Chapter V, such *plane waves* can be produced by the use of mirrors or lenses, without such a weakening of intensity.

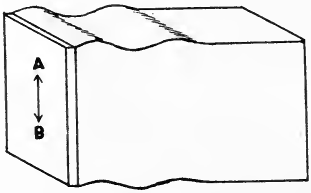

Figure 17

The nature of plane waves may be understood by imagining a block of jelly as in figure 17, to one face of which is attached a rigid board that is moved harmonically back and forth in its own plane, in the direction of the line AB. At any instant, each point in a plane parallel to the board will be in the same condition of motion, or, as we say, in the same *phase.* In general, any such surface, every point of which is in the same phase, is called a *wave-front,* whether it is at the beginning of a train of waves or not. In figure 17, every plane in the jelly parallel to the board is a wave-front, and in the case of waves coming from a point-source, every sphere with its center at the source is a wave-front.

We may have cases of a single wave, like the sound-wave sent out from an explosion, and also cases where there is a train of waves, such as those sent out by a tuning-fork. The shape of the waves may be simple or complicated. For any medium through which waves of all lengths travel with the same velocity (as is true for the free ether with light-waves) a disturbance of any kind will travel onward without changing its shape, whatever that shape may be. For example, by giving the board of figure 17 a suitable motion, waves of the form shown in figure 18 could be made to travel through the jelly.

A mathematical theorem due to Fourier proves that any such periodic form can be made up of a series of simple sine

Figure 18

and cosine forms, of different wavelengths. For this reason. we are compelled to make a special study of waves of these simple forms.

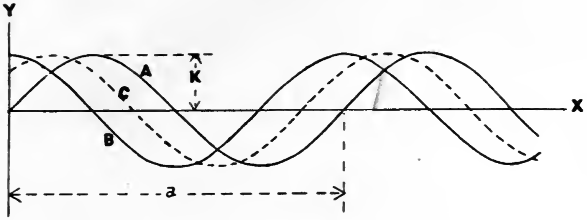

Figure 19

21. Mathematical formula for a wave.—In figure 19, A represents a sine-curve, whose equation is

$$y = K.\sin\frac{2\pi x}{a} \qquad (1)$$

B a cosine-curve, whose equation is

$$y = K.\cos\frac{2\pi x}{a} \qquad (2)$$

and C a third curve whose equation is

$$y = K.\cos\frac{2\pi}{a}(x - \alpha) \qquad (3)$$

Evidently, all three are exactly the same in form, and either can be transformed into one of the others by shifting it along the axis of x. In fact, we can represent either of the three curves by the formula (3) provided α is given a suitable value, different in each case. If $\alpha = 0$, we have the simple cosine curve. If $\alpha = a/4$ we have

$$y = K.\cos\frac{2\pi}{a}\left(x - \frac{a}{4}\right) = K.\cos\left(\frac{2\pi x}{a} - \frac{\pi}{2}\right) = K.\sin\frac{2\pi x}{a}$$

the equation of the sine curve. Whatever value α may have, its presence indicates that equation (3) represents a simple cosine curve shifted in the positive direction of x by the distance α. For y has the same value for x in (3) that it has for $x - \alpha$ in (2).

Equations (1), (2), and (3) do not represent waves, but only stationary curves. For such a curve to become a wave, it must progress steadily to the right (or left). Since the symbol α indicates a shift to the right, (3) can be changed to represent a wave instead of a stationary curve if α is replaced by a term containing the *time,* which term indicates the movement. If t represents the time and V the velocity of the waves, then in t seconds the curve must be shifted along the x axis a distance Vt. Therefore, putting Vt for a, the wave equation is

$$y = K.\cos \frac{2\pi}{a} (x - Vt) \qquad (4)$$

a is called the *wavelength,* for it is evident from figure 19 that it is the distance from crest to crest, or from trough to trough. The quantity K, called the *amplitude,* gives the maximum value y can have. Equation (4) is an equation of three variables, y, x, t. x and t are called *independent* variables, y the *dependent* variable. The equation gives the value of y for any stated distance x from the origin and for any stated time t.

It is often desirable to introduce another constant term, ϵ, within the parenthesis, making the equation read

$$y = K.\cos \frac{2\pi}{a} (x - Vt - \epsilon) \qquad (5)$$

The only difference between (4) and (5) is that the former represents a wave which, at time zero, has the position B of figure 19, while the latter is one which at that instant has the position A, B, C, or any other, depending upon the value of ϵ. This quantity ϵ is called the *phase-constant,* the whole quantity whose cosine is to be taken, viz.,

$$\frac{2\pi}{a} (x - Vt - \epsilon),$$

being known as the *phase.*

A wave advancing to the *left* would be represented by equation (4) or (5) with the sign of V changed; thus,

$$y = K.\cos \frac{2\pi}{a} (x + Vt - \epsilon)$$

So far we have regarded y as a real mechanical displacement, at right angles to the direction in which the waves are propagated. In the case of transverse waves in a string, it is indeed just that. The same thing is true of the mechanical waves in the block of jelly, illustrated in figure 17, and in all such cases the curves of figure 19 give a true picture of the contour of the waves at different times.

But there are cases of wave-motion in which this is not the case. For example, if a stiff rod or a stretched metal wire be stroked longitudinally with a rosined piece of leather, longitudinal waves are set up in which the displacement is *parallel* to the direction of propagation of the waves. In such a case we may still, for convenience, plot y at right angles to x, but the graph so obtained is *only* a graph, and not a true picture of the wave contour.

It may be that a wave does not even consist of mechanical vibrations at all. There are cases, for instance, of temperature waves, in which alternations of temperature, above and below a mean value, are sent through a material. It is quite conceivable also that waves should exist which consist of electric disturbances, for instance regions in which there is an electric intensity directed upward, separated by those in which it is directed downward, these alternating regions following one another through space with great rapidity. Since we have no certain knowledge that such electric states in the ether are necessarily accompanied by any real *motion*, of the ether or of anything else, the quantity y in such a case would have to represent the intensity of the electric field at the distance represented by x and the time represented by t. It will be shown later later (Chapter XIV) that such electric waves actually exist, that the waves of wireless telegraphy are undoubtedly such, and that we have convincing evidence that light waves are also of the same nature, differing from the waves of wireless only in length.

But many of the phenomena of waves, such as interference, diffraction, and some of the phenomena of reflection and refraction, would be the same no matter what the nature of the disturbance might be; and therefore it will be convenient, for the time being, to think of light waves as if they were really waves of mechanical displacement. Whether they are to be thought of as transverse, like those in a plucked string, or longitudinal, like those of sound, need not be considered yet, but evidence will be produced in Chapter XII to help us decide between these alternatives.

22. Interference. Fresnel's mirrors.—One of the most convincing proofs of the wave theory of light is the phenomenon of *interference*, in which two separate beams of light annul one another at certain places, producing darkness, and at other places produce a brightness much greater than either alone could cause. The theory of this phenomenon is as follows:

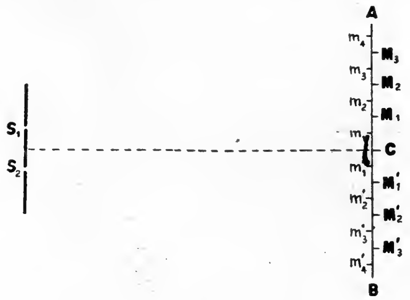

Figure 20

Suppose we have two sources of light, S_1 and S_2, figure 20, in the form of narrow slits or round holes, through each of which comes light of exactly the same wavelength.* At first, we shall also suppose that the two pencils of light are *in phase*, that is, that whenever a crest starts from one a crest will also be starting from the other. This means that, if we should write a formula of the type

$$y = K.\cos\frac{2\pi}{a}\ (x - Vt - \epsilon)$$

for each pencil, ϵ would have the same value in both. It will be easier however to discuss this case without the use of the formulæ.

The light from each slit falls on the white screen AB, and we shall first investigate what happens at that point, C, which is equally distant from the two slits. Whenever a crest reaches C from S_1 a crest will also reach it from S_2, and similarly a trough from S_1 and a trough from S_2 will reach C at the same instant. Consequently, the amplitude of the vibrations at C will be double that which would exist if light from only one slit reached it, and the screen will be very bright there. There will be other points on the screen for which the same statement

*It is customary among physicists to speak sometimes of a small hole or slit through which light is passed as the *source* of the light, although in fact the real source is a flame, a spark, an arc-lamp, or perhaps the sun. This real source is placed close to the slit or hole, or else an image of it is thrown upon the latter by the use of a lens or mirror. The object of the slit or hole is simply to provide a very narrow opening for the light to come through. If the source proper is itself small enough, the slit may be dispensed with.

holds true. For instance, if M_1 is the point just one wavelength nearer to S_1 than to S_2, and M'_1 the point one wavelength nearer to S_2 than to S_1, at each of these points crests will arrive together from the two slits, and also troughs will arrive together, and therefore these too will be points of brightness. The same may be said of M_2, which is two wavelengths nearer to S_1 than to S_2, and of any point on the screen which is an exact whole number of wavelengths nearer to one of the slits than to the other.

On the other hand, consider the points m_1, m'_1, m_2, m'_2, etc., each of which is so situated that it is either ½ wavelength, 3/2 wavelength, or in general any *odd* number of half-wavelengths, nearer to one slit than to the other. Each of these points will receive a crest from one slit at the same time as it receives a trough from the other. In other words, the pencils of light coming from the two sources will at these points be *opposite in phase*, so that they will annul one another, and the points will be dark. It need hardly be pointed out that there will be points, such as one between C and m_1, which will neither be as bright as C nor completely dark as at m_1, since here the two pencils of light meet neither exactly in phase nor exactly opposite in phase. In fact, a moving point, going from C up toward A, would first be in intense illumination, which would fade out to darkness at m_1, then brighten again to a maximum at M_1, fade to darkness again at m_2, etc.

We should expect then, according to the wave theory, to find a number of bright regions, separated by dark regions. In fact, these should be drawn out into bright and dark streaks, or *fringes*, as they are called, perpendicular to the plane for which the figure is drawn. For, even if the slits had no appreciable length in this direction, the loci of bright or dark regions would still be drawn out into lines.

Now let us see how all this would be altered if the two pencils of light were not exactly in phase as they came through the slits. We should still expect to have fringes but they would not occur at quite the same place on the screen. For instance, if the difference in phase were such that a crest would start from S_1 and a trough from S_2 at the same instant, (and vice versa of course) then the places we have marked as bright would be dark, and those we have marked dark would be bright. Herein lies a certain difficulty in subjecting these predictions to experimental proof. The trouble is not so much to keep the two pencils in the same phase, since we don't much care which points are dark and which bright, so long as the fringes stay steady long enough for us to see them. But they will not keep steady unless the two pencils at least keep the same relation to one another in phase, and this they will not do unless they were originally *part of the same pencil or beam*, that is, unless they originated in the same ultimate source. It must not be supposed that any source of light is perfectly steady. We might think of it as sending out a regular train of waves perhaps a meter in length, followed by a break and another train, perhaps longer, perhaps shorter, there being no fixed relation between the phase-constant in the first train and that in the second. In other words, the light comes in bunches of waves, rather than in a long uninterrupted series. Now, if changes of this sort are going on in the light coming from each slit, and the breaks are occurring quite independently in the two pencils, it is evident that, although fringes would be present on the screen at any instant, they would shift their position with every break in either pencil. If we assume that the average length of an uninterrupted train of waves is one meter, then since light travels 300,000,000 meters per second, there would be at least 300,000,000 shifts per second in the positions of the fringes. Consequently, no fringes could be seen, and the screen would appear uniformly illuminated.

Therefore, in order to see such fringes, it is absolutely necessary to get two pencils that have the same origin, so that whenever the phase of one pencil suddenly changes, that of the other will undergo the same change. There are several ways of accomplishing this end, the most satisfactory being one due to the French physicist Fresnel, a diagram of which is shown in figure 21. He used only one slit, S, illuminated by any source of light, but he allowed the beam from S to fall on two mirrors M_1 and M_2, very slightly inclined to one another, each of which reflected light to the screen. The light strikes the screen exactly as if it came from the two images S_1 and S_2, which may be thought of as replacing the two independent slits S_1 and S_2 of figure 20, with the one important difference that now whatever change occurs in one pencil will also occur at the same instant in the other. Thus they will always have the same relation to one another in phase, and the fringes will be steady and therefore visible.

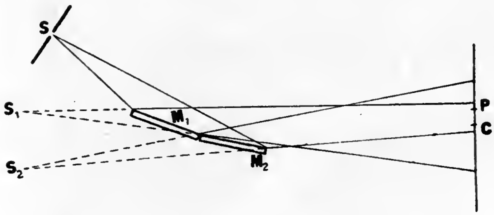

Figure 21

Figure 22

This experiment works very satisfactorily, though the adjustment is somewhat difficult, since the slit must be very accurately parallel to the line in which the planes of the two mirrors intersect. Figure 22 is a photograph of fringes taken by this method. The light used was the violet of a definite wavelength coming from the "mercury-arc."

Referring back to figure 21, let d represent the distance between the two apparent sources S_1 and S_2, D the distance from their plane to the plane of the white screen, C the point equally distant from S_1 and S_2. Let P be any point on the screen, within the plane of the diagram, and x its distance from C. We shall first calculate what values x may have in order for P to be one of the points of maximum brightness, and from this result find the distance between centers of the bright fringes.

If L represents the difference between the distances S_1P and S_2P, P will be a point of maximum or minimum illumination according as L is equal to an even or an odd number of half-wavelengths. Therefore our first step will be to express L in terms of measurable quantities d, D, and x. By simple geometry,

$$S_1P = \sqrt{D^2 + \left(x - \frac{d}{2}\right)^2}$$

$$S_2P = \sqrt{D^2 + \left(x + \frac{d}{2}\right)^2}$$

Therefore,

$$L = S_2P - S_1P = \sqrt{D^2 + x^2 + \frac{d^2}{4} + xd} - \sqrt{D^2 + x^2 + \frac{d^2}{4} - xd}$$

For convenience, we shall represent the quantity $D^2 + x^2 + \frac{d^2}{4}$, which appears in both radicals, by a single term D_0^2. Then

$$L = \sqrt{D_0^2 + xd} - \sqrt{D_0^2 - xd}$$

These two simple radicals can be expanded in series form, by using the binomial theorem, or applying the ordinary rules for the extraction of the square-root. The results are

$$\sqrt{D_0^2 + xd} = D_0 + \frac{xd}{2D_0} - \frac{x^2d^2}{8D_0^3} + \frac{x^3d^3}{16D_0^5} - \text{etc.}$$

$$\sqrt{D_0^2 - xd} = D_0 - \frac{xd}{2D_0} - \frac{x^2d^2}{8D_0^3} - \frac{x^3d^3}{16D_0^5} - \text{etc.}$$

Subtracting the lower from the upper, we get

$$L = \frac{xd}{D_0} + \frac{x^3d^3}{8D_0^5} + \text{etc.}$$

In practise, D is about the order of 100 cm., d about half a millimeter, or .05 cm., and x of course has various values, of which the greatest will perhaps be 1 cm. If we substitute these values, we see that in the first place D_0 will not differ from D itself by more than about 1/20000 cm., so that D may be substituted for D_0. Furthermore, the value of the first term in the last equation comes out to be about .0005, the next term .0000000000015, and the succeeding terms still more minute. Of course, D, d, and x need not have exactly the values here assumed, but the illustration suffices to show that in any case

D is so nearly equal to D_0 that the difference is negligible, and that in the final expression for L only the first term is large enough to measure. Consequently, we shall make no error greater than the unavoidable errors of measurement if we adopt as the correct value for L

$$L = xd/D$$

Using the letter λ to represent the wavelength of the light, we must have, in order that P may be a bright point, that L takes one or other of the values 0, $\pm\lambda$, $\pm 2\lambda$, $\pm 3\lambda$, etc. Or, since $x = DL/d$, P is a point of maximum brightness when x has the value 0, $\pm D\lambda/d$, $\pm 2D\lambda/d$, etc.

This shows us that the fringes are, at least approximately, equally spaced, the distance from center to center being $D\lambda/d$. This distance can be measured with some degree of accuracy, and also the distance D. The remaining distance d is more difficult to measure, partly because it is smaller, and partly because it is not the distance between real slits, but between two images. However, it can be measured by indirect means, and then everything necessary is known in order to calculate the wavelength. It is found, as we should expect, that the wavelength as so determined depends upon the kind of light used. If only deep red light enters the slit, the width of the resulting red fringes indicates the wavelength to be about .00007 cm.; while if deep violet is used the fringes are only a little more than half as far apart as the red, indicating the wavelength of this color to be about .000038 cm. The other colors have wavelengths between these two extremes. But the extremes themselves are not very definite, since it is found that some people can see deeper red or deeper violet than others. This fact makes us suspect the existence of wavelengths longer than the red or shorter than the violet, to which nobody's eyes are sensitive. We shall find later that there are such waves. (Sections 64 to 67)

Although this interference experiment gives us a means of measuring the wavelength of light, it is not an accurate method. More complicated interference experiments, to be described later, allow us to measure wavelengths with an accuracy of 1/1000 of 1%, and in a few cases the precision has been carried even farther.

23. Interference in white light.—Let us consider what would happen if white light, instead of light of only one color, were admitted through the slit in figure 21. Evidently every wavelength would produce its own set of fringes, the spacing being different for each wavelength, and there would be much overlapping of fringes of different color. Only one point, the point C, would be bright for all colors, since it is equally distant from the two slit-images. The central fringe would therefore be white. But the first red fringe on either side of the center would be slightly farther away than the first violet fringe. Consequently, the totality of each of these two fringes would be composed of fringes due to all the different wavelengths, for no two of which would the maxima come in exactly the same place, violet being on the side nearest to C, red on the other side. One might regard each of these fringes as a very short and impure spectrum. It would be white near its middle, with a violet inner edge and a red outer. This effect would be more pronounced for the second fringe on each side, still more for the third, etc., as if each succeeding fringe were a longer and longer spectrum. At the distance of two or three fringes away from C, they begin to seriously overlap; and at the distance of six or eight, the overlapping becomes so complex that all color-effect is lost, the fringes are no longer visible, and the screen becomes uniformly white. When white light is used, therefore, only a small number of fringes are ever visible, whereas with light of a single wavelength, spoken of as *monochromatic* light, a great number may be seen, provided the two beams overlap over a sufficiently extended region. A further discussion of interference in white light will be found in section 82.

Problems.

1. If, in Foucault's rotating mirror experiment, figure 7, a cylinder of some material, in which the velocity of light is less than in air, is inserted between the mirrors M_1 and M_2, what would be the effect upon the distance f_1f_2? What would be the effect if the material were inserted between M_1 and L?

2. Plot to scale, on the same diagram, the curves represented by formulæ (2) and (3), letting $K = 1$, $a = 4$,

$a = 1.5$. (It will suffice to plot only the peaks and troughs, and the points where the curves cross the x-axis) Show by scaling that the curve (3) lies to the right of (2) by the amount predicted in the text.

3. Plot the curve for equation (4), giving any desired value to the time.

4. If the two slit-images of the Fresnel Mirror experiment are 1/10mm. apart, how far away must the screen be to have fringes 3mm. apart, if the wavelength is .00005cm.?

5. Suppose the light coming from the two sources of figure 20 had not the same wavelength. What would be the result?

CHAPTER IV.

24. Reflection and refraction. Huyghens' principle. Index of refraction.—25. Total reflection. Critical angle.—26. Deviation through a prism.

24. Reflection and refraction.—Huyghens' principle. Index of refraction.—The laws of reflection and refraction, so far as concerns only the direction of the reflected and refracted waves and not their intensity, are easy to derive by an application of geometry to the wave theory.

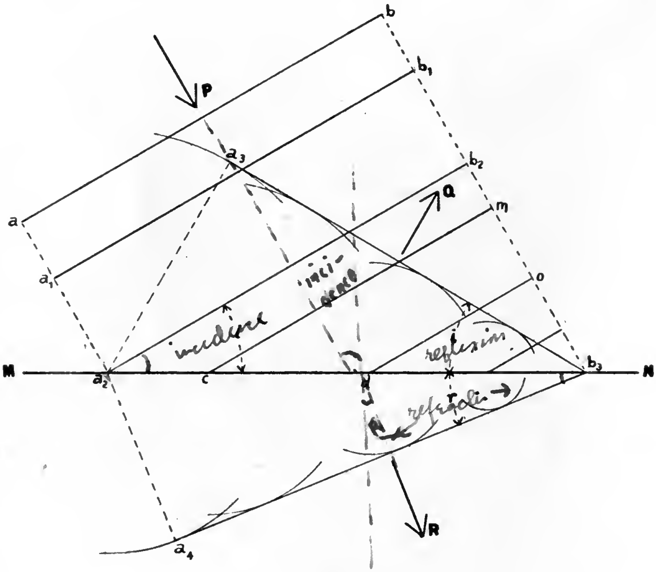

Figure 23

Let MN, in figure 23, represent the surface of a sheet of water, or the plane and polished surface of glass or some other reflecting and refracting medium. We shall suppose the water, glass, or other material to fill the space below MN, the medium above being the free ether. Plane waves are advancing through the ether, in the direction indicated by the arrow P, toward the surface. The lines ab, a_1b_1, etc., represent successive positions of an advancing wavefront, as it approaches MN. Our problem is to determine the position of the reflected wave and the refracted wave to which this incident wave gives rise.

In order to do this we shall make use of a principle enunciated by the Dutch physicist Huyghens, which applies to all types of waves. It may be stated as follows: A wave-front propagates itself by virtue of the fact that each point in the medium, as the wave-front reaches it, becomes itself a center of disturbance from which a spherical wave is sent out; and the further-advanced position of the original wavefront is nothing more nor less than the envelope of all the secondary wavelets sent out from the totality of points taken as centers.

When the wavefront ab reaches the position a_2b_2, the point a_2 therefore becomes the center of such a spherical wavelet, not only in the medium above MN, but also in that below. But the wavelets in the two media will not advance equally fast, because the velocity of light is less in the lower medium than in the free ether. Let us suppose the velocity in the lower medium to be 1/n of that in the upper. Then, while the incident wavefront is travelling onward from the position a_2b_2 till it reaches the reflecting surface at b_3, the secondary wavelet from a_2 will have acquired a radius equal to b_2b_3 in the upper medium, but a radius of only 1/n of b_2b_3 in the lower. Therefore an arc is drawn in the upper medium with radius b_2b_3, and one in the lower with radius b_2b_3/n, both having a_2 as center. What has occurred at a_2 will occur at every point on MN as the advancing incident wavefront reaches it, except that, if we want to construct the reflected and refracted wavefronts for the time when the incident wave is at b_3, we must take the radii of the secondary wavelets shorter and shorter for centers nearer and nearer to b_3. Thus, for the point c as center, we take the radius equal to mb_3 in the upper medium, mb_3/n in the lower. For, when the incident wavefront has reached c it has also reached m, and still has the distance mb_3 to travel. For d as center, the proper radii are ob_3 and ob_3/n, and so on. There should be an infinite number of such secondary wavelets, of which only a few are drawn in the figure. A plane passing through b_3 and tangent to all the secondary wavelets in the upper medium gives the wavefront of the reflected light, and another through b_3 tangent to all those in the lower medium gives that of the refracted light. Each of these advances perpendicular to its own plane, as shown by the arrows Q and R respectively. The student may ask what becomes of those parts of the secondary wavelets which do not lie on the common tangent plane. It can be shown that they mutually annul one another by interference, if we take account not only of the crests of waves but also of the troughs.

The angle $b_2a_2b_3$, which is a dihedral angle between the reflecting surface and the plane of the incident wavefront, is called the *angle of incidence.* The dihedral angle between the the reflecting surface and the reflected wavefront, $a_2b_3a_3$, is the *angle of reflection,* and that between the reflecting surface and the refracted wavefront, $a_2b_3a_4$, is the *angle of refraction.* a_2a_3 is drawn perpendicular to the reflected wavefront, a_2a_4 perpendicular to the refracted wavefront. Since a_2a_3 is equal to b_2b_3, the triangles $a_2a_3b_3$ and $b_2b_3a_2$ are equal, and the angles of incidence and reflection are equal. The triangles $b_2b_3a_2$ and $a_2a_4b_3$ are not equal, but are both right-angled triangles, the angles at b_2 and a_4 being each equal to 90°. Therefore, representing the angle of incidence by i and the angle of refraction by r, we have

$$\sin.\ i = \frac{b_2b_3}{a_2b_3}$$

$$\sin.\ r = \frac{a_2a_4}{a_2b_3}$$

Therefore,

$$\frac{\sin.\ i}{\sin.\ r} = \frac{b_2b_3}{a_2a_4}$$

By the method of constructing the figure, a_2a_4, bears the same relation to b_2b_3 that the velocity of light in the lower medium bears to that in the upper.

That is, $a_2a_4 = b_2b_3/n$, or $b_2b_3/a_2a_4 = n$. Therefore,

$$\frac{\sin.\ i}{\sin.\ r} = n$$

Since the velocity of light through a non-crystalline material such as glass or water is the same no matter what the direction of the rays may be, and depends only upon the nature of the material and the wavelength of the light, the above equation indicates that the ratio of the sines of the angles of incidence and refraction is *constant*, that is, it has the same value for all different angles of incidence. This statement, known as Snell's law, was first proved by direct measurements of different sets of angles of incidence and refraction. It holds good for all isotropic (non-crystalline) materials, but not—as we shall see later—for all crystals. The quantity n, which was originally defined simply as the ratio of the sines of i and r, is called the *index of refraction* of the lower medium in the figure, that is of the medium into which the light is refracted.

The index differs slightly for different colors or wavelengths. It is for this reason that a prism not only bends or refracts a beam of light, but also separates it into a spectrum. Since violet is bent more than red, evidently the index is greater for the shorter waves than for the longer, at least in ordinary media such as glass.

We have assumed, in discussing figure 23, that the upper medium is the free ether, but the conclusion would be exactly the same if it were any other isotropic medium, except that the corresponding index of refraction would have a different value. If the first medium were air, the change in the index would be extremely slight, since light travels almost as fast through air as through a vacuum. But if it were water or some other transparent solid or liquid the change would be great. In such a case, we say that the ratio of the sines—or, what comes to the same thing, the ratio of the light velocities—is the index of the second medium with respect to the first.

Let n_1 be the index of the first medium (with respect to the ether), n_2 that of the second, v_1 the velocity of light in the first medium, v_2 that in the second, and v_0 the velocity in the ether. Then

$$\frac{\sin. i}{\sin. r} = \frac{v_1}{v_2} = \frac{v_1 v_0}{v_2 v_0} = \frac{v_0}{v_2} \times \frac{v_1}{v_0} = \frac{v_0}{v_2} \div \frac{v_0}{v_1} = \frac{n_2}{n_1}$$

therefore, if n_{12} be used to indicate the index of refraction of the second medium with respect to the first, (light passing from the first to the second),

$$n_{12} = n_2/n_1$$

Tabulated values of the refractive indices of various solids and liquids are to be found in such collections of physical data as the Smithsonian Physical Tables, Recueil de Constantes Physique of the French Physical Society, the Physikalische-Chemische Tabellen of Landolt and Börnstein, and Kaye and Laby's Physical Tables. So far as glass is concerned, there is an almost endless variety of different glasses, having all sorts of variations of refractive index, dispersion, and absorption.

For rough calculations, we may take 1.53 as the approximate index of refraction of crown glass for yellow light, and 1.63 as that of flint glass. The index of water for yellow light is very nearly 1.33, that of diamond 2.42.

It is customary to speak of substances having very high refractive indices as being "optically dense." This is only a technical expression, for refractive index has nothing to do with real density, or specific gravity, beyond the fact that in general the heavier kinds of glass have the higher indices. Thus, we say that diamond is an extremely dense medium, and that flint glass is denser than crown.

25. Total reflection. Critical angle.—So far, we have taken up principally cases where the light passes from the less dense to the more dense medium, as from air to water, but obviously cases of the reverse type are almost as common. Whenever we see anything that lies below the surface of water, for instance, the light must pass out of water into air in order to reach our eyes. In such a case we may still write

$$\frac{\text{sin. i}}{\text{sin. r}} = \text{n}$$

where i means the true angle of incidence (on the water side of the boundary), r the true angle of refraction (on the air side), and n is the index of *air* with respect to *water*, just the reciprocal of the index of water as found in tables, that is, .75 instead of 1.33. Of course, beside the light that is refracted out into the air, there is always in addition a reflected beam going back into the water, for which the angle of reflection is equal to the angle of incidence, exactly as it would be if the light had been incident on the air side. Sometimes with light incident on the denser side of a dividing surface, it will happen when the angle of incidence is large enough, that there is no refracted light at all, *all* instead of *part* of the incident beam being reflected back into the first medium. A consideration of the formula of refraction shows that this must be so. Solving for sin. r, we get

$$\sin. r = \frac{\sin. i}{n}$$

Since, for such cases as we are now considering, n is less than unity, this equation shows that sin. r is greater than sin. i, and therefore r is greater than i, both being acute angles. It is possible, then for r to be equal to 90° and sin. r equal to 1, while i is still considerably less than 90°. If i becomes any greater, sin. r as calculated from the above equation becomes greater than 1; and this means, since the sine of a real angle cannot exceed 1, that there is no refracted wavefront. Obviously, the largest value that i can have for refraction still to occur, is that value which makes sin. r = 1. Such a value for the angle of incidence is called the *critical angle*. If we let γ represent the critical angle, we can find its value from the equation for refraction, by substituting γ for i, and 1 for sin. r. This gives

$$\sin. \gamma = n$$

As an example, let us calculate the critical angle for crown glass, in contact with air. We have taken 1.53 as the index from air to the glass, which gives 1/1.53, or .654 as the index from the glass to air. Therefore

$$\sin. \gamma = .654$$

$$\gamma = 40°49'$$

So much for the mathematical side of the question. The physical interpretation can be gotten by considering figure 23 again, with the modification that now the velocity of light in the lower medium is greater than that in the upper. Suppose, for instance, that the velocities in the two media, and the angle of incidence, have such values that while light travels the distance b_2b_3 in the upper medium it will travel a distance in the lower medium greater than a_2b_3. Under these circumstances the radius a_2a_4 of the secondary wavelet from a_2 will be so great that the point b_3 will lie *within* the sphere of the wavelet and it will be impossible to draw a plane through b_3 tangent to this sphere. When the angle of incidence has exactly the critical value, a_2a_4 the radius of the secondary wavelet from a_2, is just equal to a_2b_3.

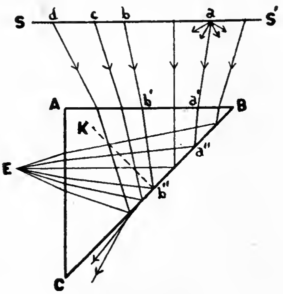

Figure 24

One of the most effective ways of showing total reflection is that illustrated in figure 24. ABC represents a right-angled prism of crown glass. The eye is held at some such point as E, close to one of the shorter prism-faces. SS′ is any rather brightly illuminated surface, such as the sky, the whitewashed wall of a room, or a sheet of paper. The eye sees, reflected in the hypothenuse BC, an image of the bright area SS′, the upper part of which is almost as bright as SS′ itself, the lower part much fainter. There is a fairly sharp boundary between the bright upper part, seen by *total reflection*, and the fainter lower part, seen by ordinary partial reflection. Following is the explanation.

Any point of SS′, such as a, sends out rays of light in all directions, but only a small bundle of these, comprising a cone, will reach such a position, after reflection at the face BC and refraction at the faces AB and AC, that they can enter the pupil of the eye and contribute to vision. For simplicity's sake, a single ray aa′a″E is drawn to represent this slender cone. Similar rays are drawn from a few other points on SS′. In every case, some light is lost by reflection at the two surfaces AB and AC, but this is not indicated on the drawing in order that the diagram may not become too intricate.

There will be some point on SS′, such as b, which is so situated that the cone of light from it which enters the eye will strike the hypothenuse with an angle of incidence, b′b″K, which is exactly equal to the critical angle. The light from any point to the left of b, such as c or d, will strike BC at an angle less than the critical angle, so that the greater part of the light in the small cone will be refracted through BC into the air, below and to the right of the prism, leaving only a small fraction to be reflected into the eye. Therefore such points as c and d will appear faint in the reflected image. On the other hand, the light from a, or any other point to the right of b, will be reflected from BC at an angle greater than the critical angle, none will be refracted through the hypothenuse, and all the light in the cone, except for the small amount reflected by the other two faces of the prism, will enter the eye. Therefore, points to the right of b will appear very bright in the reflected image of SS′, as bright as if they had been reflected by a silvered mirror, or even brighter, since a silver mirror does not reflect by any means all the light that falls upon it.

Naturally, the critical angle for any material depends not only upon the nature of the material itself, but also upon that of the material in contact with it. For example, if the prism were submerged in water instead of being in air, many of the rays which are totally reflected against air would be refracted through into the water.

The principle of the totally reflecting prism is utilized in a number of optical instruments, a few of which we shall consider later. But one of the most interesting cases of total reflection is seen in the case of a cut diamond. The index of diamond with respect to air being so large, its critical angle is correspondingly small, and a diamond owes its brilliance partly to this fact and partly to the additional fact that its index for different colors differs largely, so that for certain angles of incidence the shorter waves are totally reflected while the longer ones are not. Let figure 25 represent a cross-section of a cut diamond. Not only the ray aaa, but the very oblique one bbbb may undergo total reflection at the two surfaces XZ and YZ. Moreover, if the b ray happens to strike one of these surfaces at the critical angle for green light, for instance, the waves of shorter length will strike it at an angle greater than the critical, the waves of greater length at an angle less. For, since the index is greater for short than for long waves, the critical angle is greater for long than for short. Consequently, red rays, will escape total reflection and be par-

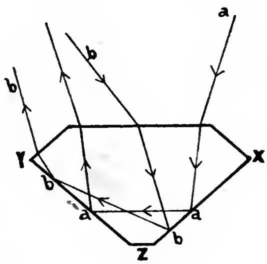

Figure 25

tially refracted out of the diamond for angles which make refraction impossible for violet rays.

Industrial applications of the total reflection principle have been made in the manufacture of so-called "shades" (light-distributors) for incandescent electric lighting, sidewalk-lights for illuminating basements, etc.

The appearance, to a fish, of things outside the water, is largely affected by refraction, and also illustrates total reflection. Figure 26 represents a pond, the eye of the fish being at E. XEY is the section of a cone, whose axis is vertical and whose half-angle K is equal to the critical angle of water, 48°45′. Any object outside the water, as at a, would be seen within the cone, for obviously light passing from air to water at an angle of incidence less than 90°, as it must be, will be refracted with an angle of refraction which is less than what would be the critical angle for light incident on the under side of the surface. The object a would appear in some such position as a′. It is clear then that the whole array of objects outside the water would appear to the fish very much distorted from their true relative positions, being crowded within a relatively small cone of view. Professor R. W. Wood, of Johns Hopkins University, has taken some curious photographs which illustrate what he calls "fish-eye views" reproductions of which can be found in his book "Physical Optics."

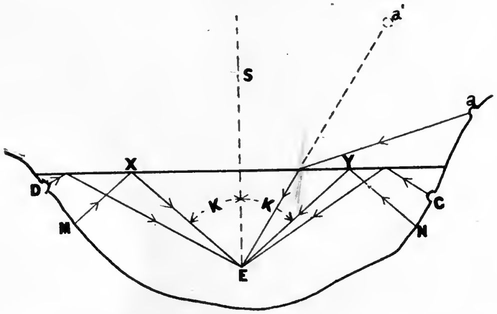

Figure 26

Whatever the fish sees by looking at the surface outside the cone XEY would be totally reflected images of objects in the water. For instance, it would see the objects C and D, not only directly, but also by reflection in the surface. In other words, the whole top surface, outside the circle whose diameter is XY, would appear as a perfect mirror. Inside this circle, the fish would see, not only objects that are outside the water, as already stated, but also faint reflections of objects within the water between M and N. It need hardly be pointed out that as the fish swims about the cone XEY moves with it, the diameter XY becoming smaller as the fish approaches the surface, larger as it sinks toward the bottom.

A man with his head under water would see things the same way as a fish but for one thing. Our eyes are adapted for seeing in the air, and the index of refraction of the cornea (the forward portion of the eye) with respect to air is such that the lens within the eye can bring to focus upon the retina objects anywhere from about eight inches to an infinite distance away. The substitution of water for air, as the medium in contact with the cornea, alters the refraction so that it is impossible, at least without extreme eye-strain, to focus upon the retina. Therefore human vision with the eyes in contact with water is very much blurred and indistinct.

26. Deviation through a prism.—We shall now consider the deviation of light by a prism (figure 27); and since, in the actual use of prisms for the production of spectra, the light waves are practically always first made plane by the use of a lens, we shall take only the case of plane waves. At each surface of the prism, not only refraction occurs, but also reflection; but in this discussion we shall ignore the reflected light. For the sake of symmetry, we call i_1 the angle of incidence at the first surface, r_1 the corresponding angle of refraction, r_2 the angle of incidence at the second surface and i_2 the corresponding angle of refraction; so that i_1 and i_2 are angles in air, r_1 and r_2 angles in glass. Then

$$n = \frac{\sin.\ i_1}{\sin.\ r_1} = \frac{\sin.\ i_2}{\sin.\ r_2}$$

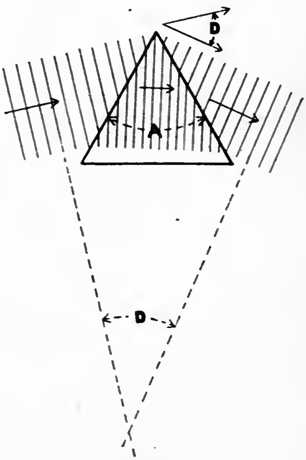

Figure 27

where n is the index of glass with respect to air. The drawing shows a series of wavefronts supposed to be just one wavelength apart (for instance the lines of the crests) both in the air and in the glass, though of course the actual length of the light waves, as compared to the dimensions of a practicable prism, is enormously exaggerated in the figure. It will be noticed that the wavelength is shorter in glass than in air. Indeed this must be so, for the following reasons: The period, or time of one vibration, must be the same in glass as in air, for only so many waves can in a given time leave the surface toward the glass side as come up to it on the air side. Also, since a train of waves advances the distance of one wavelength during the time of one vibration,

$$\text{velocity} = \frac{\text{wavelength}}{\text{period}}$$

Therefore, since the period is the same in the two media,

$$\frac{\text{wavelength in air}}{\text{wavelength in glass}} = \frac{\text{velocity in air}}{\text{velocity in glass}} = n$$

Consequently, whenever light of any wavelength λ passes from air into another medium whose index of refraction with respect to air is n, the wavelength within this medium is reduced to the value λ/n.

The angle D, between the wavefronts of the light before entering the prism and those after leaving it, or—what comes to the same thing—the angle between the *rays* before and after passage through the prism, is called the *angle of deviation*. The refracting angle of the prism itself we call A. It can be easily proved by simple geometry from the figure that

$$A = r_1 + r_2$$

$$D = i_1 + i_2 - A$$

From these two equations, together with the two gotten by applying the law of refraction to each surface of the prism, we can, if A, i_1, and n are given, solve for r_1, i_2, r_2, and D. A and n are necessarily constant for a given prism and a given wavelength of light, but by turning the prism about an axis perpendicular to the plane of the figure i_1 can be made to take any value from 0 to 90°. Changing the value of i_1 in such a manner will naturally cause changes in the value of D. If we plot the values of i_1 as abscissæ, and the corresponding values of D as ordinates, the curve will be found to have the form shown in figure 28, which shows that for a certain value of i_1 the value of D is less than for any other value of i_1. Both experiment and theory—by the application of the differential calculus—show that this minimum value of D occurs when $i_1 = i_2$ and $r_1 = r_2$, that is when the light passes through the prism symmetrically. In such a case,

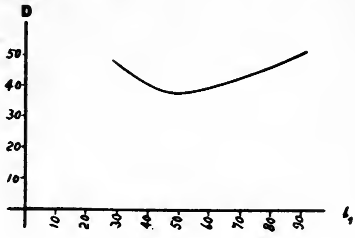

Figure 28

$$A = 2r_1$$

$$D = 2i_1 - A$$

$$r_1 = \frac{A}{2}$$

$$i_1 = \frac{D + A}{2}$$

$$n = \frac{\sin.\ i_1}{\sin.\ r_1} = \frac{\sin.\ \frac{D + A}{2}}{\sin.\ \frac{A}{2}}$$

This equation has a great deal of importance in practical optical work. For, by the use of a spectrometer, both the refracting angle of a prism and the angle of minimum deviation, D, can be measured with great accuracy. Therefore, by applying this equation we can get very accurate determinations of the index of refraction for any piece of glass that can be obtained in the form of a prism. It is the most convenient method for finding not merely the average index for white light, but the separate indices for different wavelengths.

Problems.

1. Calculate the angle of refraction when light strikes crown glass with an angle of incidence of 60°.

2. Find the index of refraction of crown glass with respect to water, for yellow light.

3. Light within a piece of crown glass strikes the surface at an angle of incidence of 40°. At what angle does it emerge?

4. Calculate the critical angle for diamond, yellow light.

5. Show that if light strikes a pile of parallel-sided plates with different refractive indices, all in contact with one another, it enters each plate with the same angle of refraction as if the others were absent, and finally emerges parallel to its original direction.

6. A star seems displaced from its proper position owing to refraction in the earth's atmosphere. Show that, despite the changes in the density of the air at different levels, we can calculate the refraction by considering that all the air has the same index as that at the earth's surface. (See problem 5).

7. Find the approximate length of wave, in water, of the extreme red and the extreme violet light, assuming the index for both to be 1.33.

8. Calculate the angle of minimum deviation for a 60° prism, the light having index 1.68.

9. What must be the properties of a body which, in air, would be invisible under any illumination? Would such a body be visible if immersed in water?

10. Certain aquatic bodies are nearly invisible in water. What are their properties?

11. Show that any colorless and transparent object would be invisible if surrounded completely by uniformly illuminated walls.

12. Prove that, in figure 27, $A = r_1 + r_2$ and $D = i_1 + i_2 - A$.

13. A real diamond will continue to glitter when immersed in water, while an imitation will not. Explain this.

14. Plot four points on a curve like figure 28, for a prism of 60°, having an index 1.54.

CHAPTER V.

27. Reflection and refraction of spherical waves at a plane surface. —28. Judgment of the distance of an image.—29. Image of an extended object.—30. Reflection and refraction at spherical surfaces.—31. Lenses. —32. Two lenses in contact.—33. Chromatic aberration.—34. Achromatic lenses.—35. Image of an extended object. Undeviated ray.—36. Magnification.—37. Micrometer.—38. Imperfections in mirrors and lenses.—39. Spherical aberration.—40. Curvature of field.—41. Astigmatism.—42. Lenses for special purposes.

27. Reflection and refraction of spherical waves at a plane surface.—It is shown in the preceding chapter that when plane waves strike a plane surface, both the refracted and the reflected wavefronts are plane. In this chapter we shall show that when the incident wavefronts are spherical (diverging from a point) and the reflecting surface plane, the reflected wavefronts are also spherical but the refracted wavefronts are not, except as an approximation to the truth. We shall also show that when both the incident wavefronts and the reflecting surface are spherical, neither the reflected nor the refracted wavefronts are truly spherical, except in the one special case that the center of the incident waves coincides with that of the reflecting surface.

In figure 29, let CBD represent the plane boundary between two media, the index of refraction of the lower with respect to the upper being n. A is the center of a system of spherical wavefronts, advancing from A toward the surface CBD. B is the foot of the perpendicular from A upon this surface, that is, the first point on CBD reached by each wavefront. If the reflecting surface had not been in its place, a wavefront, after reaching B would continue to travel with its original velocity and in a short time would reach some such

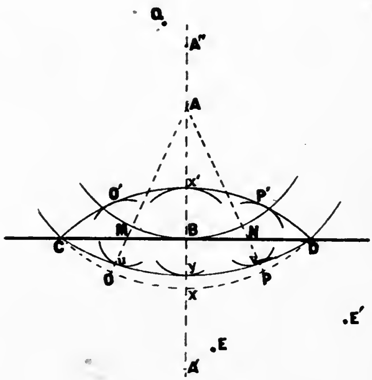

Figure 29

position as CxD. With the surface in its place, B becomes the center for a secondary wavelet in the upper medium whose radius, at the instant when the incident wavefront reaches C and D, will have acquired the length Bx′ = Bx. In the meantime, other points such as M and N will also have been reached by the incident wave and become centers of secondary wavelets, whose radii in the upper medium will be respectively MO′ = MO and NP′ = NP. The reflected wavefront will therefore be CO′x′P′D, the envelope of all such secondary wavelets. From the manner of its formation, it is obviously exactly symmetrical with the hypothetical incident wavefront COxPD, and therefore is truly spherical, with center at A′. A and A′ are equally distant from the reflecting surface, and the line AA′ is perpendicular to the latter. An eye placed anywhere in the upper medium would receive reflected light which would appear to come from A′, though really from A, and we therefore say that A′ is the *reflected image*, or *image by reflection*, of A.

The refracted wavefront is formed in a similar manner, except that in the lower medium the radii of the secondary wavelets are shorter than in the upper if n is greater than unity, longer if n is less than unity. The radius By of the secondary wavelet from B as center is not equal to Bx, but to Bx/n. The radius of the secondary wavelet from M is Mu = MO/n, that from N is Nv = NP/n, etc. The envelope of all these secondary wavelets in the lower medium, CuyvD, turns out to be, not a sphere, but a surface of higher order. Therefore, we cannot say that there is a *refracted* image in the same strict sense in which we speak of a reflected image. It is true that an eye placed in the lower medium would receive light that appeared to come from some point in the upper medium other than the true source A, but this apparent "image" would have a different position for every change in the location of the eye.

However, it is always possible to describe a sphere which approximates more or less closely to the refracted wavefront. Suppose, for example, that a circular arc be drawn through the three points C, y, and D, and a spherical segment be formed by rotating this arc about the axis AA′. This surface would coincide exactly with the refracted wavefront at the three given points, and would also pass very close to other points such as u and v. The nearer the three points C, y, and D are together, the closer would sphere and wavefront coincide for all the region between C and D, and if the three points are distant from one another by only an infinitesimal amount we may say that *in the immediate neighborhood* of these points the coincidence is exact. If A″ represents the center of this sphere, then to an eye located in the lower medium anywhere along the line AA′ or AA′ produced, or in the close neighborhood of this line, as at the point E, the light would appear to come from A″ instead of from A, and we therefore define A″ as the *image by refraction* of A. But, if the eye be located at some distance from the line AA′, as at E′, the light appears to come from a different point, such as Q.

In order to find the position of the image A″, it is best to consider the refracted wavefront just as it breaks through the surface CBD, that is, we imagine C and D to be very close together and By and Bx to be infinitesimally small. Then the radius of the sphere through C, y, and D will be the radius of the refracted wavefront just as it breaks through the surface, and equal to the distance of the image A″ from the surface.

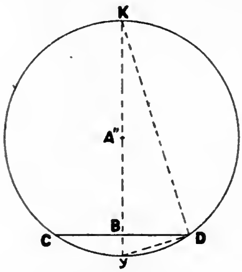

Figure 30

To find A″ is merely a matter of plane geometry. CD is a chord common to the two arcs CxD (hypothetical incident wavefront) and CyD (refracted wavefront). The line drawn from the middle point of a chord, perpendicular to the latter, till it meets the arc, is called the *sagitta* of the arc. Thus, Bx is the sagitta of the arc of the incident wavefront, By that of the refracted wavefront. We must first find what relation the sagitta bears to the radius. This can be best done by a consideration of figure 30, where the complete circle is shown. The triangles KBD and DBy are similar, therefore

$$\frac{KB}{BD} = \frac{BD}{BY}$$

If we let s represent the sagitta, R the radius of the circle, and

a the half-chord, KB $= 2R - s$, BD $= a$, and By $= s$. Therefore

$$\frac{2R - s}{a} = \frac{a}{s}$$

$$s(2R - s) = a^2$$

$$\frac{s}{a^2} = \frac{1}{2R - s}$$

We are supposing that C, y and D are very close together, so that both s and a are quantities of infinitesimal magnitude, but R remains a quantity of finite size, therefore in the limit $2R - s$ becomes equal to 2R, and

$$\frac{s}{a^2} = \frac{1}{2R}$$

or,

$$R = \frac{a^2}{2s} \qquad (1)$$

Equation (1) shows that although a and s both become infinitesimally small, the ratio a^2/s remains a finite quantity, equal to the diameter of the circle.

Now let us apply equation (1) to both the hypothetical incident wavefront and the refracted wavefront. For the former, s=Bx, R=BA. For the latter, s=By, R=BA″, the required distance. Therefore, since a is the same for both, viz., BD,

$$BA = \frac{a^2}{2Bx}$$

$$BA'' = \frac{a^2}{2By}$$

If we divide the last equation by the one above it, we get

$$\frac{BA''}{BA} = \frac{Bx}{By}$$

But, by the method of constructing the refracted wavefront, $Bx/By = n$. Therefore,

$$\frac{BA''}{BA} = n$$

or in general, if d_1 represent the distance of any point source of light from a plane refracting surface, d_2 the distance of the corresponding refracted image, and n the index of refraction of the second medium with respect to the first, then.

$$\frac{d_2}{d_1}=n \qquad (2)$$

As an application of formula (2) suppose a certain object to be 2 feet above the surface of a pond. A person above the surface would of course see a reflected image of it, apparently 2 feet beneath the surface, but a fish in the water would see a refracted image. If the fish is directly beneath the object, we can apply equation 2, putting $d = 2$, $n = 4/3$, or 1.33. This gives

$$d_2 = 2 \times \frac{4}{3} = 2.67 \text{ feet}$$

That is, the object would appear to the fish to be 2 ft. 8 in. above the surface.

If the source of light lies within the denser medium, the refracted light travels faster than the incident, the refracted wavefront of figure 29 will be bulged out more than the incident instead of being flattened, and n has a value less than unity. Thus, suppose we look straight down to the bottom of a pool 2 feet deep. Then, in equation (2), $d_1 = 2$, $n = 3/4$, and

$$d_2 = 2 \times \frac{3}{4} = 1.5 \text{ feet}$$

The pool appears to be only ¾ its actual depth. The fact that the wavefronts are not spherical is shown clearly by the observation that, if we look very obliquely to the bottom, the depth appears to have much less than ¾ its actual value. This explains a curious phenomenon which anyone standing in a pool a few feet deep with perfectly level bottom can hardly fail to notice. The bottom always appears to be bowl-shaped, with the greatest depth just underfoot, and that depth of course just about ¾ the true depth of the whole pool.

28. Judgment of the distance of an image.—The student may wonder why the curvature of the wavefront has anything to do with our judgment of the distance of an object perceived, for one is apt to think that only a single ray ever enters the eye. This is incorrect, for the pupil of the eye has a finite size and therefore always receives a finite area of the wavefront. According as the curvature of this section of wavefront is greater or less, we must exert more or less muscular strain upon the lens in the eye in order to focus the light upon the retina, and the degree of this strain enables us to judge distance to some extent. Still more important is the fact that we ordinarily see with both eyes at once, thus taking in at the same time two separate sections of wavefront. A person with only one eye is far less accurate in estimating distance than a normal person. For instance, a one-eyed man usually has much greater difficulty in hitting a nail with a hammer for this reason.

Naturally enough, estimation of distance becomes much more difficult, even for normal two-eyed vision, when the distance becomes great. It is comparatively easy to tell whether an object is five feet away or ten feet, for the difference in curvature of a five foot and a ten foot sphere is comparatively great. But it is not easy to tell whether a distant object is nearer to a mile or to two miles from us. For a small section of a sphere of either one or two miles radius is nearly flat, and there is no perceptible difference in the focussing and alignment of the eyes for such great distances. Our estimation of great distances is a result of subconscious consideration of such details as size, speed of motion (if the object seen happens to be in motion), distinctness of vision, etc., and at best it is very uncertain and subject to queer illusions. Soldiers are given a long and systematic training in judging distance.

If our eyes were set three feet apart instead of a few inches, judgment of distance would become much easier. No doubt small animals with their eyes close together are less adept in this respect than human beings. Indeed there is some reason for thinking that many animals and birds are much less keen than men in observing stationary objects, although a moving objects almost instantly arrests their attention. Most birds have the eyes set in the sides of the head, and therefore see an object with only one eye at a time. This fact must seriously hinder them in the estimation of distance; and it is no doubt to counteract this deficiency that birds have the habit, particularly when alarmed, of darting the head rapidly forward and backward, so as to get two or more points of view of any object that excites their suspicion. The *parallax* of the observed object gives some ground for a judgment of its remoteness. Thus the mechanism by which a bird estimates distance is perhaps an automatic and unconscious application of the same principle used by a surveyor in finding the width of a river, or by an astronomer in finding the distance of the nearer fixed stars.

29. Image of an extended object.—We have so far considered only cases where a single point acted as a source of light, and have found the positions of the reflected image and the refracted image of this point source. Actually, we always have to deal with objects more or less extended; but in order to find the images of such an object we have only to apply the principles already learned to each point of it. In figure 31, AB represents any object. The point A has an image by reflection, A′, and an image by refraction, A″, found by these principles, and similarly B has an image by reflection, B′, and an image by refraction B″, and so on. So far as the image by reflection goes, it is exactly symmetrical to the object, with respect to the reflecting plane.

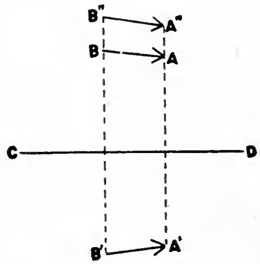

Figure 31

30. Reflection and refraction at spherical surfaces.—We shall now consider reflection and refraction at spherical surfaces, a subject which is of great importance because of the use of curved mirrors and lenses in optical instruments.

Figures 32 to 35 show four different cases of the reflection and refraction of spherical wavefronts at a spherical surface, the source of the incident waves being in each case on the concave side. All these figures are drawn on the supposition that the medium on the convex side (second medium), is denser than that on the concave side (first medium), and both media are supposed to extend indefinitely. The reflecting and refracting surface is indicated by a heavy line, the incident wavefronts by normal unbroken lines, and the reflected and refracted wavefronts by dotted lines. The two last are not truly spherical, but are near enough to be considered so as long as the incidence is nowhere very oblique. Therefore, the center of the incident wavefronts is represented by O, the center of the reflected wavefronts (except in figure 35) by I, the center of the refracted wavefronts by I′, and the center of the surface itself by C. Arrows show the direction of advance of the waves.

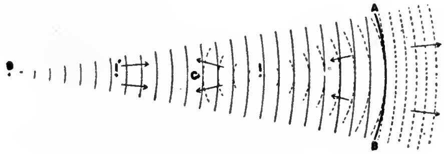

Figure 32

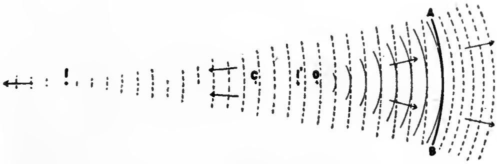

Figure 33

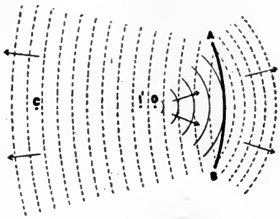

Figure 34

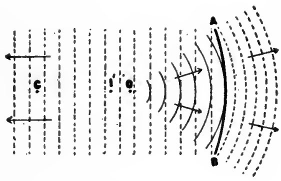

Figure 35

In figure 33, the source is farther from the mirror than is the center C. The reflected waves converge to the point I, between C and the mirror, but nearer to the former, and diverge again after passing through I. The refracted waves diverge more than the incident, appearing therefore to come from a point I′, between C and O.

In figure 32 the source O is between C and the mirror, nearer to the former. The point I, to which the reflected waves converge, and from which they later diverge, lies outside of C. The refracted waves diverge less than the incident, appearing to come from I′ between O and C.

In figure 34 O is nearer to the surface than to C. Here the reflected waves do not converge at all, but diverge at once, seeming to come from the point I within the second medium.

The center of the refracted waves is again between O and C.

In figure 35 O is just halfway between C and the surface. The reflected waves are plane, and may be said either to *converge* to an infinitely distant point on the left, or to diverge from an infinitely distant point on the right. The latter statement is preferable, since an eye placed in the path of the reflected light would see an image of O infinitely distant on the right of the figure. In fact these waves would strike the eye exactly as waves coming from a very distant star.

The points I in figures 32 and 33 are said to be *real* images of the corresponding sources O, because the light is actually converged to these points; while in the other two figures I is only a *virtual* image, the reflected light not actually being brought to focus at these points, but simply diverging as if it had come from them. In each of the figures, the refracted image, I′ is virtual.

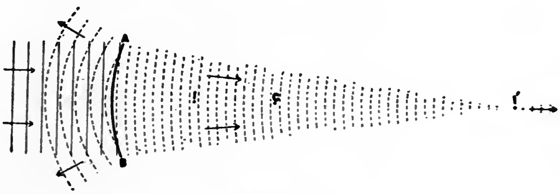

Figure 36

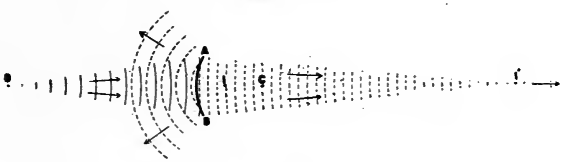

Figure 37

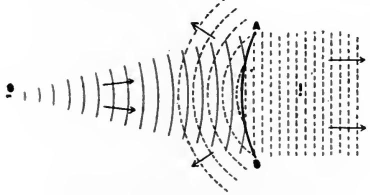

Figure 38

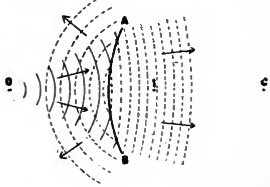

Figure 39

Figures 36 to 39 differ from the four preceding figures in that the reflecting and refracting surface is convex, instead of concave, to the incident light and the rarer medium. In each of these cases, the reflected rays diverge from the surface at once, and I is therefore a virtual image. As shown in figure 38, there is a certain position for the source O, depending upon the radius of the surface and the index of refraction, for which the refracted waves are plane. If O lies any nearer to the surface, as in figure 39, they diverge, and I′ is virtual. If O is farther from the surface, they converge, and I′ is real, as in figures 36 and 37. In figure 36 the source is infinitely distant, and the incident waves are plane.*

From what precedes, it is obvious that there is a great diversity of typical cases for reflection and refraction at spherical surfaces. The surface may be concave or convex to the incident light; the source may lie in the medium of less or of greater optical density, and may be at any distance from the surface. To develop and remember for each case a special formula giving the positions of I and I′ in terms of that of O, would be unduly laborious. Fortunately, we can derive a pair of very general formulæ, which are applicable to any case that may arise, provided we make consistent and rational conventions in regard to the algebraic sign of the distances involved.

We shall suppose all distances to be measured from the reflecting surface as a base, distances to one side being regarded as positive, those to the other side negative; and it will be most convenient to take the side from which the light comes as the positive side. We let r stand for the radius of the reflecting and refracting surface, u for that of the incident wavefronts, v for that of the reflected wavefronts, and v′ for that of the refracted wavefronts. In figures 32 and 33 all these quantities are positive. In figure 34 v has become negative, while it is $\pm\infty$ in figure 35. v′ is negative in figures 36 and 37, $\pm\infty$ in figure 38, and positive in all the other cases shown. r is always positive for a concave mirror, negative for a convex one. u is always and necessarily positive, unless the incident waves are rendered convergent before striking the surface, by means of another mirror or a lens.

*In figures 32 to 35 the index of refraction has been taken as 1.5, but in figures 36 to 39 it has been taken as 1.67 to avoid making some of these figures inconveniently long.

It is easily seen that v is positive when the reflected image is real, negative when it is virtual. On the other hand, v′ is negative when the refracted image is real, positive when it is virtual.

It is easiest to derive the two formulae by considering a case where all the quantities, r, u, v, and v′ are positive, as in figure 32. Figure 40 represents the case of figure 32 some-

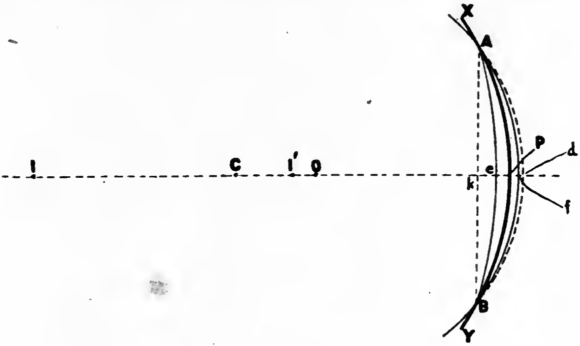

Figure 40

what exaggerated to make the diagram clearer. I, C, I′ and O have the same significance as in the preceding figures. XPY is the reflecting and refracting surface (center at C), AdB an incident wavefront as it would be if it advanced into the second medium without being retarded (center at O), AfB the actual refracted wavefront (center at I′), and AeB the reflected wavefront (center at I). In practice, mirrors are seldom used in which the diameter XY of the mirror-face is more than ⅙ of the radius of curvature CP = r. In this figure XY is made about equal to r in order that the different arcs having a common chord AB may be more clearly seen as separate.

The formula for the reflected wave is based upon the following physical fact: While the incident light would, but for retardation in the second medium, travel from P to d, the reflected light travels back from P to e. Therefore the distances Pd and Pe are equal. Putting this statement into the form of an equation,

$$Pe = Pd \qquad (3)$$

But we can write

$$Pe = PK - eK$$
$$Pd = dK - PK$$

Therefore

$$PK - eK = dK - PK$$
$$dK + eK = 2PK$$

But dK, eK, and PK are respectively the sagittas of the incident wavefront, the reflected wavefront, and the reflecting surface. We can therefore apply to each of them the general geometrical formula (1), getting

$$dK = a^2/2u$$

$$eK = a^2/2v$$

$$PK = a^2/2r$$

Making these substitutions, we get

$$\frac{a^2}{2u} + \frac{a^2}{2v} = \frac{2a^2}{2r}$$

or,

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r} = \frac{1}{f} \qquad (4)$$

This is the general formula for a mirror. The symbol f, equal to r/2, is called the *focal length* of the mirror. Its physical meaning can be shown by supposing that $u = \infty$, that is, that we are dealing with an object infinitely distant. Then $1/u = 0$, $1/v = 1/f$, and $v = f$. Then f is the distance from the mirror to that point (called the *principal focus*) where parallel rays, or plane waves are brought to a focus in the reflected light. Conversely, if $u = f$, $1/v = 0$, and $v = \infty$; that is, if the source of light is at the principal focus, the reflected waves are plane. If neither u nor v is infinite, the points O and I are called *conjugate foci.* If the source is at O, the reflected image is at I, and conversely if the source is at I the image will be at O, for equation (4) shows that the relation between these two points is reciprocal, u and v appearing in it in exactly the same way.

The formula for the refracted wave is found in a similar way. The physical fact upon which it is based is this: While the incident light would, but for the retardation in the second medium, travel from P to d, the actual refracted wave travels only from P to f, where Pd and Pf are to each other as the velocities of light in the two media. That is,

$$\frac{Pd}{Pf} = n$$

or

$$Pd = n \times Pf$$

But

$$Pd = dK - PK$$

$$Pf = fK - PK \qquad (5)$$

therefore

$$dK - PK = n(fK - PK)$$

$$n \times fK - dK = (n - 1)PK$$

But fK, dK, and PK are the sagittas respectively of the refracted wavefront, the incident wavefront, and the refracting surface. Applying to each of these the geometrical formula (1), we get

$$fK = a^2/2v'$$

$$dK = a^2/2u$$

$$PK = a^2/2r$$

With these substitutions,

$$\frac{n \times a^2}{2v'} - \frac{a^2}{2u} = \frac{(n-1)a^2}{2r}$$

or

$$\frac{n}{v'} - \frac{1}{u} = \frac{n-1}{r} \qquad (6)$$

This formula, for the refracted light, is necessarily more complicated than that for the reflected, because it involves the index of refraction, which does not affect reflection, and therefore does not appear in (4).

Formula (6) is of less common use than (4), but there are certain problems where it becomes necessary, for example the following.

A spherical globe, one meter in diameter, and made of thin glass, is filled with water. A small fish is located 40cm. from the glass wall at a certain side of the globe. (a) Where would be the image which the fish would see of himself in the farthest part of the surface? (b) Where would the fish appear, to a person outside the globe on the side farthest from the fish?

Question (a) is easily answered, for the fish would see his own reflected image, and we only need to apply equation (4) as if the glass wall were non-existent, since its thinness prevents it from affecting the problem to any extent. Then we must put $r = 50$, $u = 60$, giving $v = 42.9$cm. Therefore the fish would see, reflected from the farthest part of the boundary of the globe, an image of himself 42.9cm. from the boundary, 17.1 cm. from himself. (If we had tried to find the location of the image of the fish formed by reflection in the *nearest* part of the wall, it would have come out to be behind the fish and therefore not discernable by him as an image).

To answer question (b), we must apply equation (6), for we have to do with refracted light. Since the light passes from water to air, the appropriate index of refraction is not $4/3$, but the reciprocal of this, $3/4$. Therefore

$$\frac{3}{4v'} - \frac{1}{60} = -\frac{1}{4 \times 50}$$

Giving $v' = 64.3$. Therefore the person outside the globe would see the fish apparently 4.3 cm farther away than it really is. The fact that v' comes out positive shows that the refracted image is on the same side of the bounding surface as the fish itself. The reflected image seen by the fish is real, the refracted image seen by the observer outside is virtual.

Mirrors for experimental work in optics are usually either flat or concave, though convex mirrors are occasionally used. They are made by taking a disc of homogeneous and thoroughly annealed glass, and reducing one surface to the required radius of curvature by careful grinding and polishing. This surface is then covered with a deposit of silver by chemical deposition from a solution of silver nitrate, and the silver film is thoroughly dried and then lightly polished with chamois and rouge. The silver of course prevents any appreciable refracted light so that the major part of the incident light is turned into the reflected beam. Flat mirrors such as were used in the experiments of Fizeau and Foucault for finding the velocity of light, whose function is to reflect only part of the light and transmit the rest, must of course be ground and polished flat on both sides. One face is then covered with a very thin film of silver, the best result being obtained when the film reflects approximately half the incident light, transmitting the rest, except for some unavoidable absorption. Such a mirror is said to be *half-silvered.* When a silvered mirror becomes tarnished and dull, it is a comparatively easy matter to dissolve off the old silver with nitric acid and put a new silver coating upon it.

31. Lenses.—A lens is a disc of transparent refractive material, such as glass, bounded by two surfaces, one or both of which is curved, usually spherical. Each surface produces some reflection, which not only weakens the transmitted beam but also causes annoyance in other ways. The light reflected from the lens-surfaces is therefore a hindrance, but is absolutely unavoidable. In the following discussion of lenses we shall ignore the reflected light.

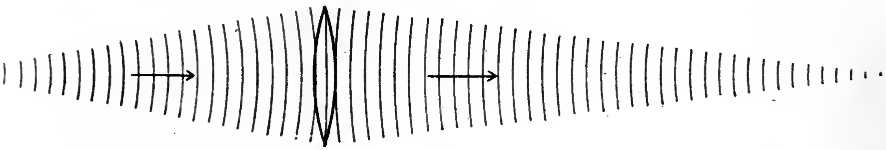

Figure 41

When spherical wavefronts, with center at O (figure 41) strike a lens, they are refracted at the first surface, and again at the second surface, finally emerging approximately spherical, so that they either converge to a point I on the side opposite to O, as in the figure, or diverge from a point on the same side as O. Our task is to derive a formula by means of which, knowing the distance of O from the lens, the radii of curvature of the two lens-surfaces, and the index of refraction, we can calculate the distance of I. This might of course be done by applying equation (6) once for each surface, taking due account of the fact that the appropriate index to be used at the second surface is the reciprocal of that for the first.

However, we shall develop our lens formula by a different method, chiefly because by so doing we can introduce a convention as to algebraic sign which will prove more convenient for our purpose than the one used in equations (4) and (6). We assume that the lens is so thin that its greatest thickness may be neglected in comparison with the distance from source to image. Call the distance from the lens to the center of the incident wavefronts u, that from the lens to the center of the refracted wavefronts v. u is considered positive when the center of the incident wavefronts lies on the side from which the light comes, that is, when the incident light is diverging, as is practically always the case. Otherwise, u is negative. On the other hand, we consider v as positive when the center of the refracted wavefronts is on the side opposite to that from which the light comes, that is, when the light leaves the lens in a converging beam. The radius of curvature of the first surface of the lens will be considered positive when that surface is convex to the incident light; that of the second surface is positive when it is concave to the incident light. By this convention, all four of these quantities will be positive in the most common case, viz., when a double-convex lens forms a real image. In figure 42 LAL′B is a somewhat exaggerated diagram of a lens. O is the center of the incident waves, or source, I is the center of the emergent waves, or image. Let r_1 be the radius of curvature of the first surface, LAL′. r_2 that of the second, LBL′.

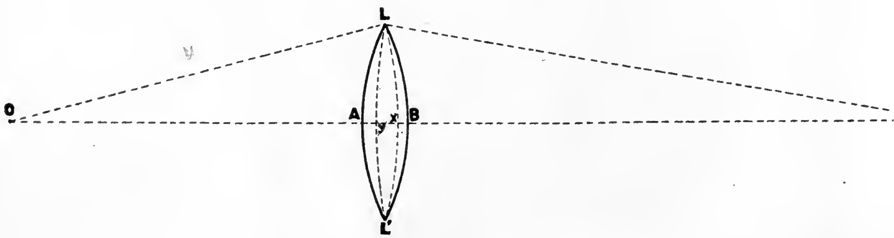

Figure 42

In developing the formula, we shall use the following principle: In order that I shall be the image of O, there must be the same number of wavelengths in every path from O to I. In particular, there are as many in the distance OL + LI as in the straight path OABI. Of course this can be true only because in part of the shorter path, viz., in the distance AB, the wavelength is shorter than in the air. We have already seen that, if λ is the wavelength in air, λ/n will be that in glass whose index of refraction is n. Therefore the total number of wavelengths in the straight path is

$$(OA + BI)/\lambda + n \times AB/\lambda = (OA + BI + n \times AB)/\lambda$$

The number in the path OL + LI is

$$(OL + LI)/\lambda$$

Equating these two expressions,

$$OL + LI = OA + BI + n \times AB$$

or

$$OL - OA + LI - BI = n \times AB$$

Now draw the arcs LxL′, from O as center, and LyL′, from I as center. Then OL = Ox, and LI = Iy. Therefore,

$$Ox - OA + yI - BI = n \times AB$$

$$Ax + By = n \times AB$$

But Ax is the sum of the sagittas of the incident wavefront and the first lens-surface; By is the sum of the sagittas of the emergent wavefront and the second lens-surface; and AB is the sum of the sagittas of the two lens-surfaces,—all with the same chord LL′. Therefore, we may substitute the appropriate values obtained from equation (1), and get

$$\frac{a^2}{2u} + \frac{a^2}{2r_1} + \frac{a^2}{2v} + \frac{a^2}{2r_2} = u\left(\frac{a^2}{2r_1} + \frac{a^2}{2r_2}\right)$$

or

$$\frac{1}{u} + \frac{1}{r_1} + \frac{1}{v} + \frac{1}{r_2} = n\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$$

or

$$\frac{1}{u} + \frac{1}{v} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \quad (7)$$

This is the approximate formula for a lens. It is admittedly not accurate, and indeed no perfectly accurate formula can be found. For the wavefront emerging from a lens is not accurately spherical. Consequently it has no true center and there is no perfect focus. It is possible, by giving the lens-surfaces a special non-spherical form, to make the emergent wavefronts really spherical; but this can be done only for a certain fixed distance of O from the lens, and the emerging waves are no longer accurately spherical if the source is moved closer to the lens or farther away. There is therefore no advantage to be gained by using the mathematically correct form, except in the case of telescopic and microscopic objectives, which are always used under the same conditions. With spherical lens-faces, formula (7) is accurate enough for all ordinary purposes, provided the lens is thin and its diameter is not more than 1/20 the distance u or v. For photographic lenses and microscopic objectives, which are thick and have relatively large diameters, it becomes very inaccurate.

Since the right-hand member of (7) contains only terms which are constant for a given lens (r_1, r_2, and n) it is convenient to replace it by a single symbol, 1/f, where f is known as the *focal length* of the lens. The formula then becomes

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \qquad (8)$$

which is identical with one of the forms of the equation (4). The meaning of the focal length f is also the same in the case of mirror and lens. That is, f is the distance from the lens to the *principal focus*, which is the point to which incident plane waves would be brought to focus by the lens, or the point such that if the center of the incident wavefronts were located there the emergent wavefronts would be plane. In fact, the only differences between (4) and (8) are—first the difference in convention as to sign, already explained,—second, the fact that the focal length of a mirror is simply half the radius, while that of a lens is a function of two radii and an index of refraction, having the value

$$f = \frac{r_1 r_2}{(n-1)(r_1 + r_2)} \qquad (9)$$

If we solve equation (8) for v, we get

$$v = \frac{uf}{u - f}$$

As already noted, u is practically always positive, so that if f is also positive, v will be + if $u>f$, infinite if $u=f$, and — if $u<f$. That is, the emergent wavefronts will be convergent if the source lies beyond the principal focus, plane if it is at the principal focus and divergent if it lies between the principal focus and the lens itself.

A negative value for f itself means that plane waves falling upon it from the left would not be converged to a point on the right, but diverged as if they came from a point on the same side as the incident light. Equation (9) shows that in order for f to be negative either r_1 and r_2 must both be negative, or the larger one must be positive and the smaller negative, since in all practical cases n is greater than 1. This is the same as saying that f is negative if the lens is thinner in the middle than at the edges, positive if thicker at the middle than at the edges. In the latter case we say that the lens is *converging* or *convex*, in the former case *diverging* or *concave*, In figure 43 are shown three different types of converging, and three of diverging lens. In order from left to right, they are named *planoconvex, double convex, concavoconvex* (or *meniscus*), *convexoconcave, double concave,* and *planoconcave.*

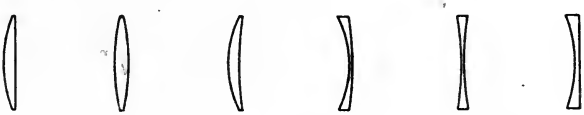

Figure 43

From the elementary theory of lenses that we have given here, it is immaterial which face is turned toward the incident light, for equation (9) shows that r_1 and r_2 can be interchanged without affecting the value of f, and such an interchange would be the only effect of turning the lens around. That is, for example, the meniscus type has the same focal length no matter whether the convex or the concave face be turned toward the incident light. But a more thorough study of lenses shows that there usually is a choice, depending upon the circumstances under which the lens is to be used. In some cases, it is best to use a meniscus or planoconvex lens, with the faces turned in a certain way, while in others a symmetrical double convex will function better, etc. The complete theory of lenses is a long and difficult study in itself, and cannot be taken up in this book.

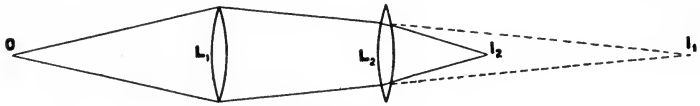

Figure 44

The quantity u can never be negative so long as the lens receives the light directly from the source. But figure 44 shows a case where, for the second lens, L_2, u is negative. (Here the wavefronts are not drawn, but the course of the light is sufficiently well indicated by the limiting rays of the beam.) O is a point-source, the light from which would be brought by lens L_1 to a focus at I_1. Lens L_2 therefore receives convergent light whose center is at I_1, and in order to find the position of the final image I_2 we must substitute for u in equation (8) the numerical value of the distance L_2I_1 with a negative sign in front of it, and then solve as usual for v.

32. Two lenses in contact.—We can now prove that when two thin lenses are placed very close together they act approximately as a single lens, the reciprocal of whose focal length is equal to the sum of the reciprocals of the focal lengths of the

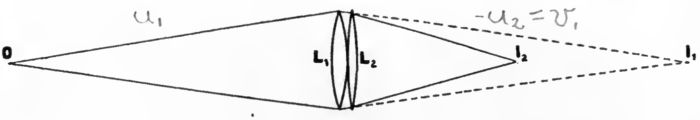

Figure 45

two given lenses. See figure 45. Let f_1 be the focal length of L_1, f_2 that of L_2. Aplying equation (8) to each lens, we get

$$\frac{1}{u_1} + \frac{1}{v_1} = \frac{1}{f_1}$$

$$\frac{1}{u_2} + \frac{1}{v_2} = \frac{1}{f_2}$$

Adding these two equations, we get

$$\frac{1}{u_1} + \frac{1}{v_1} + \frac{1}{u_2} + \frac{1}{v_2} = \frac{1}{f_1} + \frac{1}{f_2}$$

Since the center of the emergent wavefronts for the first lens, I_1, is also the center of the incident wavefronts for the second, and since, L_1 and L_2 being very close together, they are almost the same distance from I_1, v_1 and u_2 are numerically practically equal. but opposite in algebraic sign. Therefore $1/u_2$ and $1/v_1$ cancel one another, and we have left

$$\frac{1}{u_1}+\frac{1}{v_2}=\frac{1}{f_1}+\frac{1}{f_2}$$

u_1 for the first lens is simply u for the combination, and v_2 for the second is v for the combination. Therefore, if we replace $1/f_1 + 1/f_2$ by the single constant $1/f$, we get for the combination the ordinary equation for a single lens

$$1/u + 1/v = 1/f$$

where

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2} \qquad (10)$$

The reciprocal of the focal length of a lens is sometimes spoken of as its *dioptric strength*, and practical opticians adopt a lens of one meter focal length as the unit, calling it a lens of *one diopter*. A lens of two diopters would then be one of focal length 50 cm., a lens of ¼ diopter one of 400 cm focal length, etc. The above demonstration then shows that when two thin lenses are placed very close together their dioptric strengths are added. This relation holds good even if one of the lenses is diverging, provided we take the sum of the reciprocals of the focal lengths in the algebraic sense, the focal length of the diverging lens being negative. We shall find the principle very useful in discussing "achromatic," or color-free, lenses.

33. Chromatic aberration.—We have already seen that the index of refraction of a substance is different for different wavelengths, or colors; and since the focal length depends upon the index it is obvious that a lens, unlike a mirror, focusses different colors at different points. This is a serious defect in simple lenses, and it would be impossible to have very effective lenses for telescopes, microscopes, or cameras, if it were not possible to avoid it in some degree. Figure 46 is a diagram, plotted to scale, which shows the variation in focal length with the wavelength of the light, for two different lenses, one of crown glass (dotted line) and one of flint glass (full line). Each lens has a focal length of 100 inches for light of wavelength .0000589 cm. (yellow), and the ordinates show the differences between this and the focal length for any other wavelength plotted as abscissa. It is seen that the focal length for the blue differs by more than an inch from that for the yellow in the flint lens, by something less than this in the crown.

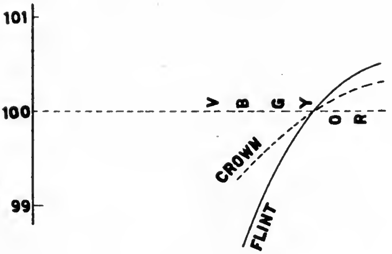

Figure 46

This defect is known as *chromatic aberration*. The figure shows that it is greater for flint than for crown lenses. Not only does flint have a greater index than crown, but its *relative dispersion*, that is, the percent change in index for a given change in wavelength, is also greater. This fact enables us, by combining a crown converging with a flint diverging lens, to produce a combination known as an *achromatic lens*, in which, though the focal length still varies for different wavelengths, the variation is relatively small. The plan adopted is to figure the two lenses so that the focal length of the combination is the same for two chosen wavelengths, say one in the brighter red and one in the greenish blue. It will then be slightly less for wavelengths intermediate between these two, somewhat greater for the deep red and the blue and violet.

34. Achromatic lenses.—In order to explain the production of achromatic lenses by a concrete example, we shall calculate in detail the radii of curvature for an achromatic of 100 inches focal length. We first choose, from a catalogue of optical glasses, two known respectively as "S.40, medium phosphate crown," and "0.335, dense silicate flint." The refractive indices of each of these glasses is given for five different locations in the spectrum, known as the points A′ (wavelength = .00007677cm., deep red), C (wavelength = .00006563cm., bright red), D (wavelength = .00005893cm., orange-yellow), F (wavelength = .00004862cm., blue-green), and G′ (wavelength = .00004341cm., deep blue). The table of indices follows:

	S.40 (crown)	0.335 (flint)
A′	1.55354	1.62621
C	1.55678	1.63197
D	1.5590	1.6372
F	1.56415	1.65028
G′	1.56953	1.66152

We are to find what must be the radii of curvature of the crown glass converging and the flint glass diverging lens, in order that the combination shall have a focal length of 100 inches for the C and also for the F light. To simplify the problem, we shall assume that the second surface of the flint lens is flat, and that its first surface fits exactly over the second surface of the crown, so that the combination will appear like figure 47, which is a very common type of achromatic. Then the radii of the four surfaces, beginning with the left hand, will be indicated by a, b, —b, and ∞.

Figure 47

Let F_c' and F_c'' represent respectively the focal lengths of the crown lens and the flint lens, for the C light. Then, in order that the combination shall have a focal length of 100 inches for C light, we must have, by equation (10),

$$\frac{1}{100} = \frac{1}{F_c'} + \frac{1}{F_c''}$$

From equation (9), or from (7) and (8), we have that for any lens

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$$

therefore, substituting the appropriate values of n, r_1 and r_2, for both lenses, we get

$$\frac{1}{F_c'} = .55678\left(\frac{1}{a} + \frac{1}{b}\right)$$

$$\frac{1}{F_c''} = -.63197\,\frac{1}{b}$$

Therefore

$$\frac{1}{100} = .55678\left(\frac{1}{a} + \frac{1}{b}\right) - .63197\,\frac{1}{b} \qquad (11)$$

Of course we get an exactly analogous equation from the fact that the focal length of the combination is also 100 inches for the F light, viz., the equation

$$\frac{1}{100} = .56415\left(\frac{1}{a} + \frac{1}{b}\right) - .65028\,\frac{1}{b} \qquad (12)$$

In the two equations (11) and (12), we may regard $1/a + 1/b$ on the one hand, and $1/b$ alone on the other, as two unknown quantities, and solve for their numerical values. The result is

$$\frac{1}{a} + \frac{1}{b} = .037269 \qquad \frac{1}{b} = .016998$$

$$a = 49.33 \text{ inches.} \qquad b = 58.84 \text{ inches.}$$

Therefore, if the crown lens be ground with convex surfaces of radius 49.33 in. and 58.84 in. respectively, and the flint lens with one surface concave of radius 58.84 in. and the other surface plane, then the combination will have exactly the same focal length, viz., 100 in., for the bright red and the green-blue light. In order to find the focal length of the combination for other colors, we can make the calculations very simply by using the already found values for the radii, and the appropriate values for the refractive indices. Thus, if F_a, F_d, and F_g represent the focal lengths for the A′ light, the D light, and the G′ light respectively,

$$\frac{1}{F_a} = .55354\left(\frac{1}{a} + \frac{1}{b}\right) - .62621\,\frac{1}{b} = .55354 \times .037269 - .69621 \times .016996 \quad F_a = 100.13 \text{ in.}$$

$$\frac{1}{F_d} = .55900\left(\frac{1}{a} + \frac{1}{b}\right) - .63720\,\frac{1}{b} = .55900 \times .037269 - .63720 \times .016996 \quad F_d = 99.94 \text{ in.}$$

$$\frac{1}{F_g} = .56953\left(\frac{1}{a} + \frac{1}{b}\right) - .66152\,\frac{1}{b} = .56953 \times .037269 - .66152 \times .016996 \quad F_g = 100.17 \text{ in.}$$

These results are plotted in figure 48, to the same scale used in figure 46. Since the lenses for which the latter figure is drawn are supposed to be made from the identical glasses which we have used in our calculation, a comparison of the two figures shows very clearly the superiority of an achromatic lens over a single-piece lens made from either of the glasses composing the achromatic. Simple lenses of crown glass are practically never used, except as spectacle-lenses, as condensers for lantern or microscopes, and in some few other cases where good definition is not required. Lenses are never made from flint glass alone. It would be less suitable than crown, not only because of its greater relative dispersion, but also because flint glasses are generally softer and more easily scratched than crown.

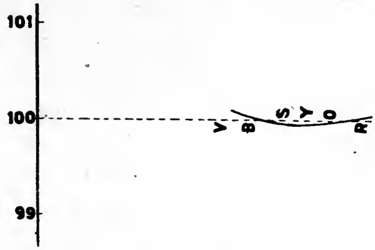

Figure 48

Whenever, as in the example calculated above, the crown and flint components of an achromatic lens have one radius of curvature in common, so that they fit to one another, they are cemented together with Canada Balsam. This procedure prevents part of the loss of light that would otherwise occur by reflection at the two surfaces.

35. Image of extended object. Undeviated ray.—Up to this point in our discussion of mirrors and lenses, we have always supposed the source to be located somewhere on the axis of the mirror or lens. But when we consider the image of an extended object we must enquire what happens when the source lies off the axis, for evidently not all points of an object of any size can lie on the axis. In this book we shall not attempt to give a mathematical treatment of this problem, on account of its difficulty, but merely state the result yielded by such an investigation, as follows: Let O and I in figure 49

Figure 49

be the positions respectively of a source on the axis of the lens and its image, as found by formula (8). Also let O′ be a point off the axis, but lying in a plane perpendicular to the axis through O. Then, provided OO′ is small compared to the distance from the lens, the image of O′, which we shall call I′, is found to lie very nearly in a plane perpendicular to the axis through I. These two planes, both perpendicular to the axis, and so situated that a point in one finds its image in the other, are called *confocal planes.* In order to locate, in the plane through I, that particular point which is the image of O′, we reason as follows: Among all the rays which diverge from O′ there will be one which, striking the first surface of the lens near the point where the axis penetrates it, will be deflected into the glass in such a way that it strikes the second surface at the same angle at which it left the first. For this ray, the lens acts merely as a flat plate of glass with parallel sides, and the ray on emerging will take a direction parallel to that which it had on entering. The desired point I′ will be the point where this ray strikes the confocal plane through I. The ray in question may be called the *undeviated* ray, for although it suffers a slight lateral displacement in traversing the lens, its direction is not changed. The thinner the lens, the smaller will be the lateral displacement of the undeviated ray, and the closer will its point of entrance into the lens and its point of exit coincide with the geometrical center of the lens. Therefore, for thin lenses, we find the image of such a point as O′ by drawing a line from O′ through the center of the lens, and another through I perpendicular to the axis, their intersection giving the location of the image of O to an accuracy sufficient for most practical purposes.

36. Magnification.—Incidentally, this construction enables us to find the size of the image of an extended object such as the arrow OO′ of figure 49. For, since OO′ and II′ subtend equal angles from the center of the lens, they must be proportional to the distance from the lens. That is,

$$\frac{II'}{OO'} = \frac{v}{u} \qquad (13)$$

It is also evident that if object and image lie on opposite sides of the lens, as in figure 49, the image is inverted, while if they lie on the same side it is erect.

Figure 49 is drawn for a converging lens arranged to produce a real image, but the facts stated above hold good whether the lens be converging or diverging, the image real or virtual.

Similar conclusions hold for the reflected images from a mirror. Here also we have confocal planes, such that a point in one has its image in the other, but there is of course no such thing as an undeviated ray, since a change of direction is always present in reflection. However, if we draw from O′ in figure 50 the ray O′P, to the point where the line OI meets

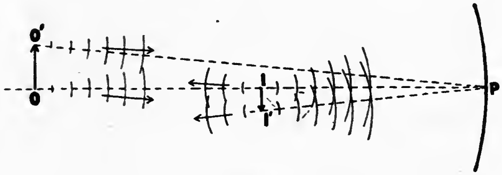

Figure 50

the mirror, the ray O′P will be reflected in the direction PI′, where, by the laws of reflection, the angles O′PO and I′PI are equal, and the intersection of this line with the plane confocal to the plane of O′ will give the image I′. It follows at once that equation (13) holds for mirrors as well as lenses. But in the case of a mirror, the image is inverted if it lies on the same side as the object, erect if on the opposite side.

An important application of the principles just stated is illustrated in figure 51. Suppose there are two stars, practically at an infinite distance, in the direction from the lens indicated by the letters C and D, the arrows indicating the

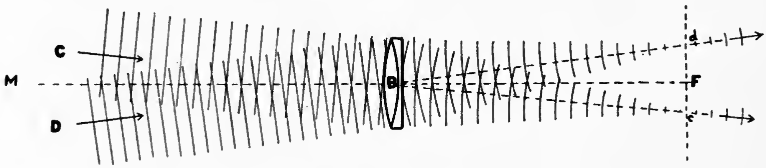

Figure 51

direction in which the light is propagated from them, in plane waves. If either star lay on the prolongation of the axis BM, its light would be focussed at the principal focus F. Otherwise, if the angles CBM and DBM are small, not more than a few degrees, the image of each star will lie in what is called the *principal focal plane* of the lens, a plane through F perpendicular to the axis. For plane waves, every line perpendicular to the wavefronts is a ray. Therefore, if we draw through the center of the lens a line perpendicular to each set of incident wavefronts, these will be the undeviated rays from the two stars; and the points where these lines meet the principal focal plane, c and d, will be the images of the two stars. The angle cBd, subtended by the images from the center of the lens, is then equal to the angle subtended by the stars themselves from the center of the lens—or indeed from any terrestrial point since the distance of the stars is so great. This angle is given in radian measure, to a sufficiently close approximation, as the quotient of the distance cd divided by the focal length.

This principle is used in practical astronomy for measuring the angular distance between double stars. There are two methods of measuring the distance cd between the images. One is to place a photographic plate directly in the focal plane, expose it to the light from the stars, and then develop it by the usual photographic processes, which leave a little blackened dot where each star-image falls. The distance between these dots is accurately measured on a dividing-engine.

37. Micrometer.—The other method is to use a micrometer, the essential part of which is a small metal frame arranged so that it can slide in a plane perpendicular to the axis of the lens. A fine spider-thread is stretched across this frame, so that it lies in the focal plane, perpendicular to the direction in which the frame slides. The whole micrometer is turned about the axis of the lens, till the direction in which the frame slides is parallel to the line joining the two star-images, and the frame is then moved by a fine-pitched screw to which it is attached, so that the spider-thread, commonly called the cross-hair, lies first on one image, then on the other. The pitch of the screw is known, so that the number of its revolutions necessary to move the cross-hair from one image to the other gives the distance. In order to make the cross-hair and star-images clearly visible, a short-focus lens, or combination of lenses, called the eyepiece, is placed just behind the focal plane. The eye sees magnified images of the cross-hair and the two original star-images.

A similar method is used in certain surveying instruments, although the conditions are somewhat different. The objects observed are not infinitely distant, and the images are consequently not formed exactly in the principal focal plane.

38. Imperfections of mirrors and lenses.—The student will no doubt have drawn for himself the conclusion that, quite apart from chromatic aberration in lenses, both lenses and mirrors are far from perfect optical instruments, since our formulæ are only approximations. Such a conclusion is undoubtedly correct, and it will be worth while to enumerate the more common faults. We shall speak principally of lenses, but what is said applies also to mirrors, for all the faults found in lenses, except those due to chromatic aberration and absorption, are also shared by mirrors, in many cases to a greater degree.

In the first place, owing to the fact that light is a wave motion, and therefore does not travel absolutely in straight lines, no optical instrument, whether it be made up of lenses, mirrors, or other elements, and no matter how perfect the workmanship may be, can produce from a point source an image which is a real mathematical point. For instance, although a star, on account of its great distance, may be regarded as a point source, its image as produced by the most perfect telescope is not a mathematical point, but a very small disc surrounded by a series of faint rings. If the lens is well made and of large diameter, the diameter of the disc and the surrounding rings is so small that the latter can be seen only by highly magnifying them, and they become a source of trouble only in the most exacting work with telescope or microscope. The nature of this fault will be considered later under the head of *diffraction*, sections 72 and 73.

39. Spherical aberration.—Another fault is known as *spherical aberration.* Quite apart from the just-mentioned difficulty of diffraction, and from chromatic aberration, the rays coming through the edges of a lens are not brought to the same focus as those coming through near the center. This follows from the fact mentioned above, that when a spherical wavefront is refracted at a spherical surface, it emerges not truly spherical. Figure 52 illustrates this defect in an exaggerated manner. The rays are drawn, but not the wavefronts.

O is a point source, from which all the rays originate. When they emerge from the lens, they do not converge to a single point. Since the central part of the emergent wavefront, say the part to which rays between those marked 5 and 7 belong,

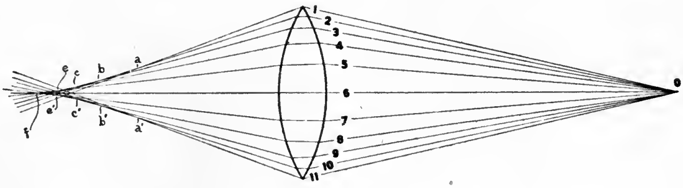

Figure 52

is very nearly spherical, these rays will all intersect nearly at a single point f. Rays 4 and 5, however, cross before reaching f, at such a point as e, and the corresponding rays 7 and 8 at e′. Rays 3 and 4 will cross still nearer the lens, as at c, 3 and 2 at b, 2 and 1 at a, etc. Consequently, instead of having a single point f as the image of O, we may say that the image is the line abcefe′c′b′a′, or rather, the *surface* formed by revolving this line about the axis of the lens. This surface is roughly conical, with a point or "cusp" at f, and by far the greater part of the light is concentrated at this point, which we commonly regard as the proper image. Nevertheless, much light fails to pass through f, and if a screen were placed at that point we should see a sort of halo surrounding the bright center, caused by light which came to focus before reaching the screen. The curve abcefe′c′b′a′ is called a *caustic,* and the corresponding surface a *caustic surface.* A familiar example of a caustic is the socalled "cow's hoof" seen on the surface of a glass of milk. It is formed by reflection from the inner surface of the rim of the glass, which acts as a concave cylindrical mirror, reflecting light from a nearby window, or any other conveniently placed source of light.

40. Curvature of field.—Still another defect is *curvature of the field.* Referring to figure 49, it was stated in the text that if O′ lies in the plane of O, its image I′ is very nearly in the plane of I, provided that the distance OO′ is small compared to the distances of O and O′ from the lens; and the two planes perpendicular to the axis were called confocal plane. More accurately, the surface which is confocal to the plane through O is not a plane but a slightly curved surface, concave toward the lens and nearly plane in the neighborhood of the axis. As a special case, suppose O and O′ are so far away that the waves reaching the lens from them are practically plane. We may then regard O, O′, and all other sufficiently distant objects as being in a plane perpendicular to the axis of the lens. Under these circumstances, if a screen be placed perpendicular to the axis at the principal focus, those of the distant objects which subtend only a small angle with the axis will be sharply in focus on the screen, but the others will be blurred, and the screen must be moved closer to the lens to bring them into sharp focus.

41. Astigmatism.—This is a fault which shows itself particularly for pencils of light which strike a lens or mirror diagonally. Under such circumstances, the image of a true point tends to become a pair of short lines, perpendicular to one another, but not intersecting. Figure 53 is intended to

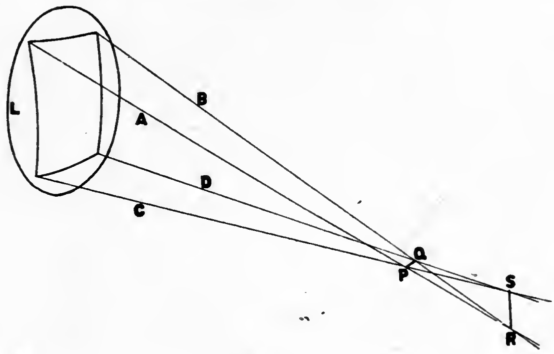

Figure 53

make this plain. The ellipse L represents a lens seen in perspective. For convenience in explanation, suppose an opaque piece of paper pasted over the face of the lens, with a square hole, so that the beam that comes through is limited to what passes through this square aperture. Only the rays coming from the four corners are shown. Rays A and B intersect at the point R, but A and C intersect at P, nearer the lens. Similarly, C and D intersect at S, B and D at Q. Rays B and C do not intersect at all, neither do A and D. There will be a horizontal focal line PQ, for the intersection of rays in a vertical plane, and a vertical focal line SR, for the intersection of rays in a horizontal plane. The two focal lines can be shown very nicely by holding a converging lens, or better still a concave mirror, so that it receives somewhat obliquely the light from the sun, and moving a white card back and forth till the two focal lines are found.

42. Lenses for special purposes.—In spite of this long list of faults, lenses function in a very satisfactory manner for most purposes. Consider for instance the lenses used as objectives in telescopes. (Telescopes are considered in detail in the following chapter.) Chromatic aberration is hardly perceptible if the lens is properly constructed of flint and crown glass in the manner already described, for the focal length is very nearly the same for all wavelengths except the blue and violet; and since these colors in most light-sources have feeble luminosity, their being somewhat out of focus hardly affects the sharpness of the images. Spherical aberration is negligible because the area of the lens is made small enough so that only that part of the emergent wavefront is used which is nearly spherical. As a rule, the diameter of a telescope objective is somewhere between 1/30 and 1/16 of the focal length, making the angular diameter of the cone of light which it transmits relatively small. Finally, the curvature of the field and astigmatism produce a negligible effect, on account of the smallness of the field. In astronomical telescopes, it is seldom necessary to use a field of as much as one degree. Consequently, only the flatter portion of the field is used, and no pencil of light that is visible in the eyepiece traverses the lens with enough obliquity to cause appreciable astigmatism. The design of a telescope objective is therefore relatively simple, and its excellence depends mainly on the quality of the workmanship and the homogeneity of the glass. This latter condition is by no means easy to fulfill in such large discs of glass as were used in making the objective of the Lick telescope (36 inches in diameter) or that of the Yerkes (40 inches).

Camera lenses are used under more exacting circumstances. In order that the lens may be "fast," that is, give sufficient illumination with very short exposure-time, it must have a diameter as great as 1/5 or 1/6 the focal length, giving great opportunity for spherical aberration. Moreover, the extent of field used is large, since the dimensions of the photographic plate are nearly as great as the focal length. Consequently, troubles due to curvature of the field and astigmatism are likely to appear. The manufacturers of photographic lenses have, however, achieved remarkable success in combating these difficulties, so that a first-class lens shows them to only a limited extent. Even with the best lenses, however, if they are used with full aperture, the corners of the picture are slightly out of focus due to curvature of field, and show a slight drawing out of points into lines, which is the result of astigmatism. A good photographic lens is made in two parts, separated by an air-space, and each part is composed of several pieces of glass. It is by altering the composition of these separate pieces and the curvature of their surfaces that the designers have succeeded in reducing largely, but not entirely, the inherent lens-defects. A photographic lens cannot properly be regarded as a thin lens.

Microscopic objectives also, if of high power, are objects of elaborate design, consisting of many pieces of glass. The thickness of such a so-called lens (really it is a combination of a number of lenses) is much greater than the equivalent focal length of the combination.

Problems.

1. If a plate-glass window, index 1.58, appears to one looking into it to be 8 mm. thick, what is the actual thickness?

2. What must be the radius of curvature of a symmetrical converging lens of crown glass, to have a focal length 1 meter?

3. An object is 4 ft. from a white screen. Find two positions in which a lens of 8 inch focus can be placed, to form an image of the object on the screen.

4. An object is 8 inches from a screen. Where should a concave mirror of 2 foot radius be placed to form an image of the object on the screen.

5. Show that problem 3 cannot be solved if the focal length of the lens is more than 12 inches.

6. A lens of 3 foot focus forms images of two stars in its principal focal plane, and a micrometer is used to find tho distance between the images. It takes 12.85 turns of the screw to move the cross-hair from one image to the other, and the

screw has 50 threads to the inch. Find the angle between the stars.

7. Show how an achromatic diverging lens can be made, and write the equations from which the curvature of its surfaces can be found, using the data for the glasses given in paragraph 38.

8. In what position must the eye be placed, to see an image formed by a lens or mirror, if a screen is not used? Why is it usually easier to find a virtual than a real image?

9. If a camera lens has a focal length of 8 inches, find the proper position for focus on an object 5 ft. away, and the length of this image if the object is 2 ft. long.

10. Show that if a camera of focus 6 inches is focussed for infinitely distant objects, any object more than 40 ft. away will be less than 5/64 inch out of focus.

11. Explain why "depth of focus" in a camera is impossible to obtain without sacrificing "speed."

CHAPTER VI.

43. The telescope.—44. Magnifying power.—45. Ramsden eyepiece. —46. Opera glass.—47. Prism binocular.—48. Reflecting telescopes.— 49. Simple microscope.—50. Compound microscope.—51. Projection lanterns.

43. The telescope.—The essential part of a telescope is two lenses,—a long-focus, large diameter achromatic, turned toward the object in view and therefore known as the *objective*, and a smaller lens (or, as we shall see later, more commonly a pair of lenses) called the eyepiece. Figure 54 shows a simple diagram of a telescope. The object viewed is supposed to be an arrow, very far away, but so large that in spite of distance it covers an angle of a degree or so. If this conception seems too artificial, we may think of the point of the arrow as representing one star, the butt another. Wavefronts are not indicated, but lines are drawn to show the course, through the instrument, of the cone of light from each end of the arrow. Dotted lines show the undeviated rays for each lens.

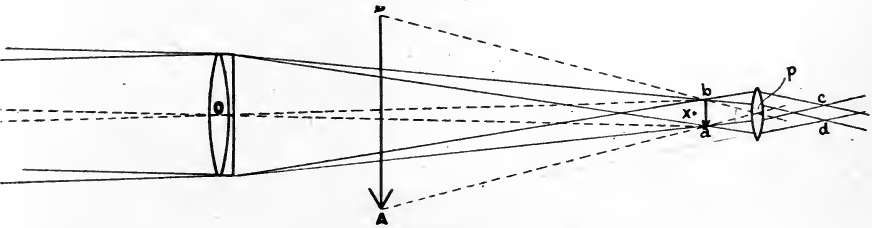

Figure 54

A real inverted image of the object is formed in the focal plane of the objective, from which the waves continue on, diverging from this image exactly as if it were a material object, except that the light is limited to a comparatively small cone. This light falls upon the eyepiece, which forms with it a second image, really an image of an image. Since the rays that form any point of the first image are limited to the cone that comes through the objective, it may well happen that the *undeviated ray* drawn from this point through the center of the eyepiece lies outside the cone and therefore does not exist as a real ray. But the position of the second image must certainly be independent of the diameter of the objective, and therefore we are at liberty in such a case to find that position by drawing fictitious undeviated rays just as if they really did exist. The figure is drawn for such a case. The position of the second image depends of course upon the location of the eyepiece which is mounted so that the observer can slide it at will through a short distance toward or away from the objective. Most observers place it so that the first image lies a little within its principal focus. Then the second image, the one which the eye sees, is virtual, still inverted, and on the same side as the first image, but farther away. This is shown by AB in the figure. If the principal focus x were placed just at the real image ab, as is sometimes done by persons of far-sighted or normal vision, AB would be thrown back to infinity, like the original object, but would still subtend a much greater angle than the latter.

44. Magnifying power.—We take as a measure of the magnifying power of the telescope the ratio of the angle subtended by the image AB to that subtended by the original object, and in calculating its numerical value we assume, for the sake of definiteness, that the principal focus of the eyepiece coincides exactly with that of the objective, that is, that x lies exactly on ab, putting AB at an infinite distance. With both the original object and the final image so far away, it does not matter what point is chosen as the apex of the angles subtended. That subtended by the object is aOb, which, in radian units, has the value ab/F, F being the focal length of the objective. That subtended by the image AB is ApB, whose value is ab/f, f being the focal length of the eyepiece. Therefore the magnifying power is

$$M = \frac{\frac{ab}{f}}{\frac{ab}{F}} = \frac{F}{f}$$

Therefore, for high magnifying power, we should use a long-focus objective and a short-focus eyepiece. Usually, a large telescope is provided with several eyepieces of different focal length, so that the magnifying power can be changed at will. For some purposes, high magnification is desirable, for others not. The higher the magnification, the smaller the visible field; that is the smaller the area that can be seen at once. For instance, with high magnification only a very small part of the surface of the sun or moon can be seen.

There are other practical limitations to the magnifying power that can be used with advantage. As we have previously stated, the real image produced by the objective is not a true picture of the object, but is slightly hazy at the edges, and is surrounded by faint diffraction rings. Any increase in magnification beyond the point where these rings or bands become visible is useless, for it merely magnifies the bands and the haziness along with the rest of the image and contributes nothing to distinctness of vision. Astronomers also find that certain conditions of the atmosphere, caused no doubt by irregularities in density, produce a haziness or fuzziness of image known as "bad seeing," which is worse than the diffraction difficulty. In fact, with a large telescope, the appearance of the diffraction bands on high magnification indicates that the "seeing is good," for poor seeing conditions cause them to be blurred out of recognition. With only moderately good seeing, an astronomer will use moderate magnification, and with very bad seeing he will abstain from observing at all.

Figure 54 shows that at a certain place cd the two cones of light from the head and the butt of the arrow cross. In fact, there is a little circle at this place through which passes every cone of light that traverses the telescope, and it is not hard to show that this circle is nothing more nor less than the image of the objective lens formed by the eyepiece. For best vision, the eye should be held so that its pupil coincides with this small circle, which, by the way, is called the *exit-pupil* of the telescope. For the figure shows clearly that if the eye is held much closer to the eyepiece, or much farther from it, only cones of light from the middle part of the arrow (cones not drawn in the figure) would enter the pupil of the eye, unless the latter were very large. To see the whole image AB at once would then be impossible, though one could see different parts of it at a time by moving the eye up or down, so as to receive the light from those parts. That part of the image that can be seen in any one position of the eye is called the *field of view*, and it is greatest when the pupil of the eye coincides with the exit-pupil. If the eye is in this most favorable position, the field of view is still limited by the diameter and the position of the eyepiece. Reference to the figure shows that if the arrow were much larger than it is there made, the cones of light from the ends of the image would partly or wholly miss the eyepiece. If they missed the latter entirely, these ends would be completely invisible in the telescope, while if only part of the cone fell on it, the image would be faint at the ends. The simple eyepiece shown in the figure is not suitable for producing a large and uniformly illuminated field of view. It is desirable that the eyepiece should come very close to the real image ab; but in order to have this occur with a single-lens eyepiece, the focal length of the latter would have to be inordinately short; and if, in addition, the diameter were made great, spherical aberration and other lens defects would become too pronounced.

45. Ramsden eyepiece.—For the reasons outlined above, much ingenuity has been applied in devising eyepieces which consist of combinations of lenses instead of single lenses, so as to give a larger field for the same magnifying power. The best known of these is the Ramsden eyepiece, which functions very satisfactorily, and is used on most telescopes. It is shown in figure 55. The objective of the telescope has been omitted from the drawing, but the real image ab and cones of light forming its ends are shown just as they are in figure 54. The eyepiece consists of two identical planoconvex lenses, mounted rigidly in a metal tube, and separated by a distance equal to ⅔ the focal length of either. It can be shown that such a pair is equivalent, so far as magnification is concerned, to a single lens whose focal length is ¾ that of either component. The front lens, called the *field-lens,* is placed very close to the real image ab, and therefore forms from it a virtual image a′b′, slightly larger, and slightly farther away. In fact, a′b′, comes just at the principal focus of the rear lens of the combination (called the *eye-lens*) or just within it. Accordingly, this latter forms the final virtual image, AB of figure 54, either at infinity

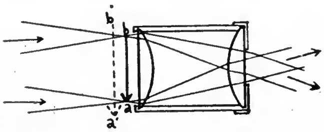

Figure 55

or at whatever distance from the eye is most suitable for the observer, adjustment being secured by sliding the whole eyepiece toward or away from the real image ab. Since ab lies close to the field lens, the latter receives and transmits light from an area of the image practically equal to the area of the lens itself, thus giving a field whose diameter is approximately equal to the diameter of the field-lens. The pencils from all parts of the image cross the axis just behind the eye-lens. This is the most convenient place for the exit-pupil, for the eye can then be placed close up to the end of the eyepiece. It is found further that this combination of two planoconvex lenses is almost free from spherical aberration, and the chromatic aberration is of such a nature as to be hardly perceptible.

If a micrometer is used with the telescope, it is placed so that the crosshairs move exactly in the plane ab in which the real image lies. The proper method of adjusting the telescope, known as *focussing*, is as follows: First, the eyepiece is pushed in or out until the crosshairs are visible, clearly and without eyestrain. Then, by means of a suitable slide in the telescope tube, the eyepiece and crosshairs together are pushed in or out till the image of the object viewed is also seen clearly, and there is no parallax between the image and the crosshairs.

One defect of the kind of telescope we have been describing is that the image is inverted. This is not a disadvantage in astronomical telescopes, but for terrestrial telescopes it is inconvenient. In such instruments the image is usually reinverted by making the tube of the instrument very long, and inserting between the eyepiece and the real image formed by the objective another lens or pair of lenses of rather short focal length, whose function is to receive the light from the real image (ab of figure 54) and form therefrom another real image which is reinverted and therefore right side up. The eyepiece then forms from this image a magnified virtual image which the eye sees. The great length of tube necessary in this form of instrument makes it inconvenient except for small spyglasses.

46. Opera glass.—In the old-fashioned opera glass, shown in diagram in figure 56, the erection of the image is provided for by using a diverging lens for the eyepiece. This form of telescope is quite short; for, in order that the eyepiece may magnify the image, it is necessary that it should intercept the light between the objective and the latter's principal focus. Thus the real image ab is not actually formed at all, for it would come behind the eyepiece instead of before it. Con-

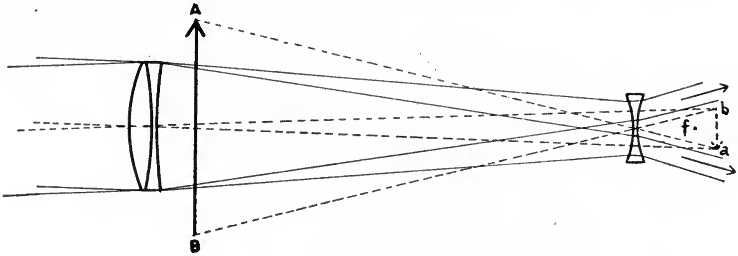

Figure 56

verging waves strike the eyepiece, with their centers on ab. Therefore, in the formula

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

or its equivalent

$$v = \frac{uf}{u - f}$$

u must be taken as negative. Since f is also negative, because the lens is diverging, the numerator of the fraction in the last equation is positive, and the sign of v depends upon the sign of u — f, where both u and f are essentially negative. In order to have a virtual image AB, v must be negative, therefore u must be greater in absolute amount than f; that is, the principal focus of the eyepiece must lie between the latter and the place where the image ab would come if the lens were removed For example, let f = —5cm., u = —5.2cm. Then v = —130 cm. The observed image AB would then lie 130 cm. away on the side from which the light comes, and would be virtual. Also, it would be inverted as regards ab, erect as regards the original object. The erectness of image, and its shortness make this type of telescope convenient, but unfortunately its field is quite small. The diagram shows that there is no real exit-pupil, as there is in the ordinary form of telescope; that is there is no place where the eye can be placed so that it will receive every cone of light that does not miss the eyepiece. The eye must be moved about in order to see the whole image of any object viewed, unless the object be relatively small. For this reason, such instruments are made only with small magnifying powers, say two or three diameters.

47. Prism binocular.—The modern prism-binocular is a great improvement over the form of telescope described above. It is essentially a telescope of the form described (or rather, a pair of such telescopes, one for each eye), in which the length is very much reduced by four reflections in totally reflecting prisms. Incidentally, the series of reflections reinverts the image, so that it is possible to use an eyepiece like the Ramsden, with its large field. An additional advantage lies in the fact that the arrangement of the two telescopes brings the objectives farther apart than the two eyes. This, being equivalent to a wider spacing of the eyes themselves, greatly increases the stereoscopic effect, or parallax, and brings the field into strong and pleasing relief.

48. Reflecting telescopes.—A telescope composed of objective lens and eyepiece is known among astronomers as a *refractor*. A *reflector* is a telescope in which a concave mirror is substituted for the objective lens. At the present time the use of reflectors is confined almost exclusively to photographic work, for which purpose they possess several decided advantages. In the first place, they are free from chromatic aberration. Secondly, they have no absorption, and this is very important in photography, for the transmission of light through glass causes much of the photographically active ultraviolet light to be lost by absorption. Finally, by making the concave reflecting surface parabolic instead of spherical, the principal focus is rendered absolutely free from spherical aberration, and a small region in its neighborhood almost so, so that most beautiful definition is secured in photographing objects of such small angular dimensions as a star-cluster or a small nebula.

The concave mirror is usually placed at one end of a long tube or frame work, the other end of which is open and pointed toward the celestial body to be photographed. Between the mirror and its principal focus, is placed a small plane mirror, set at an angle of 45° with the axis of the instrument. This reflects the light coming from the concave mirror to the side of the tube, where the photographic plate is exposed in the reflected position of the focal plane. A Ramsden eyepiece may also be placed in position to receive the light, but this has no function during the photographic process. It may be used however in pointing the instrument to the desired object. There are, however, other forms of reflecting telescopes.

49. Simple microscope.—The word microscope usually means an instrument used for magnifying small objects close at hand which, like a telescope, has two optical parts, objective and eyepiece; but in stricter language such an instrument is called a *compound microscope,* while the name *simple microscope* is applied to a single lens used as a magnifier of low power.

A simple microscope is held as close to the eye as convenient, and the object to be examined is placed somewhat within the principal focus, so that the eye sees a magnified virtual image of it at the distance which is most suitable for distinct vision. For the normal eye this is about 25cm., though it differs with different individuals. Figure 57 shows the arrangement. L is the magnifying lens, F its principal focus,

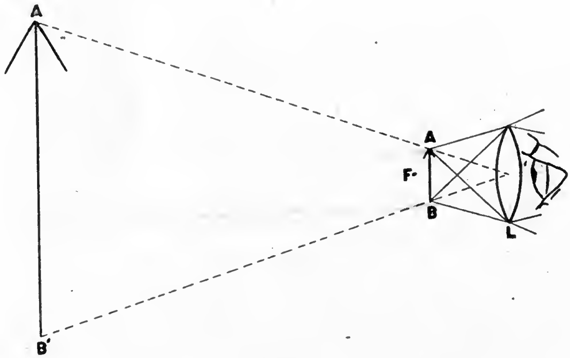

Figure 57

AB the object, A′B′ the image seen by the eye. We take as a measure of the magnifying power the ratio of the angle which the image subtends at the eye to that which the object itself would subtend if the lens were removed and the object put back where it could be seen most distinctly, that is, where the image is in the figure. Since magnifying powers need only be known roughly, and since the eye is placed so close to the lens, it will suffice to consider the center of the lens, instead of the pupil of the eye, as the apex of the angles. The angle subtended by the image is

$$A'B'/v = AB/u$$

The angle which the object would subtend if placed at the distance v is

$$AB/v$$

Therefore the magnifying power is

$$AB/u \div AB/v = \frac{u}{v}$$

From the law of lenses

$$\frac{1}{u} - \frac{1}{v} = \frac{1}{f}$$

$$\frac{v}{u} - 1 = \frac{v}{f}$$

$$\frac{v}{u} = 1 + \frac{v}{f}$$

Therefore the magnifying power is $1 + v/f$ or $1 + 25/f$ if f is expressed in centimeters. As an example, the magnifying power of a lens of 5cm. focal length is 6.

A Ramsden eyepiece used as a single lens makes a very good simple microscope.

50. Compound microscope.—It is not practicable to get very high magnifying power with a single lens, for that would require such a short focal length that the lens defects such as chromatic and spherical aberration, curvature of field, etc., would be very prominent. Therefore, wherever high magnifying power is necessary, it is provided by a compound microscope. Figure 58 shows the arrangement of parts. The "object" is usually in the form of a slide, a very thin slice of the material to be examined enclosed between two thin glass plates. The slide is represented by the short arrow A in the diagram. Slides being more or less transparent, they are examined by transmitted light. A mirror M and a condenser C concentrate upon the slide, from below, a beam of light from a window or other broad illuminated area. The set of lenses forming the condenser are of low grade, for their function is merely to provide illumination, not to form a clear image of anything.

The mirror and condensers are accessories rather than parts of the microscope proper. The latter consists of an objective O and an eyepiece FE. The former is shown in the figure as a

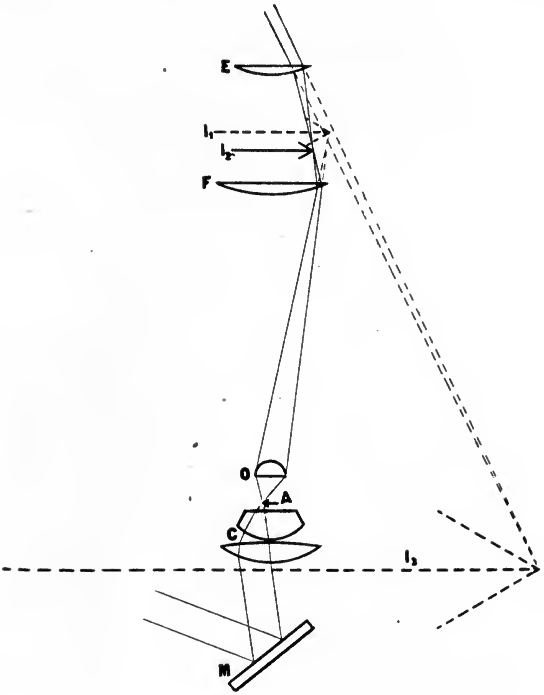

Figure 58

single hemispherical lens, but it is in fact compounded of several distinct units in order to secure chromatic and other corrections. The eyepiece shown in the figure is what is called the Huyghens type, which has certain advantages over the Ramsden eyepiece described in connection with telescopes, though it has also the disadvantage that it cannot be used in connection with crosshairs or micrometer. If it is necessary, as sometimes happens, to put a micrometer on a microscope, the Huyghens eyepiece must be exchanged for one of the Ramsden type.

The objective would, but for the eyepiece, form a magnified real inverted image of the slide at I_1, but the field-lens F intercepts the converging light directed toward this image, and converges it still more, forming the real image I_2, slightly smaller and slightly lower. Usually a diaphragm is placed in the plane of I_2 so as to limit the visible field to a small circle over which the illumination is uniform. The light diverging from I_2 then passes through the eye-lens E, which forms a magnified virtual image at I_3, the image which the eye sees. For very high powers, an "oil-immersion" objective is used. The objective comes very close to the slide, and the space between is filled with a drop of oil having an index of refraction nearly the same as that of glass, so that one might regard the object as being imbedded in the objective. Under these circumstances, the resolving power of the microscope is somewhat increased, and the brightness is also increased because less light is lost by reflection from the bottom surface of the objective and the top surface of the slide.

Certain particles too small to be seen with the bright background illumination commonly used in a microscope, such as the particles in certain colloidal solutions, can be seen as bright points against a dark background if the illumination comes from the sides instead of from below. This is the principle of the so-called "ultramicroscope."

In the figure, the fainter lines represent the full beam of light from the point of the arrow through the objective and eyepiece. The same rays are shown below, from before they impinge upon the mirror till they strike the slide. Similar rays from the butt of the arrow, or from any other part of it, could be drawn, but they are omitted for the sake of simplicity.

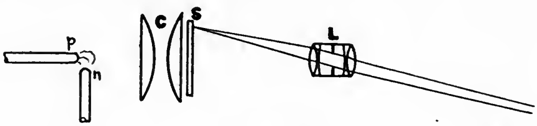

Figure 59

51. Projection lanterns.—Figure 59 shows the ordinary projection lantern, for throwing upon a screen an enlarged image of a lantern-slide. The slide S is so placed with reference to the projecting lens L that the screen comes at the conjugate focus, and the focal length of L must be chosen with due regard to the distance of the screen and the desired magnification. The rest of the apparatus is for obtaining the necessary illumination of the slide. Since the latter is more or less trasparent, the illumination is supplied by transmitted light.

p is the positive, n the negative carbon of an arc-lamp, and C a condenser consisting of two plano-convex rough lenses. The slide should lie as close to the condenser as convenient, and the arc should be so placed that its light is focussed by the condenser through the slide upon the center of the projection lens, forming an image of the arc there. The lens L should be achromatic, in order to avoid chromatic aberration at the screen; and it should consist of two units with a diaphragm between, otherwise the picture on the screen may show some distortion. Except for these two points L need not be a high-grade lens. The pencil of light coming from any point on the slide is so narrow, as shown in the figure, that there is little opportunity for spherical aberration, astigmatism, or curvature of field to show any bad effect.

The focussing is done by moving the lens L. If the illumination of the image is not uniform, this is an indication that the arc is not in the proper place, or that the negative carbon is shutting off some of the light from the positive carbon. Lately it has become more common to substitute a high-power filament lamp instead of the arc. This makes the lantern much easier to operate, and the illumination, though slightly weaker, is strong enough in most cases.

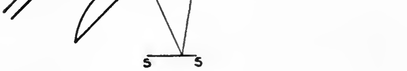

Figure 60

The opaque projection lantern, one form of which is shown in figure 60, is used for projecting images of postcards, pictures or printed matter in books, etc. It differs from the slide-lantern in two essential points. First, it is necessary to introduce a reflection (mirror m in the figure) between picture and screen, to prevent the image from being either upside down or right side to left. Second, the original picture must be illuminated from the front, since transmitted light is out of the question, and this makes it difficult to get the illumination strong enough. The projection lens L must be of exceptionally large diameter for its focal length, and it must be well corrected for all the defects of lenses, since it receives a full beam of light. The arc is made to give an exceptionally large amount of light by using large carbons and a very heavy current, and the condenser system is made to cover a very large angle from the arc. Usually a glass cell containing water is interposed, to cut out much of the infrared light, which would unduly heat the picture. For use in small rooms, where the picture on the screen need not be very large, a high-power tungsten filament lamp may replace the arc, but in such a case it is necessary to use a special screen made of filled canvas covered with aluminum paint, which reflects more strongly than a simple white screen.

Problems.

1. The objective of a telescope has a focal length of 30 ft. What is the magnifying power, when an eyepiece of focal length ½ inch is used?

2. Explain why it is that dirt, or even a large opaque obstacle, on the surface of the objective of a telescope, is never visible to a person looking through the eyepiece, the only apparent effect being a general dimming of the image.

3. Prove that the "exit-pupil" is the image of the objective as formed by the eyepiece.

4. A projection lantern is being planned for use in a certain room. The screen is to be 30 ft. from the slide, and it is desired that the image of the slide on the screen shall measure 58.5×72 inches. (A slide is 3.25×4 inches). What must be the focal length of the projection lens?

5. Explain completely why strong illumination is so much harder to obtain with the opaque projection lantern than with the ordinary lantern for slides.

CHAPTER VII.

52. Prism spectroscope.—We have already considered the spectroscope to some extent, in the chapter on color, but we are now in a better position to understand its principles. The essential parts of a prism spectroscope, shown in figure 61, are the collimator C, the prism P (or a train of prisms), and the telescope T. The collimator is a tube, at one end of which is an achromatic lens, at the other end a fine slit. The latter is carefully adjusted at the principal focus of the former, so that light which enters the slit and passes through the lens emerges in accurately plane waves. The beam then passes through the prism and is *dispersed*, that is, the different wavelengths are deviated to different amounts. It is best to have the prism turned so that the region of the spectrum to be examined traverses it at minimum deviation.

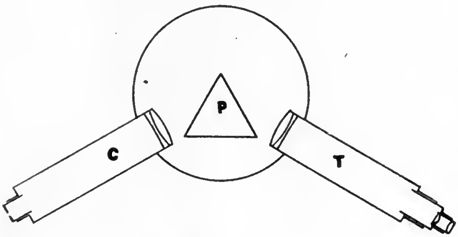

Figure 61

It should be borne in mind that when the light leaves the prism, although the different wavelengths take different directions, yet all the rays of any one wavelength remain parallel to one another till they strike the objective of the telescope. Therefore the latter converges the waves of each particular length to a definite place in the principal focal plane. Consequently, there will be in this plane an image of the slit for each particular wavelength that enters the slit, and the whole array of these images constitute the *spectrum* of the light in question. The eyepiece of the telescope forms an enlarged virtual image of the spectrum, just as it would form a virtual image of any very distant object which the objective focussed in its focal plane.

A *spectrometer* is a spectroscope provided with a large divided circle, so that the angular position of the prism, of the telescope, or of both, can be accurately measured. There is also a crosshair at the principal focus of the telescope. The spectrometer is used for making accurate measurements of the refractive indices of prisms, and for other angular measurements.

A *spectrograph* is a spectroscope so arranged that the spectrum can be photographed instead of viewed directly. Any spectroscope can be converted into a spectrograph by removing the eyepiece from the telescope and putting a photographic plate in the focal plane of the objective; but it is better to remove the whole telescope, replacing it by what is really a long-focus camera, a light-tight box with a specially corrected photographic lens at one end and a holder for photographic plates at the other. An ordinary hand camera, focussed for distant objects, might be used instead, but in most cameras the focal length is rather short, and this causes the spectrum also to be short. since its length is proportional to the focal length of the projecting lens, other things being equal.

53. Bright-line spectra.—The character of the spectrum seen in a spectroscope varies greatly with the chemical nature and physical condition of the body emitting the light that passes in through the slit. The flame of a Bunsen burner, except for the small bluish inner cone, is practically invisible, and if such a flame is placed before the slit, nothing is seen on looking into the spectroscope, as we should expect. But if a piece of asbestos soaked in a solution of common salt (sodium chloride, NaCl) or of any other compound of sodium, is put into the edge of the flame, the latter immediately becomes yellow in color. If this yellow light enters the slit, the spectrum shows two fine yellow lines, that is two yellow images of the slit. The light has a slightly different wavelength in these images, .00005890cm. in one and .00005896cm. in the other.* We interpret the appearance of these lines as follows: atoms of sodium pass into the flame and, under the conditions existing there, start into vibration with two different periods, thus starting in the ether waves of the two different wavelengths given.

The fact that these particular lines appear in the spectrum, whatever compound of sodium be used, proves that it is the sodium, not the chlorine, that is responsible for them. Chlorine produces no color in a flame, though it can be excited to radiation by an electric spark. The appearance, in the spectrum of any source of light, of two bright lines of wavelength .00005890 cm. and .00005896 cm. is therefore a sure and delicate test for the presence of sodium. The merest traces of sodium, far too slight for detection by chemical methods, produce the distinctive coloration in a flame. One cannot, however, assume that a yellow color alone indicates sodium, for yellow includes a considerable range of wavelengths, and some other elements have in their spectra yellow lines, of wavelength different from those attributable to sodium. For instance, if the light from a Cooper-Hewitt electric lamp (mercury arc) be examined with a spectroscope, it shows a number of lines, two of which have wavelengths .00005769cm. and .00005790cm., which brings them in the yellow region, but a slightly different part of the yellow from the sodium lines.

In the Bunsen flame, sodium never shows anything but the two above-mentioned lines and a very faint green one, but there are circumstances in which it gives in addition a number of other lines, as when metallic sodium or a salt of sodium is put into the crater of a carbon arc-light.

*The two lines are so nearly the same in wavelength that when a single small prism is used in the spectroscope they appear as a single line. Spectroscopes of higher power show them as separate and distinct, and the most powerful even show that the wavelength is not absolutely definite in either line. Each line has a small but perceptible *width*, showing that for each the wavelength varies between certain narrow limits. This statement is also true of all other spectrum lines, and it is impossible to obtain a beam of light all of which has *exactly* the same wavelength.

In order to study the spectra of the different elements, various means must be employed to get them into a condition where they emit their characteristic radiations. There are only a few elements which give their spectra in a flame, like sodium. An effective method in the case of metallic elements which do not rapidly oxidize or suffer other chemical change in air, is to pass an electric spark between points made from the metal in question, the spark being operated by an induction coil or transformer in parallel with a condenser. Another method is to take two carbon rods, bore a hole in one of them, fill it with the metal, and connect both rods to the terminals of a direct-current supply of not too low voltage with some resistance in series. When the ends of the rods are touched together and separated about a quarter of an inch, the intense heating at the point of contact vaporizes the carbon and forms a bridge of glowing vapor called the *arc,* across which the current continues to flow. Some of the element packed into the hole in the rod also vaporizes and contributes its vapor to the formation of the arc. If the light from the arc itself (the bridge of vapor, not the glowing ends of the rods) is passed through the slit of the spectroscope, a large number of lines appear, some of which are due to the element in question, some to gaseous compounds of carbon, and some to such impurities as are always present in the rods.

The spectrum of a gas is usually obtained by the use of a so-called *vacuum-tube.* This is a glass tube with a restricted middle portion, into opposite ends of which are sealed metallic terminals. The tube is evacuated of air, and enough of the gas is put in to exert a pressure of a few millimeters of mercury, after which the tube is hermetically sealed. The electric discharge of an induction coil is sent through the tube, from one terminal to the other, causing the gas inside to become luminous and emit its characteristic wavelengths.

Every known element has more than one line in its spectrum. No two elements have identical spectra, and so far as is known no two elements show the same line in common, with the possible exception of hydrogen and helium.

Besides the elements, certain compounds also emit spectra composed of lines. Thus, the blue inner cone of a Bunsen burner shows the spectrum of carbon monoxide, and there are several groups of lines in the spectrum from the carbon arc which are believed to originate in cyanogen gas, produced by the action of atmospheric nitrogen on the carbon poles. The lines in the spectra of compounds are arranged in groups of more or less regular order, technically known as *bands*.

54 Spectral series.—The fact that a single element emits light of several wavelengths shows that the atom is capable of vibrating in several different frequencies, just as a stretched string or column of air can execute the vibrations that produce sound waves in several definite frequencies. A string can vibrate not only with its fundamental frequency, which we may call N, giving a sound of wavelength L, but also with a frequency twice as high, 2N (the octave), giving a wavelength L/2, a frequency 3N, giving wavelength L/3, and so on indefinitely. Therefore we might represent the whole series of sound wavelengths emitted by the string with the single formula

$$\lambda = \frac{L}{n}$$

where L is a certain constant for the string, and n may have any integral value from 1 to ∞. When $n = 1$, λ is the wavelength of the fundamental, when $n = 2$, λ is the wavelength of the first overtone, etc.

A most natural question is the following: Are not also the different wavelengths of light, given out by such an element as sodium or hydrogen, related to one another in some simple numerical way, so that a single formula will represent all of them if different integral values are given to one of the symbols? This cannot be answered by an unqualified *yes* or *no*, but we can say that in *some* of the elements (particularly the metallic ones of small atomic weight) *some* of the lines can be represented in this way, though not by so simple a formula as applies to the acoustical vibrations of a string. The simplest case is that of hydrogen. Figure 62 is a photograph of the visible, and part of the ultraviolet, regions in the spectrum of this gas. A careful examination shows that the lines may be classified into two groups: First, there is a large number of lines without any apparent regularity of arrangement whatever. Second, there are several lines, marked α, β, γ, etc., on the photograph, which show regularity in two respects. The one of greatest wavelength, α, is the strongest line in the whole spectrum, and each succeeding line of the group is weaker than the one before it. Moreover, when the wavelengths

Figure 62

of these lines are measured, it is found that in passing through the series, from α through β, γ, δ, etc., they come closer and closer in wavelength, like a mathematical series approaching a limit. Balmer showed that the wavelengths of the whole group, including lines too far in the ultraviolet to show in this photograph, can be represented with a considerable degree of accuracy by the single formula

$$\lambda = .00003646 \frac{n^2}{n^2 - 4}$$

where n may take any integral value from 3 on. If $n = 3$, we get the wavelength of the α line,—if $n = 4$, we get β,—and so on*. This whole group of lines, represented by a single formula, constitutes a spectral *series*. Our knowledge of series has been greatly increased by the work of Rydberg, Kayser and Runge, and many other investigators. The hydrogen spectrum shows another series in the far ultraviolet region, and still another in the infrared. Other elements also show series in their spectra, although a slightly more complicated formula is necessary to represent them. A comprehensive account of series in spectra is found in French in the "Rapports Presentés au Congres International de Physique," 1900. A very good resume in English is contained in pages 559 to 621 of Baly's "Spectroscopy," second edition. In this chapter we shall not take up the attempts to explain the peculiar form of

*The measured wavelengths of the first four of these lines, in centimeters, are .00006563, .00004862, 00004341, .00004102.

the series relation, since the subject comes more appropriately after the introduction of the electromagnetic theory of light.

55. Continuous spectra.—Spectral lines tend to become widened when the density of the radiating gas or vapor is increased. The hydrogen lines produced when a spark is passed between platinum points in the gas at atmospheric pressure are much broader and hazier than when it is in the rarefied condition of the vacuum-tube. If a large amount of sodium is present in the carbon arc, the two strong lines in the yellow, instead of being fine and sharp, become very broad, and can easily be made to run together and extend some distance beyond their original positions on both sides, causing them to have the appearance of a single broad yellow band with hazy edges. This of course means that the wavelengths emitted are no longer confined even approximately to two definite numerical values, but extend over a relatively wide range. Several causes contribute to this effect of increased density, one of which is probably the fact that any given atom is hindered, by the very close proximity of other atoms, in its natural free vibrations. At any rate, when an exceedingly dense gas, or a solid or liquid body, becomes luminous, the widening of the characteristic lines is so extreme that all possible wavelengths are emitted within the range of the visible spectrum and beyond, and all appearance of definite lines is lost. The spectrum is then said to be *continuous,* since it extends throughout a very wide range of wavelengths without a break in continuity anywhere. In contradistinction from this, the kind of spectrum that we have found to be given out by rare gases, metallic vapors, etc., is called a *bright-line spectrum.* As an example, the hot carbon pole of an arc light gives a continuous spectrum, but the bridge of vapor between the two poles gives a bright-line spectrum.

The following application of the principles of the spectroscope to astronomical problems is interesting and instructive. A nebula is a celestial object which appears in the telescope as a cloud of gas, but the possibility exists that it may really be a swarm of stars, so close together and so far from us that the telescope is incapable of resolving them into discrete bodies. The spectroscope, however, shows that some nebulæ have a continuous spectrum, others a bright-line spectrum, so that with the aid of this instrument it is easy to pick out those that are gaseous.

56. Dark-line spectra.—The sun and most of the stars show still a third type of spectrum, which may be said to be an exact reversal of the bright-line type. While the latter is an assemblage of scattered bright lines, in colors appropriate to the spectral region in which they fall, against a blank—that is a black—background, the former is an assemblage of fine black lines against an otherwise continuous colored background. It is therefore called a *dark-line* spectrum. It may be described as a continuous spectrum with certain definite wavelengths missing.

The following experiment explains the cause of these dark lines in the solar spectrum. An arc lamp, a Bunsen burner, a converging lens, and a spectroscope are set up in line, so that the lens forms an image of the bright carbon pole on the slit of the spectroscope, and the light passes through the flame of the burner before reaching the slit. Some of the light passes into the collimator, and of course produces a continuous spectrum. A little common salt is inserted in the flame, and immediately the sodium lines appear, not bright, however, as they would be in the absence of the light from the arc, but as apparently *dark* lines against the bright continuous spectrum of the arc. The sodium vapor absorbs from the light passing through it those particular wavelengths which it is capable of emitting, and absorbs more than it emits, thus making the lines appear black against the brighter arc spectrum, though in reality they are not absolutely black. Undoubtedly, the dark lines of the sun's spectrum are produced by absorption in the same way. The main bulk of the sun is believed to be an exceedingly dense gaseous mixture, as viscous as a liquid, and like a hot liquid it gives a strictly continuous spectrum. But surrounding this dense luminous portion is an envelope of cooler and rarer vapor containing many chemical elements. These absorb from the light passing through them just those wavelengths which they can emit. When the light reaches the earth it is therefore deprived of these particular wavelengths, and the spectroscope shows black lines at the corresponding positions in the spectrum. By comparing the positions of the black lines with the positions of the bright lines emitted by various terrestrial elements, it has been possible to identify on the sun most of the elements known to us on the earth. The moon and the planets, since they send us only light which they receive from the sun, have the same spectrum. Most of the fixed stars have spectra of the same character as the sun's, and in some cases only a careful examination can distinguish them from the latter. Thus the spectroscope proves that the fixed stars are bodies of the same general physical condition as our sun, in spite of great differences in size, mass, and temperature; and it also shows that the chemical elements present in the earth are distributed throughout the universe and probably make up the major part of its material, a fact which could hardly have been proved by any other means.

57. Absorption by solids and liquids.—Solids and liquids also produce absorption of light, as we have already seen in the chapter on color. If a colored liquid, such as a solution of copper sulphate, potassium permanganate, or chlorophyll, is placed in front of the slit of a spectroscope, and the light from a source that would of itself alone give a continuous spectrum is passed through it, certain parts of the spectrum are absorbed in whole or in part. The black, or darkened, regions which then appear in the spectrum are not fine and sharp, like those

Figure 63

produced by sodium vapor in a Bunsen flame, but broad and hazy. They are called absorption bands. Figure 63 (A) is a photograph of the absorption spectrum of an alcoholic solution of chlorophyll, the green coloring matter of plants. Figure 63 (B) is a photograph of the complete spectrum, with the chlorophyll absent, to serve as a comparison.

58. Continuous spectrum of an absolutely black body.—The strictly continuous spectra given out by hot solids and liquids differ from one another only as regards the distribution of the energy in different wavelengths. Thus, the spectrum from the pole of a carbon arc is not only brighter throughout than that from the filament of a tungsten incandescent lamp,—it also has a greater proportion of its energy in the shorter wavelengths than that from the same body when cooler; but differences in material and surface condition have also great effect. The only kind of continuous spectrum that can be theoretically studied is that from what is called an *absolutely black* body, which is defined as one which absorbs all light, of whatsoever wavelength, that falls upon it, reflecting none. Strictly speaking no objects are absolutely black, though such materials as lampblack and platinum-black nearly fulfill the definition. But it can be shown by theory that the inside of an enclosure, the walls of which are kept at a uniform and constant temperature, acts exactly like a theoretical absolutely black body; and it is possible to make use of this fact for experimental purposes by making a small hole in the wall of such an enclosure, through which light from the interior can pass out and enter the slit of a spectroscope. Figure 64 is a series of graphs, drawn for different temperatures of the radiating enclosure, of the results of measurements of such black-body continuous spectra. For each point on a curve, the abscissa represents the wavelength, the ordinate the corresponding energy. It will be noticed that for higher temperatures

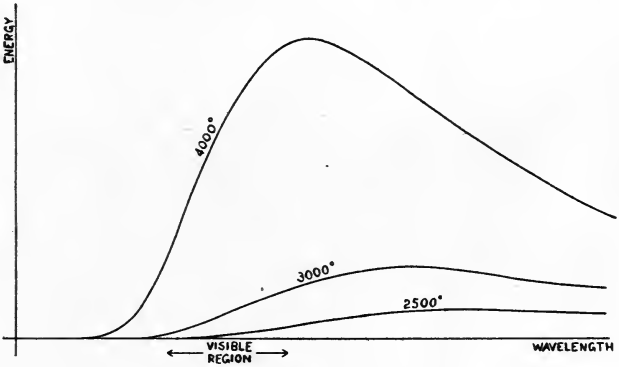

Figure 64

the energy is greater throughout, but particularly in the shorter wavelengths.

59. Planck's theory of "quanta".—The relation between the energy of black body radiation and the wavelength, for a given temperature, as represented by the curves of figure 64, has been the subject of many exhaustive theoretical investigations. It is extremely difficult to give a complete theoretical explanation of the exact form of these curves. Indeed, the man who made most progress toward this end, Max Planck, came to the conclusion that an explanation is impossible unless we make a very remarkable hypothesis in regard to the behavior of a radiating atom, which amounts to this—that although an atom can *absorb* energy steadily and continuously, it can *radiate* only if and when it has acquired by absorption a certain definite *quantum* of energy, or an integral number of times that quantum; and when it does radiate it radiates away all the energy it contains. Just as by adopting the atomic theory of matter we abandon the ancient notion that matter is continuous, so Planck's hypothesis would lead to the conclusion that energy also, so far as the radiation of it is concerned, is composed of discrete amounts. For an atom can radiate one quantum, or two, or three, etc., but not one and a fraction.

This hypothesis, which has been named the "quantum theory," is so very different from our previous notions about energy that, in spite of Planck's success in deriving a formula for black body radiation which fits the experimental curves, it is doubtful if it could obtain much support were it not that a number of other phenomena, including such diverse things as the variation of specific heats with temperature, X-rays, and the explanation of spectral series, are made much more understandable by means of the same hypothesis. Planck offers no explanation of why an atom should radiate in such a manner, and the whole question of the quantum theory is one of the puzzles of modern physics.

The size of the quantum is not the same for all wavelengths, but is directly proportional to the frequency, or inversely to wavelength. That is, for any frequency v the smallest unit of energy radiated is

$$hv$$

The multiplier h is an absolute constant, whose numerical value, in the c. g. s. system of units is

$$h = 6.55 \times 10^{-27}$$

It is commonly known as "Planck's constant."

60. The plane grating.—So far, we have learned no way of measuring wavelengths except by simple interference experiments, such as that of Fresnel with the two mirrors, as described in section 22. That such a method is not capable of much accuracy can be seen from the following considerations. Referring again to figure 20, section 22, it will be recalled that we found that certain points C, M_1, M_1', M_2, M_2', etc., are very bright, and the points midway between dark. The determination of wavelength is made by measuring the interval between two successive bright spots, and also the distance between the two sources S_1 and S_2, and their distance from the plane of the screen. Not only is the distance S_1S_2 difficult to measure, but also the distance between the bands, CM_1, M_1M_2, M_2M_3, etc., cannot be measured accurately, because the bright points are not sharply defined but shade off gradually into darkness.

The difficulty might be illustrated graphically by plotting abscissas to the right of the line AB in figure 20, the length of each abscissa representing the intensity of the illumination at the corresponding point on the screen. The graph that would be obtained would be like figure 65 (X). Evidently the location of the places of maximum brilliancy is subject to considerable error, which would be much lessened if, instead of these broad maxima, we had sharp and clearcut bright lines separated by broad dark spaces, as indicated in figure 65 (Y). Two other decided advantages would also accrue: first, the maxima would be much brighter, since the light would be confined to a very narrow instead of a broad band,—second, if more than one wavelength were present, the maxima due to the different colors would be much less likely to overlap.

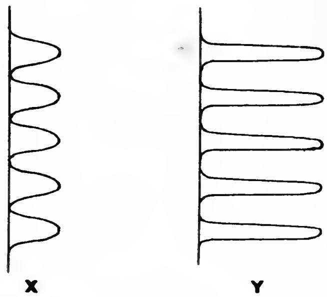

Figure 65

The desired result, making the maxima narrow, sharp and bright, can be secured by getting interference *from more than two points at once.* For instance, if Fresnel's mirror experiment could be arranged so that there were three regularly spaced apparent sources of light, instead of only the two, S_1 and S_2, the maxima would be sharper and brighter,—if there were four, they would be still sharper,—and so on; but one of the best devices for the purpose is what is called a *grating.*

In its theoretically simplest form, a grating is an opaque plate containing a large number of slits, parallel and spaced close together at equal intervals. Light from a narrow source, like a distant slit, or a star, falls upon it and passes through the many narrow slits, producing interference bands on the other side. Let AB, figure 66, represent a section of such a

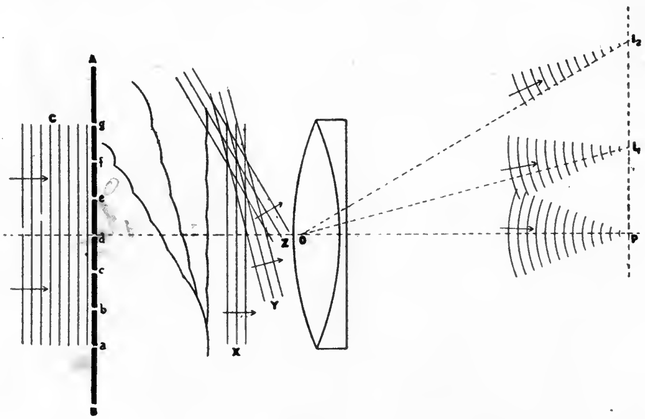

Figure 66

grating. The slits, shown in section at a, b, c, etc., are supposed to run perpendicular to the plane of the paper. We shall suppose that monochromatic plane waves are falling perpendicularly upon it, the advancing wavefronts being indicated at C. The slits, or transparent portions of the grating, may be regarded as centers from which new wavelets start out. As these get farther from the grating, where their curvature is less, they tend to combine into several different sets of plane wavefronts, moving in different directions. For instance, the 12th wave out from a, together with the 12th from each of the other openings, tends to form a plane wave, parallel to the incident waves C. There will be a continuous train of such waves, a few of which are indicated farther out at X. These, after passing through the lens, will be brought to the principal focus P, and will form a bright spot there, exactly as if the grating were removed and the original wavefronts C fell directly upon the lens. An entirely different set of wavefronts will be formed by a combination of the 12th wavelet out from a, the 11th from b, the 10th from c, and so on, the resulting wavefront being inclined at a certain angle to the original wavefronts. A few of the wavefronts formed in this manner are shown at Y. After traversing the lens, they will be brought to focus at a point I_1 in the principal focal plane. This point will of course be the intersection of the plane with the undeviated ray for this set of waves, which is a line drawn through the optical center O perpendicular to the wavefronts Y. Still a third set of wavefronts will be formed by a combination of the 12th wavelet from a, the 10th from b, the 8th from c, etc. These, a few of which are drawn at Z, are still more inclined to the original wavefronts, and are brought to focus farther out in the focal plane, at such a point as I_2.

It is not difficult to prove that if θ_1 represents the angle between the Y and the X wavefronts (which is the same as the angle I_1OP) and θ_2 the angle between the Z and X wavefronts (the angle I_2OP), λ the wavelength of the light, and σ the distance between the centers of slits in the grating,

$$\sin.\ \theta_1 = \frac{\lambda}{\sigma} \qquad\qquad \sin.\ \theta_2 = \frac{2\lambda}{\sigma}$$

Of course, the bright points on one side of P would be duplicated by corresponding bright points on the other side, since everything is symmetrical about the axis of the figure. It is also possible that there may be other bright points, for which the sine of the angle is $3\lambda/\sigma$, $4\lambda/\sigma$, etc. There is, however, a limit to the number of bright points obtainable; for the sine of an angle cannot be greater than 1, and if, for instance, $4\lambda < \sigma < 5\lambda$, there will be four bright points on each side of P, but not 5.

The formation of the wavefronts by monochromatic light in passing through the grating can be very nicely illustrated by the following experiment with ripples in a basin of mercury.

To one prong of a tuning-fork is fastened a piece of sheet-metal cut like a comb with some 16 teeth spaced about 3/16 inch apart, as shown in figure 67. The fork is then placed so that the teeth just dip below the surface of the mercury in the basin, and is set into vibration. Each tooth becomes a center for a series of ripples which emanate from the teeth just as secondary light wavelets emanate from the openings in the grating of figure 66. The only difference is that the teeth are the actual sources of the ripples, while the openings in the

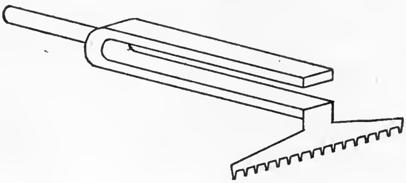

Figure 67

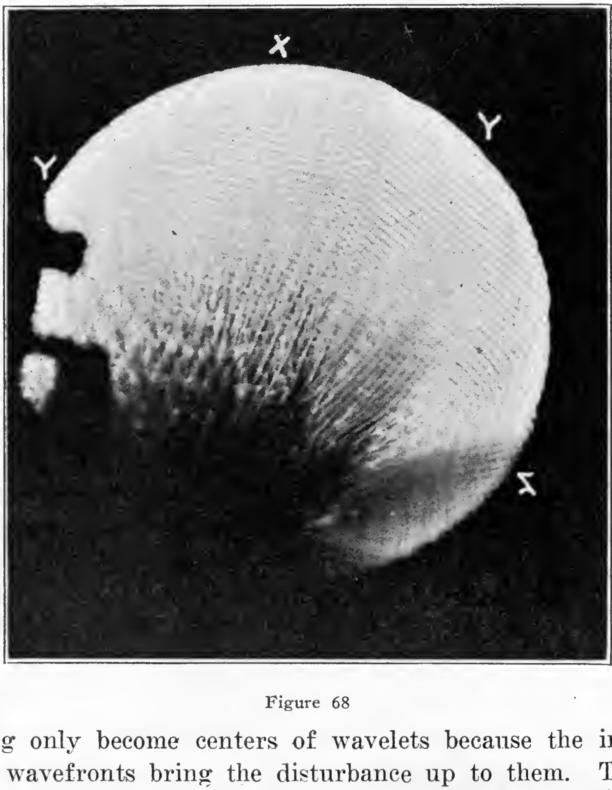

Figure 68

grating only become centers of wavelets because the incident plane wavefronts bring the disturbance up to them. The appearance of the mercury surface becomes like figure 68, which is an instantaneous photograph. The heavy dark place of irregular shape is the shadow of part of the tuning-fork. Close to the comb, the vibrations are too complicated to be analyzed, but farther away five separate sets of plane wavefronts can be clearly seen. The central one, X, advancing perpendicular to the comb, corresponds to the wavefronts X of figure 66. The two marked Y correspond to the Y wavefronts of figure 66 and a symmetrical set on the other side of the axis, and similarly for those marked Z. If the wavelength of the ripples had been somewhat less, or the space between teeth greater, other sets of wavefronts might have been seen. The comb was so constructed that $2\lambda<\sigma<3\lambda$, so that only two spectra could be expected on each side of the central axis.

It was not practicable in the case of the ripples to arrange any device for focussing, and one may say that figure 68 corresponds to figure 66 with the lens removed.

Now, suppose that the incident light contains, beside the wavelength already considered, λ, another, longer or shorter, which we shall call λ'. The grating would produce several sets of wavefronts for this wavelength also. The central set would be parallel to the incident wavefronts C, and would therefore be brought by the lens to the principal focus P, like those for wavelength λ. The other sets for λ', however, would not be brought to the same points as the corresponding sets for λ, for we have seen that the angles θ_1, θ_2, etc., depend upon the wavelength. If $\lambda'>\lambda$, each bright point for λ' will be focussed farther from P than the corresponding point for λ, and conversely if $\lambda'<\lambda$. Therefore, however many wavelengths may be present in the incident light, a series of spectra will be formed on each side of the central spot P. These are known as the first spectrum, second spectrum, etc., to the right or left, as the case may be. The central spot P, since it contains in itself all wavelengths of the incident light, is sometimes spoken of as the "spectrum of order zero." This manner of speaking is consistent with the general formula for the grating,

$$\sin.\theta=\frac{n\lambda}{\sigma}$$

For the first spectrum, $n = 1$, for the second $n = 2$, and for the spectrum of order zero $n=0$, so that $\theta=0$ for all wavelengths.

In order to measure wavelengths with the grating, we must first know σ, the distance between the centers of the slits, sometimes known as the "grating-space." This can be found by placing the grating under a high-power microscope and measuring with a micrometer. Then we must measure one of the angles θ_1 or θ_2, etc., for the wavelength desired. If the lens of figure 66 is the objective of a telescope, and if, as in the figure, the axis of the telescope is perpendicular to the plane of the grating, the angle may be found by means of a micrometer. But when, as is usually the case, the grating is mounted on the table of a spectrometer, so that the telescope may be swung about an axis through the grating perpendicular to the plane of the diagram, then it can be turned so as to receive upon the cross-hair at the principal focus any wavelength in any order of spectrum, so that the angle θ can be read off from the verniers attached to the telescope.

61. Why the lines are sharp.—Of course, these measurements would still be very inaccurate, and lines, or maxima, due

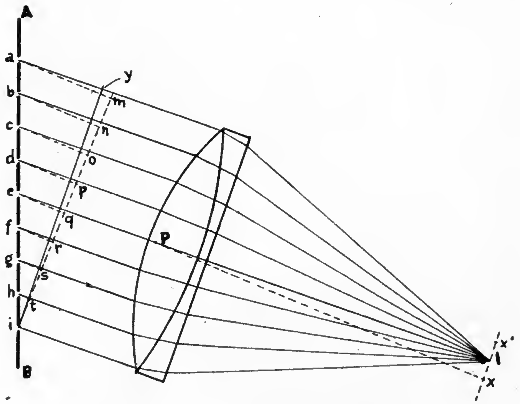

Figure 69

to different wavelengths would seriously overlap, unless these maxima were quite sharp, as indicated in figure 65 (Y). It remains to show why the large number of slits in the grating produces the desired sharpness.

Consider figure 69. AB represents a grating, which for the sake of definiteness and simplicity, we shall suppose has 9 slits or openings, although a practical grating usually has many thousand. We suppose there is only a single wavelength present in the light. The only wavefront drawn, iy, is one for the first spectrum. Rays are drawn from each slit, running through the lens to the point I, where they are brought to focus. The lens is turned in such manner that these rays are parallel to the lens-axis, so that I is really the principal focus. From what we have learned about lenses, it is evident that there is the same number of wavelengths in each ray, measured from the wavefront iy to the point I, and brightness occurs at I because, measured from the slits to this wavefront, each ray has one more wavelength than the one next below it in the figure, so that all arrive at I in the same phase. Now let im represent a hypothetical wavefront, slightly inclined to iy, so that the perpendicular distance am contains 9 wavelengths, whereas ay contains only 8. Such a wavefront would be brought to focus at a point x, very close to I because im is so slightly inclined to iy. We shall show that practically no light reaches the point x from the 9 slits because the rays that might come there mutually interfere, or annul one another. The proof is as follows: Between the wavefront im and x, there will be the same number of wavelengths in every ray, otherwise x would not be the proper focus for a wavefront in the position im. Hence we need consider only those portions of the nine rays that lie between im and the grating. There are 9 wavelengths in am, $9 \times 7/8$, or $7\,7/8$ in bn, $9 \times 6/8 = 6\,3/4$ in co, $9 \times 5/8 = 5\,5/8$ in dp, $9 \times 4/8 = 4\,1/2$ in eq, $9 \times 3/8 = 3\,3/8$ in fr, $9 \times 2/8 = 2\,1/4$ in gs, $9 \times 1/8 = 1\,1/8$ in ht. Then am is just $4\,1/2$ wavelengths longer than eq, and therefore the rays from a and e will interfere and annul one another, so far as effect at x is concerned. Similarly, and for the same reason, the ray from b will annul that from f, the ray from c that from g, and the ray from d that from n. There remains only the ray from i, not annulled by any other ray, to produce illumination at x. In a later chapter, (see section 91), it will be shown that in such a case the illumination is proportional to the square of the number of effective elements; therefore, since 9 slits conspire in phase to produce illumination at I, and only 1 at x, the illumination at the latter point will be only 1/81 that at the former.

Now, suppose that, instead of 9 slits, the grating contained 999. Then ay would contain 998 wavelengths, and we should draw im so that am contained 999. The point x would then come very close indeed to I, and the brightness of illumination at x would be only $1/(999)^2$ that at I, because only one opening in the grating would contribute to it, all the others annulling one another's effects in pairs. We should then be fully justified in saying that x is practically a dark point, and it is quite easy to prove, by similar reasoning, that there is another point x′, close to the other side of I, which is equally dark. Since there is a dark spot so very close on each side of the bright spot I, the maximum of illumination must be very sharp. Moreover since, if the grating-space σ is quite small, the maxima of two successive orders, such as I_1 and I_2 of figure 66 are quite far apart, we have just the conditions for accurate measurements, viz., sharp and widely separated maxima of brightness.* In our proof, we have taken only the case of a grating having an odd number of openings, but the result would be the same in all essentials if the number were even. In fact, the dark points would then be absolutely dark, instead of only relatively so.

In actual gratings there are sometimes as many as 120000 openings, and the grating-space σ has been made as small as 1/20000 inch, though 1/15000 inch is more common. It is of course obvious that actual slits could not be cut so close together as this, through a solid plate. Gratings are made by ruling upon a plate of glass an immense number of fine parallel scratches with a diamond. One may regard the scratches as being opaque, the clear unscratched spaces between taking the place of the slits. Strictly speaking, this is incorrect, for the scratched grooves are really not opaque. But the diminished thickness of the glass where the grooves are cut causes the light passing through these places to be out of phase with that which passes through the clear spaces, and this has the same effect, so far as the location of the spectra is concerned, as if the grooves were really opaque.

*A more thorough treatment shows that between the sharp bright maxima of different orders there are a great many faint secondary maxima, but with practical gratings these are too feeble to be observed.

62. Reflection gratings.—The best gratings are ruled, not on glass, but on a polished plate of metal. In such cases, it is the reflected light that produces the spectra. A study of figure 66 will show that if the spaces a, b, c, etc., reflected light, while the spaces between did not, spectra would be formed in exactly the same way, but on the opposite side of the grating. The great difficulty about a grating is to make the rulings, or "slits," perfectly uniform in spacing. This involves a great many practical difficulties which it is impossible to overcome entirely. Still, a good grating gives beautifully sharp and clear spectral lines, making it much superior to a prism for producing spectra, except in one particular,—since a grating produces many spectra, no one of them contains more than a fraction of the available light, and therefore grating spectra are in general not so bright as those produced by prisms.

63. The concave grating.—Henry Augustus Rowland, who developed the manufacture and use of gratings for the measurement of wavelengths and other spectroscopic investigations, conceived the idea of ruling gratings upon concave reflecting surfaces, in order to avoid the use of lenses to focus the light. This is of particular importance in the regions of very short wavelength, for glass absorbs much of the ultraviolet light, and it is difficult to make satisfactory lenses out of materials which are free from this objectionable quality. The idea proved very valuable, and the so-called "concave grating" is a most useful instrument, simple and convenient in manipulation. For the theory of the concave grating, the student should consult some more advanced text, such as Preston's "Theory of Light," or Baly's "Spectroscopy," where the following relations are proved:

Suppose that, in a plane through the middle of the grating and perpendicular to the rulings, a circle be drawn whose diameter is the radius of curvature of the concave surface, so that the circle is tangent to that surface at one end of the diameter, and passes through the center of curvature at the other end. Then, if, the slit be placed anywhere on this circle, the different spectra will be sharply focussed at various points about the same circle. Moving the slit will of course cause the spectra to shift their position, but so long as the former remains on the circle, the latter will also. Rowland adopted a very simple and ingenious method for making sure that the slit remains on this focal circle, as shown by diagram in figure 70. Two beams, AS and BS, are set up making an angle of

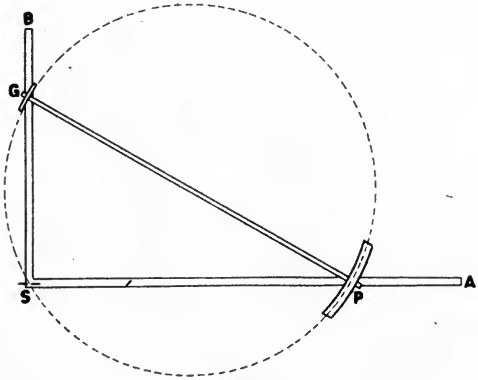

Figure 70

90°. The slit is placed at the intersection of these two beams, i. e., at S. The grating G, and a plateholder P, to hold a long narrow photographic plate, are mounted at opposite ends of a third beam, whose length is equal to the radius of curvature of the grating. Thus, the center of curvature of the grating comes just at the middle of the photographic plate. The ends of the beam GP are mounted on slides or carriages, so that the end G can slide along the beam BS while the end P slides along the beam AS. This beam therefore always forms the hypothenuse of a right-angled triangle, the acute angles of which can be changed. From the geometrical proposition that an angle inscribed in a semi-circle is a right-angle, it follows at once that S always remains on a circle of which GP is a diameter. Of course, the photographic plate P is bent to fit this circle.

64. The ultraviolet region. Fluorescence. Phosphorescence. Photography.—It was stated in section 22 that there are waves shorter than the violet, and others longer than the red. The former are called "ultraviolet," the latter "infrared."

Although ultraviolet light does not itself produce vision, it can, by aid of the phenomenon called "fluorescence," give rise to visible light and thus make its presence known. Fluorescence is a property of a great many substances, including the dyes fluorescein and uranin. It is the power to absorb certain wavelengths and re-emit the energy in longer wavelengths, instead of turning it into heat. Fluorescein and uranin, for instance, absorb ultraviolet light and re-emit the energy as yellowish-green light. It is very easy to show the existence of waves beyond the violet in the spectrum of the sun or of an arc-light, by holding in that position a thin-walled flask containing a solution of one of these dyes, or a glass plate which has been colored with one of them.

There are some substances which continue to give off light long after the incident light has been cut off. This phenomenon is called "phosphorescence," because it suggests the glow shown by a piece of phosphorus that has been rubbed, in the dark. The latter glow, however, is not true phosphorescence, because it is caused by slow oxidation, rather than by previously absorbed light.

The principal method for studying the ultraviolet region is by photography, for these waves are particularly effective on a photographic plate. In order to go very far into the ultraviolet, however, it is necessary to avoid the use of glass for prisms and lenses, since it absorbs very strongly all but the longer ultraviolet waves. Either quartz or fluorite are used for this purpose, these substances being transparent for much shorter waves than glass, but for the shortest waves gratings and mirrors must be used. Even a little air absorbs such waves, and Schumann and Lyman, who got wavelengths as short as .00001cm., were obliged to work in a vacuum or a hydrogen atmosphere. Very recently, Millikan and Sawyer have found waves as short as .00000272cm. from a spark in a high vacuum.

65. The infrared region.—Photography has also been applied to the study of the shorter infrared waves, for plates can be specially prepared suitable for this purpose. But the usual method, and the only method suitable for the very long waves, is to detect them by the heat produced when they are absorbed. The infrared waves are sometimes erroneously called "heat-waves," because the major part of the energy radiated from the sun or any hot solid is composed of waves longer than the red, and they are therefore responsible for most of the heat produced when the light is absorbed. But it is conceivable that some source of light might radiate in such a way that most of the energy was in the ultraviolet, in which case it would be the very short waves that were responsible for most of the heating. In fact, there is no such thing as a heat-wave, at least in the sense in which we understand a wave when speaking of light, sound, etc. For heat consists of entirely irregular molecular motions, which have nothing in common with waves in the ether except that they possess energy. Light energy does not become heat energy till the light has been absorbed and the waves as such have ceased to exist.

The distribution of waves in the infrared, or for that matter also in the visible and ultraviolet, might theoretically be mapped out by moving a thermometer with a small blackened bulb through the whole focal plane of the spectrum, and noting the therometer reading at each point, but a mercury or alcohol thermometer is far too insensitive for such a purpose. The devices that have been most commonly used are the bolometer and the thermopile, but any other very sensitive temperature-indicator might be used in which the object whose temperature is measured is in the form of a strip, narrow enough to cover only a small section of the spectrum at a time, and blackened so as to absorb and turn into heat whatever radiant energy falls upon it.

66. The bolometer.—In the bolometer, this strip is a thin and narrow piece of blackened platinum, which is connected up with three other conductors so as to form a Wheatstone bridge, to which a battery and a very sensitive galvanometer are also connected in the usual way. The bridge is balanced so that there is no deflection of the galvanometer when no radiation is falling upon the strip. If the strip is placed anywhere in the spectrum so that energy falls upon it, it is heated, its resistance is thereby changed, the bridge becomes unbalanced, and a deflection of the galvanometer results. The magnitude of the deflection indicates the intensity of the radiation of wavelength corresponding to the point in the spectrum where the platinum strip is at the moment.

67. The thermopile.—The thermopile depends upon the principle that in a metallic circuit consisting of two dissimilar metals, an electric current will flow if one junction be at a different temperature from the other. Usually, at least for such purposes as are now under discussion, the circuit is compounded of many pieces of two different and suitable metals, for instance a piece of antimony, then one of bismuth, another of antimony, another of bismuth, etc., till there may be a dozen or more such junctions in the whole circuit. A sensitive galvanometer is also included, to detect and measure the current. Every alternate junction is brought into a line, close to one another, and these junctions are exposed to the radiation while the others are shielded from it. The exposed junctions in line are of course blackened so that they will absorb and not reflect the radiation.

With the infrared, as with the ultraviolet, it is necessary to avoid the use of glass. Usually a prism of rocksalt is employed, and concave mirrors are substituted for the lenses of the ordinary spectrometer, to focus the spectrum and to collimate the light.

68. The Doppler principle. Motion of the stars.—An interesting application of the spectroscope is to the determination of the rate at which stars are approaching or receding from the solar system. This is based upon a principle known as the "Doppler effect," which applies not only to light, but also to sound or any other wave phenomenon. If a source of waves is approaching an observer, the length of the waves which he receives is less, if it is receding from him greater, than if there is no motion.

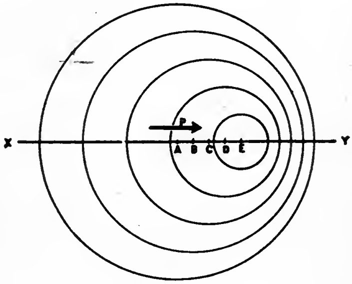

Figure 71

The reason is easily explained by figure 71. Suppose that the source of waves is moving in the direction of the arrow P, with a velocity v, and let V be the velocity of the waves themselves. Bear in mind that V depends only upon the properties of the wave-carrying medium, not at all upon the velocity of the body emitting the waves. If that body sends out a crest when it is at A, then that crest will expand into a growing circle with A as center, despite the fact that the source moves away from A in the meanwhile. Let A, B, C, etc., be positions of the source at instants differing by a period, so that a crest is started from each of these points. Then at any later instant the wavefronts will be circles, as shown, but not concentric circles, the center of each being displaced toward the right from the center of the preceding one by the distance the source moves during one period. If λ represents the normal wavelength, as it would be if the source were at rest, the period is λ/V, and the motion of the source during this time is $v\lambda/V$. Evidently, the wavelength received by an observer to the left, in the direction of X, will be *increased* by this amount, while that received by an observer to the right, in the direction Y, will be *decreased* by the same amount. That is, the observed wavelength, to a stationary observer in the line of motion, will be

$$\lambda' = \lambda \pm \frac{v\lambda}{V} = \lambda \frac{V \pm v}{V} \qquad (1)$$

the upper or lower sign being used according as the source is receding or approaching.

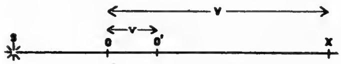

Figure 72

It may be, however, that the observer instead of the source is moving. In such a case, the analysis must be somewhat different, as will be shown by use of figure 72. Let S be the position of the stationary source, and O the initial position of the observer. Suppose the latter is moving away from the source, toward the right, and let O′ be his position one second later, so that $OO' = v$. Lay off also the distance OX equal to V. Within this last distance there are V/λ waves, and the observer would have received all these waves in one second had he remained stationary at O. But since he has, during the second, moved to O′, he actually receives a number less than this, for the waves lying between O and O′ have failed to reach him. The number he actually receives per second is therefore $(V - v)/\lambda$, and this is the *frequency* of the waves as he receives them; whereas the frequency with which they are emitted from the source is V/λ. This case differs from the preceding in that the wavelength is not actually changed, but it seems to the observer to be changed; for since his spectroscope or other device for measuring wavelengths is moving along with him, it acts as if the waves had the velocity V with respect to the observer together with a frequency $(V-v)/\lambda$. Since wavelength equals volocity divided by frequency, the spectroscope therefore indicates the wavelength to be

$$V \div \frac{V-v}{\lambda} = \frac{V\lambda}{V-v}$$

If the observer is *approaching* the source, instead of receding from it, it is easy to prove in a precisely analogous way, that the wavelength appears to be $V\lambda/(V+v)$, or in general the apparent wavelength is

$$\lambda' = \frac{V\lambda}{V \pm v} \qquad (2)$$

the lower sign being taken if the observer is receding, the upper if he is approaching.

The two formulas, (1) and (2), are not quite alike, and the student is usually inclined to think that they should be: for he argues that motion is only relative and it should make no difference which of the two things, source or observer, is at rest and which in motion. But there is a third thing to take into account, namely the medium that carries the waves. It must make a difference, for instance, in the case of sound, whether an observer is moving through stationary air toward a stationary sounding body, or the sounding body is moving through stationary air toward a stationary observer. Just so, it must make a difference in the analogous case of light, *provided that we may regard the ether as a material medium which has a definite location in space.* Certain considerations bearing upon this proviso will be brought out later (section 85).

For all applications to spectra, formulas (1) and (2), though not identical, yield results so nearly identical that there is no distinguishing between them. If we divide the numerator by the denominator in the second member of (2) we get

$$\lambda' = \lambda\left(1 \pm \frac{v}{V} + \frac{v^2}{V^2} \pm \frac{v^3}{V^3} + \text{etc.}\right)$$

which differs from (1) only by the terms in the square and higher powers of v/V. The greatest velocity found in a star is of the order of 300 kilometers per second, that of light is 300000 kilometers per second, and therefore v/V is of the order 1/1000. Since this makes v^2/V^2 of the order 1/1000000 and the other terms still smaller, they are beyond the limits of measurement and can just as well be dropped out.

If a star then is moving toward us, or we toward it, all the wavelengths in its light are apparently shortened, that is all the lines in its spectrum are shifted toward the violet end, and if the star is moving away from us, or we from it, the lines are shifted toward the red. In either case, the amount of the shift is proportional to the relative velocity or rather to that component of it which is along the line joining star and observer. It is possible, from the great array of lines shown in the spectra of many stars to identify them with certain elements in spite of their displaced position, and thus to know their proper wavelengths. To determine the velocity then, it is only necessary to photograph the star's spectrum side by side with the spectrum of some terrestrial source such as that of a hydrogen vacuum tube or a spark between iron terminals, and measure the apparent wavelengths by comparison. Obviously, one cannot say whether a certain measured shift of the lines is to be attributed to motion of the star or of the earth, or partly to each, except that we know the earth to be moving with a tolerably great velocity in its orbit about the sun. The sun itself may also be moving, but it is customary to treat all such measurements of line-of-sight velocities as if the sun were at rest. The measured velocity is corrected for the earth's orbital motion, and the result stated as the velocity of the star "with respect to the sun."

It is senseless to ask if the sun has any "absolute" motion, for the motion of one body cannot be intelligently thought of except with reference to some other body, but one cannot help speculating as to whether the sun is moving or at rest with respect to the *ether,* though such a question presupposes that the ether is to be regarded as having a definite location in space. In section 84 a certain experiment will be discussed which bears upon this question.

Problems.

1. Suppose the slit of a spectroscope to be 5mm. long, the focal length of the collimator to be 30cm., and that of the telescope to be 40cm. How wide will the spectrum, in the focal plane of the telescope, be?

2. A dense flint prism, of angle 60°, has the following indices of refraction: 1.7774 for wavelength .00004713, 1.7695 for λ .00005016, 1.7537 for λ .00005896, 1.7444 for λ .00006708. Plot a curve with λ as abscissa, index as ordinate. How long would be the stretch of spectrum, between the longest and the shortest of the above wavelengths, if the focal length of the telescope were 30cm.? (It will be sufficient, though not strictly correct, to consider that all the light passes through the prism at minimum deviation, and use the formula of paragraph 26).

3. Calculate the wavelengths of the first four lines of the Balmer formula for hydrogen, and compare with the experimentally found values given in the foot-note to paragraph 54.

4. What would be the spectrum of light reflected from a white wall, in daylight? Under illumination by tungsten-filament lamps?

5. Some of the dark lines in the solar spectrum are due to absorption in the earth's atmosphere. How can these be distinguished from the true solar lines?

6. From the curves of figure 64, show why a heated solid first becomes dull red, only becoming "white-hot" at a very high temperature. Also show why incandescent lamps are more efficient when operated at high temperatures.

7. Calculate the energy in a "quantum," for light of wavelength .00003 cm., and of wavelength .00008 cm.

8. What wavelengths in the first, second, and fourth order spectra come at the same place as λ .00005000 cm. in the third order, with a grating.

9. Show that, in a plane grating, if the incident plane waves strike with an angle of incidence i instead of zero, the formula becomes $\sin. i \pm \sin. \theta = n\lambda/\sigma$.

10. If a certain star is moving away from the earth at a speed of 100 kilometers per second, find the change in wavelength of a line whose proper wavelength is .000065 cm. (red). Also for one whose proper wavelength is .000040 cm. (violet).

CHAPTER VIII.

69. The approximately rectilinear propagation of light.—70. Shadow of a straight edge.—71. Shadow of a wire.—72. Diffraction through a rectangular opening.—73. Resolving-power.

69. The approximately rectilinear propagation of light.— We have already seen, in section 18, that light does not travel absolutely in straight lines; that is, it does not cast absolutely sharp shadows, even when the light-source is as narrow as we can make it. In fact, we should rather expect that light, like other forms of wave motion, would bend rather freely about an obstacle in its path, and our first interest lies in explaining, not why it bends at all, but why it does not bend more.

The statement that light travels approximately in straight lines really amounts to this, that a comparatively small object,

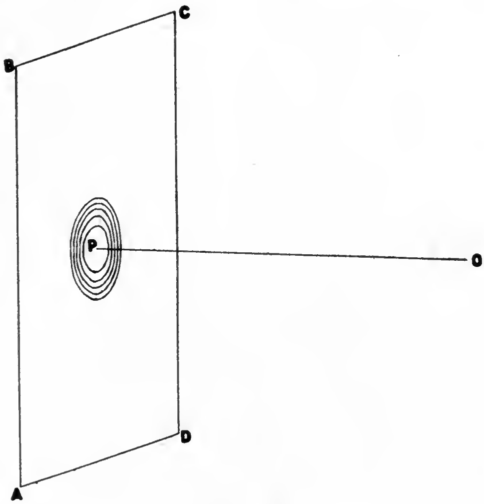

Figure 73

even when held some distance from the eye, will screen off all, or at least most, of the light coming from a point. For instance, a ten-cent piece six feet from the eye, will pretty effectually blot out the light of a star, although the same coin at one hundred feet will not. In this respect, light forms a marked contrast with sound, for even an obstacle several feet in diameter, when six feet away, produces very little effect on the intensity of a sound. Our first problem, then, will be to explain why the very much shorter wavelengths of light cause this great difference. For simplicity, we shall assume that we are to deal with only plane waves, and only monochromatic light, that is, light all of one wavelength. The area ABCD figure 73 represents in perspective a portion of a plane wavefront advancing from the left toward the point O. According to Huyghen's principle, the effect at O may be regarded as made up of the summation of all effects produced by secondary wavelets, one coming from each point in the wavefront. Some of these secondary wavelets will annul one another upon reaching O, but there will be a net residual effect, which we shall prove is the same as if all but a small portion of the wavefront were annihilated. Let P be the foot of a perpendicular drawn from O to the wavefront, and r the distance OP. Imagine a number of spheres drawn with O as center, and with radii equal to $r + \frac{\lambda}{2}$, $r + \lambda$, $r + \frac{3\lambda}{2}$, $r + 2\lambda$, $r + \frac{5\lambda}{2}$, etc., increasing by $\lambda/2$ each time. Each sphere will intersect the wavefront in a circle with P as center, and the whole wavefront will be divided up into a large number of regions known as "Huyghens' zones," the innermost one a circle, the others rings. We shall number them 1, 2, 3, etc., beginning at the center. The most important characteristic of these zones is this, that the average distance from O to the points in any one zone is half a wavelength shorter than that for the next zone beyond. If we consider the effect of any one zone, at the point O, it is evident that points on the inner boundary, being one-half wavelength nearer than those on the outer boundary, will tend to neutralize the latter's effects; but points lying nearer the middle line of the ring will not completely neutralize one another, so that there will be a certain net effect at O due to the whole zone. Each zone, however, tends to neutralize the effect of the next zone within it. Therefore we can represent the effects of the odd zones by positive quantities, d_1, d_3, d_5, etc., and those of the even ones by negative quantities, $-d_2$, $-d_4$, etc., and the effect of the whole system of zones, that is, the whole wavefront, will be properly represented by a series,

$$D = d_1 - d_2 + d_3 - d_4 + d_5 - d_6 + \text{etc.}$$

which will have an infinite number of terms if the wavefront is not limited in extent.

We shall now prove that d_1 is slightly greater than d_2, d_2 than d_3, etc., so that the series consists of alternate positive and negative terms of decreasing magnitude, and is therefore convergent and has a definite value.

The effect of any zone depends upon three things, its area, its distance from O, and the inclination at which its light comes to O. Leaving aside for a moment the influence of the inclination, the effect should be proportional to the area of the zone divided by its distance from O. If R represents the radius of the outer boundary of the first zone, we have from simple geometry $R^2 + r^2 = (r + \lambda/2)^2$, giving

$$R^2 = r\lambda + \lambda^2/4$$

therefore the area of the first zone is πR^2 or $\pi(r\lambda + \lambda^2/4)$. It is easy to show in a similar way that the square of the radius of the outer boundary of the second zone is $2r\lambda + \lambda^2$, and the area of the corresponding circle $\pi(2r\lambda + \lambda^2)$. The area of the second zone is the difference between the areas of these circles, or $\pi(r\lambda + 3\lambda^2/4)$. Similarly, the area of the third zone is $\pi(r\lambda + 5\lambda^2/4)$, that of the fourth $\pi(r\lambda + 7\lambda^2/4)$, etc.

The average distance of points in the first zone from O is $r + \lambda/4$, that of points in the second zone $r + 3\lambda/4$, etc., so that if we divide the area of each zone by its average distance from O we get in each case the quotient $\pi\lambda$. Therefore, the effect of zone distance and zone area, in causing a variation in the terms of the series, mutually counterbalance one another, so that the only thing to consider is the inclination. Since the light from the outer zones comes to the point O at a greater inclination with the line PO than does that from the inner zones, it follows that the magnitude of the terms d_1, d_2, etc., continuously decreases with higher order, and the series is convergent. The difference between successive terms, however, is quite small, particularly in the case of the first few terms.

The easiest way to get an idea of the value of the whole series, representing the effect of the whole wavefront upon the point O, is by a graphical method. Lay off on a straight line, as in figure 74, the distance oa equal to d_1. From a lay off in the opposite direction the distance ab, equal to d_2. Then $ob = d_1 - d_2$. Lay off bc equal to d_3, cd equal to d_4, and so on. Then

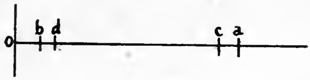

Figure 74

$oc = d_1 - d_2 + d_3$, $od = d_1 - d_2 + d_3 - d_4$, and so on. It is evident, that, since each term is smaller than the one preceding it, the whole series will have a value less than d_1, and if we make an assumption about the manner in which the terms decrease which is certainly reasonable, it can be shown that the sum of an infinite number of terms will be just equal to half of d_1. The physical meaning of this mathematical result is that the effect of the whole wavefront is only half that which the first zone alone would exert, so that if an opaque screen were placed in the path of the light, with a hole in it only large enough to let the first zone through, the illumination at O would be actually increased. This surprising prediction is actually verified by experiment, but it must be remembered that it is only at the point O, and in its immediate neighborhood, that the illumination is increased. Other points in the plane of O become darker. Furthermore, since the size of a zone depends not only upon the wavelength, but also upon the distance r, a hole big enough to let through only the first zone as calculated for the point O would let through the first and second as calculated for some point on OP nearer to P, the first, second and third for a point still nearer, etc. Therefore, at a certain point the illumination would be $d_1 - d_2$, or nearly zero, while at another it would be $d_1 - d_2 + d_3$, nearly equal d_1 or d_3, and so on. Therefore, different points along the line OP should be alternately bright and dark. This statement also is found to be true.

The problem with which we began was to explain why a comparatively small obstacle will cut off light but not sound. Suppose then that instead of an opaque screen with a small hole, we place in the path of the light a small opaque disc, which will cut out some of the central zones but allow all the rest to pass on. For numerical calculation, take the diameter of the disc to be 1 cm., and the distance from the point whose brightness we are considering to be 200 cm. Let n be the number of zones cut out. The square of the radius of the nth circle is $nr\lambda + n^2\lambda^2/4$, and this must be equal to the square of .5 while r is 200 and λ may be taken as .00005, about the brightest part of the spectrum. Then,

$$(.5)^2 = .25 = .01n + .000000000625n^2$$

The coefficient of n^2 is so small that we may neglect it, since we want only the approximate value of n, which is

$$n = .25/.01 = 25$$

Since the first 25 zones are eliminated, the illumination at O will be given by the series

$$D = d_{26} - d_{27} + d_{28} - d_{29} + \text{etc.}$$

whose value is approximately equal to half the first term, that is $d_{26}/2$. Now although the values of the d's decrease rather slowly, still the 26th term is very much smaller than the first, consequently the illumination at O is very faint, though not absolutely zero.

As a comparison, let us calculate how large a disc would be required to screen off 25 zones from a sound wavefront, 200 cm. away, the wavelength being taken as 64.45 cm., which corresponds to a sound of 512 vibrations per second. Letting x represent the required radius of the disc, and using the formula

$$x^2 = nr\lambda + n^2\lambda^2/4$$

we get x = 985 cm., that is, the disc must be nearly 20 meters in diameter. The example shows how the great difference in wavelengths between light and sound accounts for the difference in the effectiveness of a small obstacle.

70. Shadow of a straight edge.—We are now in a position to explain the bright and dark bands that appear near the shadow of an obstacle of considerable width with a sharp straight edge, which were mentioned in section 18. Figure 75 is the same as figure 16, with the addition of the wavy line from X to Z, which is a graph representing the distribution of illumination as it appears on the screen AB. The peak p indicates that at the point P on the screen the illumination is particularly bright, that is, there is a bright band running along parallel to the edge e, at a distance PE from what would be the edge of the shadow if light did travel absolutely in straight lines. On the other hand, the drop in the curve to the right of C indicates that C is relatively dark, and a dark band

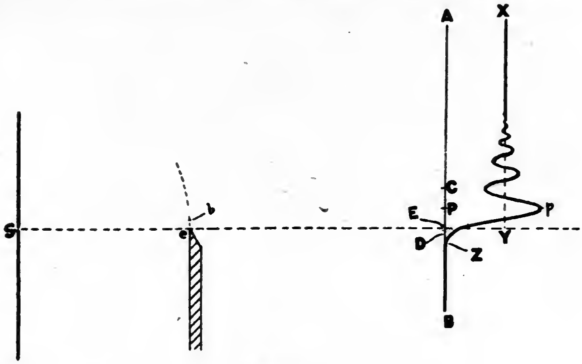

Figure 75

runs along the screen there, also parallel to the edge. The other peaks and depressions in the curve show that there are a series of bright and dark bands, becoming less and less pronounced at greater distances from the edge, until, at some point beyond X, all traces of bands are lost, and the screen appears uniformly illuminated, just as if the obstacle at e were removed.

Consider first the illumination at the point E, just on the straight line through the slit S and the edge e. If we take the plane of the wavefront just as it reaches the edge e, and construct on it the Huyghens' zones for the point E, the center of the zones will lie at e, and half of each zone will be cut off by the obstacle. Therefore, the effect at E, instead of being half that due to the first complete zone, will be only ¼ that of the first complete zone, that is, E will be distinctly darker than if the obstacle were not present.

For a point further up on the screen, such as P, the center of the zone system would be at such a place as b. Let us suppose that P, and therefore b, are high enough so that the whole of the first zone is uncovered, but that the obstacle cuts off a segment of each of the others. Then the effect at P is greater than it would be with the obstacle removed. For we have seen that the effect of all the zones from the second on, when they are complete, is to nullify half the effect of the first zone. In this case the first zone is complete and the others incomplete, and therefore these are incapable of nullifying half the effect of the first. Thus the first bright band at P is explained.

At a still higher point, as C, the first two zones are clear, the rest partly covered. Here the second zone, being less opposed than it would normally be by the zones of higher order, is more free to oppose the action of the first, therefore the illumination at C is less than if the obstacle were not in its place, and the first dark band is explained.

The second bright band occurs where three zones are completely free, the second dark one where four are free, etc. Evidently these bands become weaker and weaker as more and more zones are uncovered, since the zones of higher order are weaker than those of lower.

Now take a point D, below E. For this point, the center of the zone system is hidden by the obstacle, and no zone would be complete. In this case it is better to construct the zones on an entirely new plan. Instead of taking a line from D perpendicular to the wavefront as the basic distance for drawing the spherical surfaces that cut out the zones, we take De, the shortest distance to that part of the wavefront that is not hidden, and draw our spheres of radius De $+ \lambda/2$, De $+ \lambda$, De $+ 3\lambda/2$, etc. All of the resulting zones will be incomplete segments of rings, and the resulting illumination at D will of course be faint. The farther down D is moved into the shadow, the fainter it will be, but there will be no maxima or minima of illumination. The brightness fades out continuously and rather rapidly, becoming inappreciable at some such point as Z. From there on, no illumination can be seen in the shadow.

The formation of bands at the edge of a shadow, together with a number of similar phenomena, are classed under the general name of *diffraction.* Bands like those of figure 75 are always formed about the edges of shadows, whenever the source of light is small enough, or when, if not small, it is far enough distant to subtend a very small angle, as in the case of a street light fifty or one hundred yards from the opaque obstacle forming the shadow. The source may be a long line of light, or a slit, if it is parallel to the edge of the obstacle, in which case of course the bands will be brighter.

71. Shadow of a wire.—One of the most interesting cases of diffraction occurs when the obstacle is a narrow straight object, such as a wire. If a slit is used as source, it must be parallel to the wire, and the shadow may be cast on a whitened wall or a sheet of paper. Bands are seen on the outside of the shadow, like those of figure 75, and in addition there are always one or more bright bands running through the center of the shadow, formed by interference between the light bending into the shadow from the two sides.

72. Diffraction through a rectangular opening.—A very important case of diffraction, having much to do with the effectiveness of telescopes, microscopes, spectroscopes, and other optical instruments, is the kind produced when light passes through a limited opening. We shall take up only a case in which the conditions are somewhat simplified, so as to make the theoretical discussion easy.

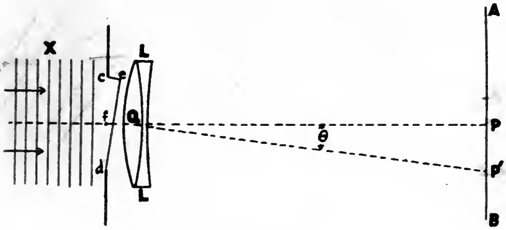

Figure 76

In figure 76, X represents a system of plane waves advancing toward the lens LL. The waves may come from a star or a distant point of light, or they may come from a slit placed at the principal focus of another lens, not shown in the diagram. The lens LL tends to concentrate the light at its principal focus P, on the screen AB. cd is an opening, like a moderately wide slit, in an opaque screen, so that the only light that reaches the lens is in a beam, as wide as cd in the plane of the paper, and as long as may be desired in a direction at right angles to the paper. That is, the opening is in the form of a slit with straight sides, whose width cd is not great. It is found that the light is not all brought to the point P, but—if the source is also a slit—there is a bright band at P, parallel to the slit-source, flanked on each side by a number of fainter bands. If the aperture cd is widened, the central bright band becomes brighter and *narrower,* and the fainter bands become much more closely spaced. If, on the other hand, the aperture is *narrowed,* the central bright band at P becomes *widened* and the whole system of bands widens out and becomes fainter. The explanation is as follows:

Since the principal focus P is not the only point that receives light from the section of wavefront cd, we shall select a point at random, P′, in the focal plane, and find what illumination comes to it. Let θ represent the angle between the axis of the lens and the line from the optical center to P′. P′ will be a point of absolute darkness only if all the secondary wavelets, coming from every point in the section of wavefront cd, neutralize one another. Draw de perpendicular to QP′ and ce parallel to it. Let f be the center of the opening cd. From our previous study of lenses, we know that a wavefront de would be focussed at P′, showing that from the line de on to P′ there are the same number of wavelengths in every ray. Then whatever differences in path exist in the secondary wavelets are accounted for in the space between cd and de. If ce is just a wavelength, the point P′ will be *dark,* exactly contrary to what one would at first thought suspect. For we must remember that not only the points c and d, but *every point between* c and d sends a ray toward P′. For each point in the half df, there is a corresponding point in the half fc whose ray has a *half-wavelength* farther to travel in getting to P′. Therefore one half of the wavefront cd neutralizes the effect of the other half, and P′ is dark. If ce is two wavelengths, P′ will again be dark, for cd may then be divided up into four equal parts. The upper two of these will annul one another, each point in one being $\lambda/2$ farther from P′ than a corresponding point in the other, and for the same reason the lower two will annul one another. Similarly, if ce is equal to any integral number of wavelengths, except zero, the point P′ will be dark. If ce $= 0$, P′ will of course coincide with P, all the secondary wavelets will have the same distance to travel, and the point will be the brightest possible. On the other hand, if ce is equal to an odd number of half wavelengths, complete darkness at P′ is impossible. Suppose, for instance, that it is $3\lambda/2$.

Then cd could be divided into thirds. Each point in the uppermost third would annul the effect of a point in the middle third, being one-half wavelength farther from P′, and thus the two upper thirds would cancel one another. The lowermost third, however, would have a residual effect at P′ and would therefore cause a small illumination there, which accounts for the first maximum of brightness on the lower side of the principal focus. Another maximum would occur at such a position of P′ that ce $= 5\lambda/2$, etc. Since the angle cde $= \theta$, and ce $=$ cd. sin. θ, we may express what we have found in the form of equations as follows: Complete darkness occurs at such angles θ that

$$\sin.\ \theta = n\lambda/cd$$

where n is any integer except 0. The greatest brightness occurs when

$$\sin.\ \theta = 0$$

but secondary maxima of brightness occur when

$$\sin.\ \theta = n\lambda/2cd$$

where n is any odd integer except 1. (When ce $= \lambda/2$ the point is not dark, but neither is it a point of maximum brightness.) From the symmetry of the figure, it is plain that the points of maximum brightness or of darkness on the lower side of the principal focus would be duplicated by similar points, symmetrically placed, on the upper side. The graph of figure 77 shows the distribution of brightness on the screen. The symbols to the right of the vertical line indicate the values of ce at the corresponding points on the screen, while to the left are plotted the corresponding brightness.

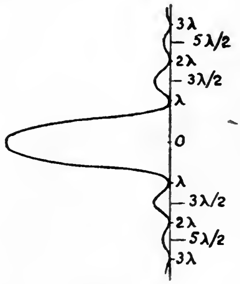

Figure 77

If the screen is removed, and an eyepiece placed in position to focus on the focal plane of the lens LL, so that the latter with the eyepiece form a telescope, then the bright and dark bands are seen in the eyepiece just as they would be on the screen, with whatever magnification the eyepiece produces.

If the aperture cd is quite wide, the bands are very fine and sharp, so that a high magnification is necessary to see them.

73. Resolving-power.—The importance of the considerations in the preceding section can be understood from the following example. Suppose that in figure 76, instead of a single set of plane wavefronts X, we have *two* sets, nearly but not quite parallel to one another, say with a small angle a between them. For instance, the light might be sodium light which has passed through a prism or grating, so that the two distinct wavelengths present in this light will have their wavefronts slightly inclined to one another, the rays of course also being inclined to one another at the same angle. Even if there is no diaphragm such as cd to limit the two beams, still the limited size of the prism itself puts a limit on the width of the beams, and the result will be the same as if there were an opening just large enough to pass them. Therefore each beam will produce in the focal plane of the lens LL a diffraction pattern like that of figure 77. These two patterns will not fall in exactly the same place. In fact their centers will subtend, from the optical center of the lens, an angle a, the same as that between the two wavefronts before they reached the lens. Instead of two sharp spectrum *lines,* then, we have two bands of a certain width, each flanked by a number of fainter bands, and this would be true even if the slit of the spectroscope were infinitely narrow and each beam of light absolutely monochromatic. The wider the beam of light, the more nearly these bands would approximate to real lines. If the angle a is quite small, and the width of the beams is also small, the two central bands will run together, so that only a single band will be seen, no matter how much we try to magnify the image with a high-power eyepiece. The question then arises, how small may be the angle between the wavefronts, a, in order that the two *lines* may be *resolved,* that is, seen as distinct. It is difficult to say exactly when the two images run together, but for purposes of comparison between different instruments we say that the limit of resolution is reached when the central maximum for one image comes just where the first minimum of the other image occurs. We have seen in the preceding section that the angle between the center of a diffraction pattern and the first minimum is given by $\sin.\theta = \lambda/cd$, or, since θ is so small that $\sin.\theta = \theta$ very approximately, we may write $\theta = \lambda/cd$. Therefore, for resolution of the spectrum lines we must have

$$a \geqq \lambda/cd$$

The lines will therefore be resolved only when the angle between the wavefronts is at least as great as an angle whose arc is equal to the wavelength, the radius being equal to the width of the beam. In order that a spectroscope may have high *resolving-power,* it is therefore not sufficient that the prism or grating shall produce a great dispersion of the different wavelengths. It is also necessary that prism or grating shall be wide enough to transmit a wide beam of light, and the lenses must also be wide enough not to cut this beam at the corners.

The same sort of problem comes up in the use of telescopes, though there is a slight difference due to the fact that the beam of light has then a round cross-section. It is evident that when a telescope is pointed toward a star everything is symmetrical about the axis, and therefore the diffraction pattern, instead of being a central band flanked by parallel fainter bands, will be a central *disc,* surrounded by a system of faint *rings.* It corresponds very closely indeed to what would result if figure 77 were revolved about its axis of symmetry, except that, because the opening is round and not rectangular, the actual diameters of the dark rings are somewhat enlarged. The radius of the first dark ring subtends from the optical center of the objective an angle whose value is $1.22\lambda/cd$ instead of λ/cd, if cd represents the width of the beam of light, that is the diameter of the objective. If the telescope be pointed toward a close double-star, each star will form in the focal plane an image of itself which consists of a central disc surrounded by faint rings, and we say that the images are just resolvable when the center of one disc falls on the first dark ring of the other diffraction pattern. A double-star is resolvable then if the two components make an angle which is equal to or greater than 1.22λ divided by the diameter of the objective.

A telescope of large diameter is said to have a large resolving-power. Similar considerations apply to a microscope, though the treatment is somewhat different because with that instrument the light does not enter the objective in plane waves but in a highly diverging beam. All optical images are affected by diffraction, and this explains the statement made in section 38 that no lens, however perfect in design and workmanship, can produce a point image from a point source.

The question is often asked, by those unacquainted with optical instruments,—"How small may an object be, or how far away may it be, and yet be visible in a given microscope or telescope?" Such a question is not pertinent, because any microscope will render visible any object, however small, and any telescope will reveal any object however small or remote, provided that object gives out a sufficiently strong beam of light. The proper question is,—"How close together (in linear measure for a microscope, in angular measure for a telescope) may two small objects be and yet be seen as distinct and separate in the instrument?" The efficiency of a telescope in this respect depends entirely upon the diameter of its objective lens, not at all upon its focal length, consequently the diameter is by far the more important dimension.

The eye, like any other optical instrument, has a definite limit to its resolving-power, depending upon the diameter of the pupil. Anyone can convince himself of the truth of this statement by the following simple experiment: Make two clear dots upon a blackboard, about ¼ inch apart, and then walk backward away from the wall with the eyes fixed upon the dots. At a certain distance, the two seem to run together, and can no longer be distinguished as separate dots. Strange as it may seem, it is possible that they may be distinguished at a greater distance in a relatively dark than in a very bright room, for in very bright light the pupil contracts, and the resolving-power is decreased.

The diffraction-rings produced in the eye are finer in structure than the structure of the retina itself, if the pupil is normally open, and therefore are not ordinarily visible, but they become visible if an effective decrease in the diameter of the pupil is made artificially by looking through a very small hole in a card or other opaque object. The following experiment is instructive: Make a single straight cut with a knife in a stiff card, so as to make a sort of slit. Hold this close to the eye and look through it at a single filament of an incandescent lamp, keeping the slit parallel to the filament. The diffraction bands described in connection with figures 76 and 77 will be clearly seen. It will be noticed that the bands become narrowed if the card is sprung so that the opening is widened, and conversely if the opening is narrowed. Here the lens of the eye takes the place of the lens LL in figure 76, the opening in the card that of the opening cd, and the retina that of the screen AB. Notice that the central maximum is 50% farther from either of the secondary maxima on either side, than any two adjacent secondary maxima are from one another.

Problems.

1. A certain prism of dense flint glass separates the two yellow sodium wavelengths by an angle of 2 seconds of arc. How wide must the beam be, in order that the two lines may be seen separated in the spectroscope?

2. A "double-star" has components whose angular separation is only 0.16 second of arc. What is the diameter of the objective of the smallest telescope that can resolve them?

3. Find the resolving-power of the eye, when the pupil is $\frac{1}{8}$ inch in diameter.

CHAPTER IX.

74. Young's interference experiment.—The arrangement of mirrors, invented by Fresnel to show the interference of light, and described in section 22 (see figure 21), is not the only device for the purpose, though perhaps the most satisfactory. The earliest was due to Thomas Young, one of the early champions of the wave theory. It consists of a slit through which light streams to an opaque screen containing two very small holes close together. A white screen farther on receives the light from the two holes, which could be replaced with advantage by two fine slits parallel to the original slit.

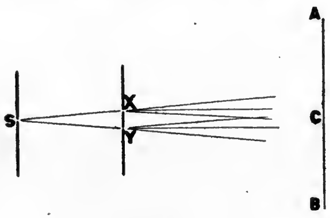

Figure 78

Figure 78 shows this simple arrangement. S is the source slit, X and Y the two round holes or small slits. These last are small enough so that diffraction causes the light coming through them to spread out. Light from the two holes will therefore overlap in the neighborhood of C on the white screen AB, and will produce interference fringes there. The student will observe that the opaque screen with two slits through it is to all intents and purposes a very coarsely ruled grating with only two openings. Although this arrangement is easy to construct, very little light comes through the two slits, and the fringes are therefore very dim indeed.

75. The biprism.—This is another device of Fresnel. It consists of two identical glass prisms, of very small angle, cemented together base to base. Figure 79 shows how it is set up. B is the biprism, S the slit-source, ab a white screen. The light passing through the upper half of the biprism is bent downward, that through the lower half upward, so that the light comes to the screen as if it came from two points X and Y, which may be regarded as virtual images of the real slit S, formed by refraction through the upper and lower halves of the biprism respectively. Where the two beams overlap near the center of the white screen, interference fringes are produced.

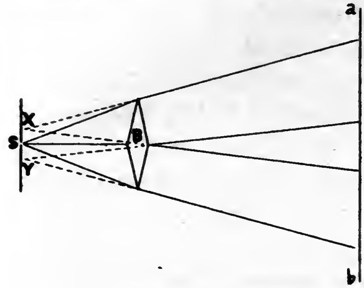

Figure 79

76. Interference in thin uniform films.—Nature herself provides us with the finest cases of interference, in the beautiful iridescent coloring of soap-films, films of oil on water, fissures in the interior of crystals, etc. In all such cases, it is found that there are two reflecting surfaces separated by a very small distance, and the interference is between the light reflected from the first surface and that reflected from the second. It is further necessary, if the colors are to be seen anywhere but in certain particular spots when the eye happens to be in the correct position, that the light shall come from a widely extended source, such as the sky, the whitened wall of

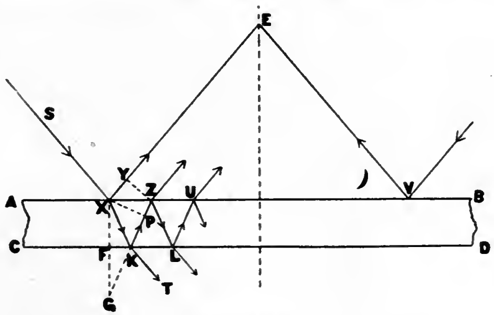

Figure 80

a room, or a broad flame, giving out light in all directions from every point of it.

We shall first suppose that the film has plane and parallel surfaces, so that it is of uniform thickness. Let AB and CD, figure 80, represent the upper and lower surfaces of such a film, and E the position of the eye. Let t be the thickness of the film, and n its index of refraction. At first we shall suppose that the extended radiating surface, not shown in the diagram, but well above the position E, gives light of only a single wavelength λ. If only the upper surface of the film reflected light, the eye could look in any direction, such as EX, and see part of the extended source reflected in this surface, the course of a ray of light being given by the lines SX and XE. But only part of the light is reflected at the first surface, while the rest enters the film along the ray XK, and part of it is reflected at the second surface along the ray KZ. When the ray KZ strikes the upper surface, part of it is again reflected, but part is refracted out and follows a ray parallel to the ray first reflected, XE. There will in fact be still another ray which will emerge parallel to XE after three reflections, another after five, and so on, each one, however, weaker than the rays that have undergone fewer reflections. Since the film must be very thin to show interference, all these rays will lie very close together, and may therefore enter the pupil of the eye together. If the eye is focussed for infinitely distant objects, that is for parallel rays, they will all be brought to focus at the same point on the retina.

Let us first see what difference in phase exists, at the retina, between light in the ray XE and that in the parallel ray from Z, which has suffered only one reflection within the film. Draw YZ perpendicular to the two rays. From Y and Z on, there are the same number of wavelengths in each path, since a wavefront such as YZ would be focussed upon the retina. Also draw XP perpendicular to KZ. A wavefront in the position XP would be reflected to the position YZ on emerging into the upper medium. Therefore there are the same number of wavelengths in XY and ZP, and the path-difference between the two rays is simply the distance XK + KP, in the medium of refractive index n. Draw XF perpendicular to the surfaces AB and CD, and prolong it till it meets ZK produced at G. The right-angled triangles XFK and GFK are equal, therefore GK = XK, and the path-difference XK + KP = GP. The angle at G is the angle of refraction, r, and XG = 2t. Therefore the path-difference is

2t. cos r. Since the wavelength within the film is λ/n, the number of wavelengths in the path-difference is

$$2t.\cos r \div \frac{\lambda}{n} = \frac{2nt.\cos r}{\lambda}$$

77. Change of phase on reflection.—But difference in path is not the only difference in the two rays SXY and SXKZ, for they have suffered reflections of quite different kinds. While the reflection at X occurs on the rarer side of the surface AB, that at K occurs on the denser side of the surface CD, and it is an important fact that a reflection on the rarer side of a boundary is always accompanied by an abrupt change of phase which is equivalent to the introduction of half a wavelength in the path. That is, a crest is reflected as a trough, and vice versa. No such change of phase exists when the reflection occurs on the denser side of the boundary. Crest is reflected as crest and trough as trough. An actual proof of these statements involves difficult mathematics, but their reasonableness may be made plain by a consideration of the much simpler case of waves running along a string. In figure 81 XY represents a light string, joined at Y to the end of a much heavier string YZ. Since transverse waves

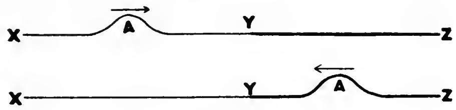

Figure 81

travel faster along a light than along a heavy string, the light string takes the place of the air in figure 80, the heavy string that of the film, and the point Y represents one of the bounding surfaces. In the upper diagram of figure 81 a single crest A is shown advancing toward the boundary point Y. Suppose for a moment that the end Y were clamped tightly. Then, when the crest A struck it, the reaction at this fixed point would throw the cord below the normal position, and a reflected trough would start back toward the left. With the end Y not fixed rigidly, but merely fastened to the heavier cord, a trough is still reflected to the left along the lighter cord, while a crest goes on along the heavier one. There is a certain analogy here with the case of a light ball striking squarely a heavier one which was originally stationary, both being perfectly elastic.

The direction of motion of the striking ball is reversed, and part of its motion is given up to the other.

If the ball originally in motion is the heavier, its direction of motion is not reversed upon striking the lighter one, but only somewhat checked, both balls moving off in the same direction after impact. The analogue to this case is provided when a crest approaches the point Y along the heavier string, as indicated by A in the lower diagram of figure 81. Both the reflected and the continuing waves will then be crests.

Of course, in either case, the reflected and the continuing waves will both have amplitudes less than that of the incident wave. In fact the energy of the incident wave is divided between the two.

Let us apply these principles to the optical case of figure 80. The incident wave coming along from S, like that along the lighter string in the upper diagram of figure 81, will give rise to a reflected wave of reversed phase travelling toward E and a continuing (refracted) wave without reversal of phase travelling toward K. These two will divide between them the energy of the original wave. On the other hand, the wave coming from X to K, like that in the heavier string in the lower diagram of figure 81, will produce a reflected wave travelling toward Z and a continuing (refracted) wave travelling toward T, neither of which has its phase reversed.

Return now to consideration of the difference in phase between the ray reflected at the upper surface of the film, and the one which has undergone one reflection inside the film. We have found that the difference in path, expressed in wavelengths, is

$$\frac{2nt. \cos r}{\lambda}$$

To this must be added the equivalent of one half wavelength on account of the dissimilarities in the two cases of reflection. Therefore, interference will occur when

$$\frac{2nt. \cos r}{\lambda} + \frac{1}{2} = N + \frac{1}{2}$$

or when

$$\frac{2nt. \cos r}{\lambda} = N$$

N being any whole number.

As a matter of fact, the ray from Z is considerably weaker than that from X, and therefore cannot completely neutralize it. But if we consider the ray from U, which has undergone three internal reflections, we see that its path exceeds that from Z by exactly the amount by which the latter exceeds that from X, viz., in wavelengths,

$$\frac{2nt. \cos r}{\lambda}$$

but that there is no occasion in either of the rays from Z or U for the change of phase on reflection. Consequently the rays from U and Z will differ by a whole number of wavelengths in phase, and therefore will be in condition to assist one another, when the rays from Z and X are opposite in phase, so as to interfere. In the same way the rays that have undergone 5, 7, or any odd number of internal reflections, will also be in phase with that from Z. It can be shown that the sum of the amplitudes of the rays that have undergone 1, 3, 5, 7, etc., internal reflections is just equal to the amplitude of the ray that has had only one external reflection. Consequently, when the thickness t, the angle r, and the wavelength λ are such that

$$\frac{2nt. \cos r}{\lambda} = N$$

a whole number, the eye will see no light whatever in this direction.*

*There are certain considerations of continuity which would lead us to believe, quite apart from the demonstration of figure 81, that a change of phase must be caused by either internal or external reflection, and not by both. As a film becomes thinner and thinner, it should approach, in optical qualities, the condition of no film at all; in other words, a film whose thickness is very small compared to the wavelength of light should fail completely to reflect light, allowing it to pass through unimpeded. In fact, it is easy to test this point, for a soap-film that is allowed to drain and evaporate becomes in some places much thinner than the wavelength of visible light. Such places are known as "black spots," and they have the appearance of irregularly shaped holes through the film, although the fact that the whole film does not collapse shows there is not a real hole. (Incidentally, this experiment indicates that the diameter of a molecule must be much smaller than the wavelength of visible light).

But, if there were no difference in phase introduced by the two kinds of reflection, the only cause of phase difference would be dif-

Now since, the thickness of the film being constant, the condition for interference varies only with the angle of refraction, or—what comes to the same thing—with the angle of incidence, then if the eye observes darkness in any such direction as EX it will also observe darkness in any direction, such as EV, for which the angle of incidence is the same. That is, it will observe a dark *ring,* subtending a cone whose angle is XEV. Since darkness occurs whenever

$$\frac{2nt. \cos r}{\lambda} = N$$

where N may have any one of the integral values, 0, 1, 2, 3, etc., there will be a series of such rings. It has already been intimated that the eye must be focussed for parallel rays, as if looking at an infinitely distant object, in order that these rings may be seen. There is also another reason why they look as if they were very distant: since the appearance of a dark ring depends only upon the angle of incidence, and not at all upon the position of the eye, if the latter be moved the rings will move with it, instead of appearing fixed within the film. In fact, they seem to be seen *through* the film, just as one sees distant clouds or landscape through a window. For these reasons, it is said that the rings of interference seen in thin films of uniform thickness are "located at infinity."

The ring for which N = 1 is called the ring of *first order,* that for which N = 2 the ring of *second order,* etc. Of course, the dark rings are separated by bright rings where the light reflected from the upper and lower surfaces are more or less in phase.

If white light instead of monochromatic falls upon the film, a series of colored rings will be seen, instead of merely dark rings separated by bright rings all of one color. In a real soap-film the colors are of irregular outline instead of in a ring pattern, because the film is never of uniform thickness and the surfaces are usually not plane. Where the angle of

ference in path, and for vanishing thickness of the film all the reflected beams would have the same length of path. Thus, experiment, as well as general reasoning from the principle of continuity, indicate an abrupt change in the phase of one of the reflected rays.

incidence and the thickness are such as to produce interference for light of a certain wavelength, other wavelengths are reflected without interference, and thus the colors are produced.

78. Non-uniform films.—When a film is not of uniform thickness, its two faces will not be parallel, and the theory of the interference becomes much more difficult. One thing about the interference bands produced in such a case, however, is in striking contrast with those for films of uniform thickness, viz., the fact that they are not *located at infinity*, but seem to lie in, or very close to, the film itself. Figure 82 will show a reason for this fact. AB and CD are the two surfaces of a non-uniform film, Sx an incident ray, xE a ray reflected from the upper surface, and zE′ a ray emerging after one reflection inside the film. Under these circumstances, xE and zE′ are not parallel. In order that both, after entering the pupil of the eye, shall be brought to focus at the same point of the retina, the eye must be focussed, not upon infinity, but upon the point where the two rays cross, which is either within the film or close to it. The difference in phase varies with both the angle of incidence and the thickness of the film, but principally with the latter. Consequently a single dark fringe seen on the surface maps out very closely the points of equal thickness.

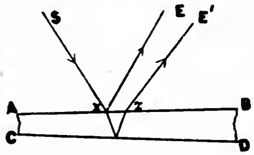

Figure 82

Suppose one sheet of plate glass be laid upon another, and viewed by the reflected light of a sodium burner. A very thin film of air is left between the plates, which is usually quite far from uniform in thickness, because the surface of commercial plate glass is never really plane. Regions of equal thickness of the air film are mapped out by curved bright and dark lines. Suppose that at a certain fringe the thickness of the film is 5λ. Then, along the next fringe on one side it is 5.5λ, along the next on the other side it is 4.5λ. The thickness increases by half a wavelength from fringe to fringe, because the light reflected from the lower surface of the film traverses the film twice. If one of the glass surfaces is known to be plane, the method of fringes in reflected light may be used to test the other for planeness, and to show what parts must be ground down.

79. The Michelson interferometer.—A very valuable instrument, whose action depends upon the same principle as interference in thin films, is the Michelson interferometer, shown diagrammatically in figure 83. AB and CD are plane glass mirrors, heavily silvered on their front surfaces, the former fixed in position, the latter so mounted that it can be moved backward and forward, in a direction perpendicular to its own plane, by a fine-pitched screw. EF is another plane glass mirror, set at an angle of 45°, and fixed in position. On its upper right-hand surface it is covered with a very thin coat of silver, so that it will reflect about half of all light falling upon it and allow about half to pass through, that is, it is a half-silvered mirror. An extended source of monochromatic light, such as a sodium flame, (Bunsen burner fed with common salt) is placed at L. A small part of the light is reflected from the lower unsilvered surface of the mirror EF, but this is weak enough to be ignored. Of that which reaches the half-silvered surface, half passes through to the fixed mirror AB, is there reflected back to EF and is again half reflected, to the eye placed at I. The other part of the light which, coming from L, strikes the half-silvered surface, is reflected to CD and back again, and half of it passes through EF to the eye. The difference in path between these two beams causes interference fringes to be seen. The fourth glass GH is added because without it one of the interfering beams would pass through three thicknesses of glass, the other only through one. GH must be of the same thickness as EF and parallel to it, but it is not silvered.

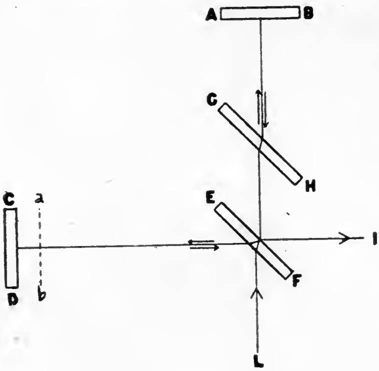

Figure 83

So far as the eye is concerned, we may regard the fixed mirror AB as replaced by *its image,* seen by reflection in EF, which would come in some such plane as ab. If AB be adjusted, by means of screws provided for that purpose, so that its image is accurately parallel to CD, we shall have in effect light reflected from two plane and parallel surfaces, ab and CD, just as in figure 80, with the single exception that when we are dealing with a real film, as in figure 80, multiple reflections occur within the film, causing a series of reflected beams, while here we have only two. The interference fringes will therefore be a system of concentric circular rings, apparently located at an infinite distance.

Any displacement of the movable mirror CD will change the thickness of the hypothetical airfilm between CD and ab, and cause a corresponding change in the rings, either an increase or a decrease in their diameter. In order to determine whether an increase in the distance ab to CD causes the rings to enlarge or diminish, let us fix our attention on a certain dark ring, say of order 200. We have seen that the difference in path, expressed in wavelengths, will be

$$\frac{2nt. \cos r}{\lambda} = N$$

Here, n = 1, because the film is air, and therefore the angles of refraction and incidence are equal. Also, N = 200. Therefore

$$\frac{2t. \cos i}{\lambda} = 200$$

t is of course the distance ab to CD. If t increases, the factor cos i must decrease by the same amount, in order that the ring of order 200 shall still be seen. But a decreasing cosine means an increasing angle, and therefore, for greater thickness of the airfilm, one must look more obliquely in order to see this ring, that is the ring enlarges. All the rings enlarge as CD moves away from ab, and spots appear in the center one after another and expand to form new rings. Conversely, if CD moves toward ab, the rings all decrease, and one after another they shrink to mere spots in the center and disappear.

The appearance or disappearance of each bright spot at the center indicates that another half wavelength has been added to or subtracted from the thickness of the airfilm. For in the center the incidence is normal, that is, i = 0, and the path-difference becomes merely 2t. It is therefore easy to determine accurately how far, in terms of light wavelengths, the mirror CD is moved, by simply counting the number of rings that appear or disappear at the center and dividing by 2. Professor Michelson has used this instrument to find the length of the standard international meter, in terms of the wavelength of a red line in the spectrum of cadmium. The actual details of the experiment were complicated and laborious, but the degree of accuracy obtained was marvelous. A very readable account of this and of other scientific uses of the interferometer will be found in Michelson's book "Light Waves and their Uses."

If the fixed mirror AB is so adjusted that its image ab is not parallel, but slightly inclined to CD, we shall have the case of a non-uniform film, and the bands, instead of being circular, will be nearly straight, mapping out regions of approximately equal thickness. Instead of being located at infinity, they will seem to lie near ab and CD. A movement of CD causes them to move across the faces of the mirrors.

80. Newton's rings.—The interference phenomenon known as "Newton's rings" is a matter of considerable historical interest. As its name indicates, it was known to Sir Isaac Newton. It is clearly an example of interference in an airfilm whose thickness is not uniform, but Newton, refusing to entertain the wave theory, gave a different and somewhat cumbersome explanation which now has no interest for us.

These rings are produced by placing upon a plane piece of glass a plano-convex lens of very slight curvature, convex side in contact with the plane glass. Between the two pieces there is a film of air, whose thickness varies from zero in the middle, at the point of contact, to a great many wavelengths at the edges. Figure 84A is a photograph of this arrangement as it appears when illuminated from the front by monochromatic light. In this photograph the light was of violet color, and was obtained by separating the light from a mercury-arc into its spectrum. The fringes appear somewhat elliptical in the figure, because the photograph was necessarily taken somewhat obliquely. They are in fact circular, marking regions of equal thickness. The center, where the two glasses come into contact, is always black, corresponding to zero difference in path, unless dust or some other obstruction prevents actual contact. Figure 84B is the same thing when illuminated by white light, instead of monochromatic. The difference will be explained later.

81. Fabry and Perot interferometer.—If the student will refer again to figure 80, he will readily understand that not only the *reflected* light, but also that which is *transmitted* through the film, should show interference fringes. For some of the light passes through the film without being reflected, some passes through after two internal reflections, some after four, etc., and these different rays, all parallel on emergence from the film, should be in a condition to interfere for certain angles of incidence. Such fringes are indeed seen in the transmitted light, but they are rather weak, because the ray that passes through without reflections is so much stronger than any of the others that the interference is not complete. Only faint brighter and darker rings on a rather bright background are seen.

Fabry and Perot conceived the idea of diminishing the intensity of the first ray, and increasing that of all the others, by lightly silvering the surfaces of the film. Under these circumstances, the contrast between bright and dark rings becomes much greater, and besides the bright rings become much sharper, something like spectral lines,—that is, the width of the bright part is much less than that of the dark part. On this principle, they have developed a type of interferometer that is known by their names, which for some purposes is superior to that of Michelson. Its essential parts are two discs of perfectly plane glass, parallel to one another, one of them fixed and the other movable by means of a screw in a direction perpendicular to its own plane. The two surfaces turned toward one another are lightly silvered, and the layer of air between is the film in which the interference takes place. Only the transmitted light is used. The thickness of the air-film is altered by turning the screw. This instrument has been much used in recent years for determining wavelengths. It is more accurate than a grating for this purpose. The technique of its use, however, is somewhat involved.

82. Interference in white light.—In any interference experiment, it is found to be impossible to observe interference fringes if white, or composite, light is used, except when the difference in path amounts to only a few wavelengths. In most experiments, matters are so arranged that in part of the field of view the difference in path of the interfering rays is zero or nearly zero, while in other parts it may be many wavelengths; but in some, as with a thin film whose thickness is uniform but amounts everywhere to say more than six or eight wavelengths, the difference in path is always relatively large. In cases of the latter sort, fringes are absolutely invisible with white light, though they may be very clear with monochromatic light. In cases of the former sort, white light shows a few fringes, usually not more than ten or a dozen, where the differ-

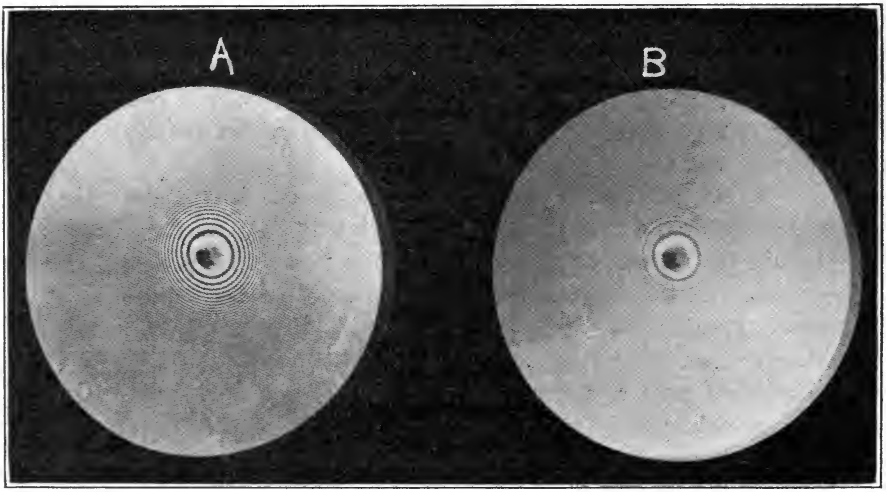

Figure 84

ence in path is quite small, but none whatever in the rest of the field. Consider, for example, Newton's rings. Here the thickness of the air-film ranges from zero, where the lens and the flat disc come into contact, to larger and larger values as we recede from that point. Figure 84B is a photograph of the rings when the light comes from a tungsten filament lamp, and is therefore composite white light. It does not reproduce the visual appearance quite faithfully, because the range of wavelengths for photographic sensitivity is somewhat different from that for visual sensitivity, but the general character is the same. 84A, on the other hand, is taken with monochromatic light, and the difference is striking. In the original photograph, fine clear rings may be counted clear out to the edges of the disc in 84A, while in 84B only a few can be seen. These last of course were brilliantly colored in the original object.

In order to explain the difference, suppose that, by means of a lens, an image of the ring-system in white light is focussed upon the end of the collimator of a spectroscope, and that the slit can be moved sideways so as to admit to the spectroscope light from any chosen part of the ring-system. First, let it be placed to receive light from a point where the air film is .000018cm. thick. We suppose the illuminating beam of light falls upon the lens and disc perpendicularly, so that the difference in path for the interfering beams is twice the thickness, or .000036cm. We have seen that, because of the reversal of phase on reflection from the rarer side of a boundary, destructive interference takes place when this difference in path is one wavelength, or any integral number of wavelengths. Therefore absolutely no light of wavelengths .000036cm., .000018cm., .000012cm., etc., will enter the spectroscope. Only the first of these would be visible light, deep violet at that,—the rest are far in the ultraviolet and have no effect upon the visible appearance of the fringes. For light of wavelength .000072cm., (deep red), the difference in path is $\lambda/2$, and since the reversal in phase of one of the interfering beams is equivalent to another half wavelength very strong light of this color would enter the spectroscope. Therefore the spectrum would be very strong at the red end, fading off into absolute darkness at the violet end. Without the use of the spectroscope, this part of the ring-system would appear reddish, or rather orange, for the presence of some color of wavelengths shorter than red would undoubtedly alter the appearance somewhat toward the yellow side.

Now let the slit be shifted to receive light from a place where the thickness of the film is twice as great, .000036cm. The difference in path is now .000072cm., and the wavelengths destroyed by interference .000072cm. (deep red), .000036cm. (deep violet), .000024cm. (ultraviolet), .000018cm. (ultraviolet), etc. The spectrum would then be quite dark at both ends, shading into brightness in the middle, about the green. The color at this part of the ring-system as seen by the unaided eye would undoubtedly be green.

Change the slit again, so as to receive light from a place of thickness .000072cm., path-difference .000144 cm. The wavelengths destroyed by interference would be .000144 (infra-red), .000072 (deep red), .000048 (blue-green), .000036 (deep violet), and certain ultraviolet wavelengths. The visible spectrum would be dark at each end and have a dark band near the middle, at the blue-green, but it would be very bright in the yellow and also in the blue of shorter wavelength, the brilliancy shading off into darkness at the two ends and in the middle. The color as seen without the spectroscope would be a compound of strong yellow and strong deep blue, with some orange, green and blue-violet. It would probably be nearly white, somewhat yellowish.

In the first case, the spectrum has a single dark band, at the violet end,—in the second, a dark band at each end,—in the third one at each end and one in the middle. Of course, as the bands become more numerous, they also become narrower, for there is always a bright region between two dark regions. In each case, one of the dark bands has come at wavelength .000036, but that is only because we chose places where the difference in path was equal to that wavelength or a multiple of it. With the choice of another point, it would have been found that the deep violet was bright instead of dark. Since an increase in the thickness of the film causes more dark bands in the visible spectrum, let us see how many there would be at a place where the film is considerably thicker, say .000216cm. The path-difference is .000432, and the wavelengths destroyed by interference are found by dividing this length by the separate integers, 1, 2, 3, 4, etc. They are

.000432		.000072		.000033	
.000216		.000062		.000031	ultra-
.000144	infrared	.000054		.000029	violet
.000108		.000048	visible	etc.	
.000086		.000043			
		.000039			
		.000036			

There would then be seven dark bands in the visible spectrum, with bright bands between. What would be the color of such light? White light with the violet end of its spectrum suppressed would be reddish, with the red end suppressed bluish, with both ends suppressed greenish, with the middle alone suppressed purple, and so on. But when a large number of wavelengths, distributed regularly throughout the spectrum, are cut out, there is no reason why the remaining light should show one color more than another, since it contains constituents from all the spectral regions in the more general sense, though not light of every particular wavelength. Therefore, at such a place the color would be simply white, indistinguishable to the eye alone from the light with which the two discs, flat plate and lens, are illuminated. Evidently the same condition holds over all parts of the air-film where the thickness is several times the wavelength of red light or greater, and since all such parts are white, they will appear uniformly illuminated and no fringes will be distinguishable, although in monochromatic light they could be clearly seen.

The proof given above for the case of Newton's rings will evidently hold good just as well for any sort of interference experiment, so that we never observe interference fringes with white light unless the difference in path between the interfering beams is not more than a few wavelengths.

83. Rainbows.—The colors of rainbows are caused partly by interference, partly by a dispersion of light passing through raindrops, very similar to the action of a prism. The approximate positions of the principal bows is given by considering only the dispersion by the raindrops. The modifications introduced into the theory by the interference phenomena are difficult to understand, and they will be omitted from the discussion here given.

We shall first take up the primary bow, the only one that is usually seen. In order to see it, the back must be turned to the sun, and then it appears, if there are any raindrops in the right position and the direct sunlight reaches them, as an arc, whose center lies on a prolongation of a line from the sun through the eye of the observer. If the observer changes his position, the bow seems to move with him, so that the bow, the sun and the observer always keep the same relative position. Since this is what would happen if the bow were infinitely distant, we say that, like the rings seen by reflection in a thin uniform film, the bow is located at infinity. This is true even when it is formed by water drops quite close at hand, as in the spray from a fountain. The bow seems to move through the fountain as the observer moves, instead of staying fixed in it.

The bow is caused by light that has suffered two refractions and one reflection: that is, by light that has been refracted into the raindrop, reflected once inside, and then refracted out again. Of course, the refractions cause a dispersion of the light, just as a prism would, but the case is less simple than that for a prism, because the surfaces of the raindrop are curved while those of a prism are plane. When parallel rays fall upon a prism, all the light of a given wavelength will be refracted into the same direction, but in the case of the spherical drop they are spread out into a number of different directions. This can be seen from figure 85, where three parallel rays from the sun are drawn, striking the sphere at slightly different positions, therefore having different paths through the drop and emerging in quite different directions below. The angle D, between the direction of a ray before entering the drop and its direction after leaving, called the deviation of the ray, is different for rays striking the surface at different angles of incidence. It is not difficult to show that for any ray

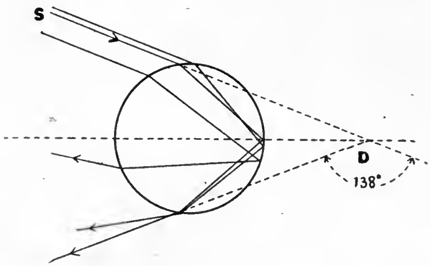

Figure 85

$$D = 180° + 2i - 4r$$

where i is the angle of incidence and r the corresponding angle of refraction. If we calculate and plot the values of D for various values of i, using of course the relation that

$$\sin i = n . \sin r$$

n being the index of refraction of the drop for the particular wavelength we are considering, the graph will come out to be like that of figure 86. For increasing values of i, the value of D first decreases, then increases. If we take n = 1.33, which is correct for a certain part of the yellow, the minimum value of D is 138°, and it occurs when i = 61°. This property, that D has a minimum value, is of great importance, for evidently rays having a deviation in the immediate neighborhood of 138° would be far more numerous than those having a deviation considerably more than this, and there are none at all with a deviation less than this. Therefore, although not all of the light of the wavelength in question is sent out from the drop in one direction, most of it is sent out very nearly in one direction.

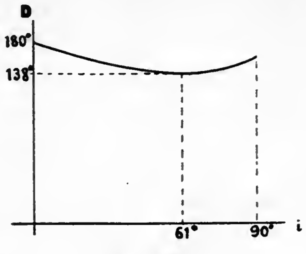

Figure 86

Now a ray having a deviation 138° would come to the observer's eye as from a direction making an angle of 180° — 138° = 42° with a line drawn from the sun through the eye, as shown in figure 87. Therefore all drops situated upon a cone of half-angle 42°, with the eye as apex, would send to the eye a greater amount of light of this wavelength than any other drops, and the observer would see a yellow circular arc. If we take red light, whose index of refraction is smaller, the angle of minimum deviation would be smaller than 138°, and therefore the half angle of the cone on which lie the drops giving maximum red light would be greater than 42°, that is, the red circle would be outside the yellow, and correspondingly the circles for shorter wavelengths would lie inside. Therefore the bow is red on the outside and violet on the inside. Notice that the drops giving the maximum red light are not the same as those giving the maximum of other colors, and that if the observer changes his position the drops which formerly sent to him any particular color will no longer be in the proper direction to do so, and their function will be supplied by other

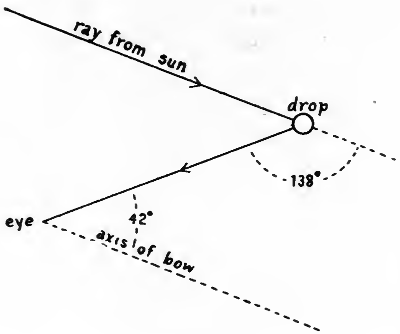

Figure 87

drops which are in proper relation to his new position. This accounts for the bow appearing to be at an infinite distance. Notice also that, because a drop sends not all, but only the greater part, of its light in one general direction for a given wavelength, there will be a great deal of overlapping of colors. A given color overlaps all the other colors that lie within it, that is, red overlaps all the others, orange overlaps yellow, green, blue and violet, yellow overlaps green, blue and violet, etc. Only the red is even approximately pure.

The angular radius of the bow is somewhat modified from the figures given above (42° for the yellow, etc.) by the interference phenomena already mentioned, which we shall not discuss here, and by the fact that the sun is not a point-source but subtends a definite angle.

Not all the light that enters a raindrop traverses such a path as is shown in figure 85. Some is reflected at the first incidence, without entering the drop. Some is refracted out where it strikes the surface the second time, without any internal reflections, and this light might also form a bow, but that there is for this case nothing like a minimum deviation. Some light is refracted out of the drop after two internal reflections, some after three, etc., and theoretically there should be a bow for each such case. The second bow, that formed after two internal reflections, is in fact often seen. It is formed outside the first bow, and has its colors in reverse order, and it is of course fainter. The theory of its formation is quite similar to that for the first, except that the light which forms it comes into the drop from the lower side and emerges from the upper side. The third and fourth bows, possible in theory, are still fainter than the second, and to make matters still more unfavorable they come in the bright part of the sky near the sun. Therefore they cannot be seen. Bows of still higher order are inherently too faint for visibility.

Rainbows are often classed with other natural optical phenomena of the atmosphere, such as halos, coronas, mock-suns, mirages, etc., under the general heading of "meteorological optics." A good account of many of these phenomena is given in Wood's "Physical Optics," or in W. J. Humphreys' "Physics of the Air," part III.

84. Motion relative to the ether.—In section 68 it was shown how the Doppler effect, applied to the spectrum of a star, gives us a means of finding the relative velocity of the star with respect to the earth. It was also stated that velocities measured in this way are usually corrected for the earth's orbital velocity so as to give the star's velocity referred to the sun. Whether the sun itself is in motion or at rest is a question without sense, for we cannot specify motion without referring to some body which we regard as fixed. In other words we can conceive of motion only as a relative matter, and absolute motion is an absurdity.

But it is not absurd, on the face of things, to speculate whether the sun is at rest or in motion with regard to the ether, since up to the present we have regarded that medium as if it partook in some measure of the nature of ordinary material things. A passenger in an airplane or submarine could readily tell that he was moving with respect to the surrounding air or water, even without taking note of surrounding solid objects, and it is conceivable that some optical experiment might enable us to detect and measure our velocity with respect to the ether.

In 1887 Michelson and Morley carried out an experiment devised with the above purpose in view. It was based on the assumption, already tacitly made in the discussion of the Doppler effect, that the velocity of light-waves in the ether, like that of sound-waves in air, does not partake of the velocity of the body which emits them. If, in figure 88, a body moving from Q toward P emits a wave when it is at the point X, this wave spreads out into an enlarging circle whose center remains at X and does not follow the emitting body. If the circle, with X as center, has a radius equal to c, the velocity of light, then it would represent the wavefront one second after emission. If the velocity of the body itself is v, then at the end of the second it would be at the point Y, where XY = v. If

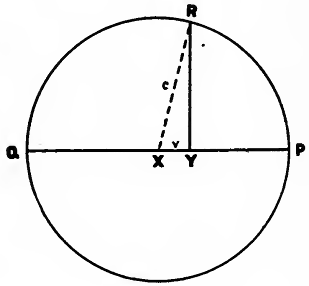

Figure 88

this is the case, then to a person moving with the emitting body the velocity of light should appear to be unequal in different directions. Toward P the velocity would be $YP = c - v$, toward Q it would be $YQ = c + v$, and toward R, at right angles to the line QP, it would be $YR = \sqrt{c^2 - v^2}$. Now refer to the diagram of the Michelson interferometer, figure 83, remembering that the interference is between two beams going from the diagonal mirror EF, one up to AB and back, the other to CD and back. Let the instrument be adjusted till these two paths have exactly the same length, and then suppose the whole interferometer to be mounted upon some base which is moving rapidly to the right in figure 83 with a velocity v.

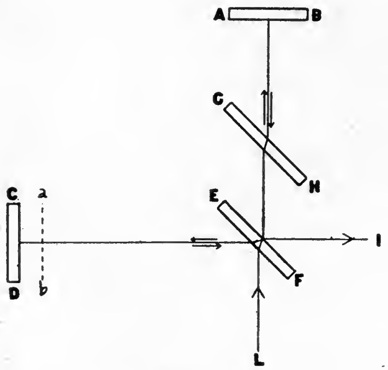

Figure 83

The two paths, in spite of being equal in length, would then not include the same number of wavelengths, because the velocity of the light, with respect to the interferometer, would be different in different directions. From the diagonal mirror to AB and back, it would be $\sqrt{c^2 - v^2}$; from the diagonal mirror to CD it would be $c + v$; from CD back to the diagonal mirror it would be $c - v$. Consequently, the fringes would be slightly shifted from their position when the instrument was at rest with respect to the ether. If it were turned through 90°, so that the path from EF to AB lay along the direction of the velocity v, the shift would be in the opposite direction. Instead of calculating the amount of the shift, we shall merely calculate the difference in the time for the two interfering beams of light to come together again after the separation. If d is the distance from the diagonal mirror to either of the mirrors AB and CD, the time required for the beam that goes up to AB and back is

$$\frac{2d}{\sqrt{c^2 - v^2}} = \frac{2d}{c}\left(1 + \frac{v^2}{2c^2} + \frac{3v^4}{4c^4} + \text{etc.}\right) = \frac{2d}{c}\left(1 + \frac{v^2}{2c^2}\right)$$

approximately, while that required for the other beam is

$$\frac{d}{c+v}+\frac{d}{c-v}=\frac{2cd}{c^2-v^2}=\frac{2d}{c}\left(1+\frac{v^2}{c^2}+\frac{v^4}{c^4}+\text{etc.}\right)=\frac{2d}{c}\left(1+\frac{v^2}{c^2}\right)$$

approximately. The terms in the fourth and higher powers of v/c have been dropped because any attainable velocity of matter is so small compared to that of light.

The difference in these two times is v^2d/c^3. If we use for v the orbital velocity of the earth, this difference is exceedingly small. But the interferometer is an exceedingly sensitive instrument, and Michelson and Morley were able to increase its sensitiveness by introducing a number of additional reflections whose effect was to increase the distance d. The whole was then mounted upon a stone slab floated in mercury so that it could be turned about a vertical axis. It was turned so that first one arm, then the other, was parallel to the earth's orbital motion, but the expected shift of the interference fringes *did not take place.*

85. The relativity theory.—This failure of Michelson and Morley's experiment was a crisis in the history of physics. There can be no doubt that the earth is actually in motion, and that the apparatus should have detected this motion if no blunder was made in the theory. If light takes on the velocity of the emitting body, as it would if Newton's corpuscular theory were correct, then the failure of the experiment is what we should expect, but this seems impossible if light consists of waves. Some physicists indeed have favored throwing over the wave theory, and the ether with it, but this cannot be done in view of all the phenomena of interference. Fitzgerald and Lorentz have pointed out that the Michelson and Morley experiment can be reconciled with the wave theory if we assume that a body in motion is very slightly shortened in all those dimensions that lie parallel to its motion. Later a complete doctrine known as the "Relativity Theory" was worked out, based on two fundamental postulates: first, that it is impossible to detect by any means any relative motion between matter and the ether,—second, that the velocity of light in the free ether will always come out the same, no matter under what circumstances it is measured. From these postulates the change of dimensions of a moving body suggested by Fitzgerald and

Lorentz follows as a necessary consequence, but it also follows that the mass of a body is changed, and that the time-unit is altered, by motion. All these changes are exceedingly small unless the velocity becomes comparable with the velocity of light.

Einstein has in recent years extended the relativity principle to accelerated, as well as to steady motion, as a result of which he was led to predict, among other things, that a beam of light in passing near a body with a strong gravitational field, like the sun, would be deflected. This prediction has been apparently verified by observations during a recent solar eclipse. The complete theory involves fundamental changes in our idea of time and space, and is very metaphysical and mathematical, as well as physical.

Problems.

1. Show that, in the circular fringes produced by uniform films, a ring of small angular diameter corresponds to interference of a high order, that is, a high value of N. Find the highest order of interference for a film of glass, index 1.54, 1/10 mm. thick, for light of λ .00005 cm.

2. What would be the effect upon the fringes seen in the Michelson interferometer, of inserting a very thin slip of glass into one of the arms, say between CD and EF of figure 83?

3. Show that, in the phenomenon of Newton's rings, the radii of successive dark rings are approximately proportional to the square-roots of the successive integers, assuming that the thickness of the film alone determines the interference.

4. Prove the statement in paragraph 83, that $D = 180° + 2i - 4r$.

CHAPTER X.

86. Simple harmonic motion.—87. Velocity in S. H. M.—88. Acceleration in S. H. M.—89. Energy in S. H. M.—90. Two parallel S. H. M.'s.—91. Application to cases of interference.—92. Two S. H. M.'s at right-angles.—93. Lissajous figures.

86. Simple harmonic motion.—In all wave phenomena, we are much concerned with a particular kind of vibratory motion known as *simple harmonic motion,* and this name will occur so frequently in this chapter that we shall at the outset adopt for it the abbreviation S. H. M. Its definition is as follows:

The motion of a body P, figure 89, is simple harmonic along a line MN when it moves so that if a body O be imagined travelling at a uniform rate in a circle with NM as diameter, P keeps always at the foot of a perpendicular drawn from O to MN. The time in which O completes the circuit, which is the same as the time in which P passes from M to N and back again, is called the *period,* and will be represented by the letter τ. The distance CM, half the range of motion P, called the *amplitude*, will be represented by K. The distance CP, represented by y, from the middle position of P to that position which it momentarily occupies, is the *displacement;* and the corresponding angle OCM, represented by ϕ, is the *phase.* K and τ are constants for the motion, while y and ϕ are variable with the time. We consider y to be positive when P is above C, negative when it is below, so that y varies from K to —K, oscillating between these values.

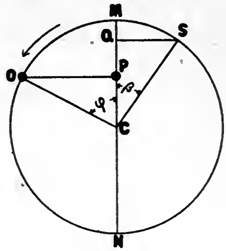

Figure 89

Our first problem will be to find a mathematical relation between the displacement y and the time, represented by t. Since P is at the foot of the perpendicular from O, cos. ϕ = CP/CO, therefore

$$y = K. \cos \phi$$

Since O moves uniformly in a circle, the time required for it to move through the angle ϕ will be to the time necessary to complete the circuit, as ϕ is to 2π. Therefore, if we are to count time from the instant when P is at M, the end of its path, $t : \tau :: \phi : 2\pi$, or

$$\phi = 2\pi t/\tau$$

and

$$y = K. \cos \frac{2\pi t}{\tau}$$

Often, however, it is convenient to count time, not from a particular instant such as that when P is at the end or at the middle of its path, but from some instant chosen at random, as when P is at such a point as Q and the corresponding position for O is at S. Our formula may easily be modified to suit this more general condition. Represent the angle SCM by β. Then, since, during the time t, O has moved through the angle SCO $= \phi + \beta$, instead of

$$t : \tau :: \phi : 2\pi$$

we must write

$$t : \tau :: (\phi + \beta) : 2\pi$$

giving

$$\phi = \frac{2\pi t}{\tau} - \beta$$

and

$$y = K. \cos\left(\frac{2\pi t}{a} - \beta\right) \qquad (1)$$

This is the required mathematical relation between y and the time, and may be called the *equation* of the S. H. M. y could easily enough have been put into the form of a sine instead of a cosine, by simply considering the complement of ϕ instead of ϕ itself. We might then have simply defined a S. H. M. at the start as a motion in which the displacement is a sine or cosine function of the time, and so avoided speaking of the body O of the figure, which is purely imaginary, and was introduced only to make the definition easier to think of physically.

There is a close relation between equation (1) above and equation (5) of chapter III, which is

$$y = K. \cos \frac{2\pi}{a} (x - Vt - \epsilon)$$

The latter, since it is the complete equation for a wave, involves, besides the variables y and t, the third variable x, which is the distance from a fixed origin, measured in the direction in which the wave is progressing. But, if we fix our attention upon a definite position in the medium, and consider only the motion there, x will be regarded as, for the time being, a constant, and therefore we may bracket it with the other constant ϵ. The equation then becomes

$$y = K. \cos \left[\frac{2\pi Vt}{a} - \frac{2\pi}{a} (x - \epsilon) \right]$$

(The sign of the quantity whose cosine is being taken has been changed, because the cosine of a positive angle and the cosine of an equal negative angle are always the same.) Now V is the velocity of the wave and a is the wavelength, as can be seen by referring back to section 21, and therefore $a = V\tau$ or

$$V/a = 1/\tau$$

Therefore we can write

$$y = K. \cos \left[\frac{2\pi t}{\tau} - \frac{2\pi}{a} (x - \epsilon) \right]$$

This is exactly the same form as equation (1) above, the place of β being taken by $2\pi(x - \epsilon)/a$. Therefore any definite part of the medium must, when a monochromatic wave passes through it, go through a S. H. M. The value of the phase-constant β is greater the farther along be the point considered; that is, although a point far from the source goes through the same motion as a point nearer, it is behind it in phase.

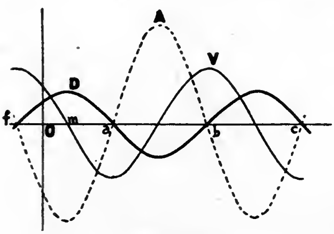

Figure 90

The relation between y and t, expressed in equation (1), is very well shown in a graph such as the heavy curve D in figure 90, where t is the abscissa and y the ordinate. It is the same kind of curve as those shown in figure 19, except that there the abscissa was distance, while here it is time. Evidently the maximum height or depth of the curve, measured from the horizontal axis, is the amplitude K, while the distance ac is the period τ. Those portions of the curve that slope upward to the right indicate that P in figure 89 is moving up, while those that slope downward to the right indicate that P is moving downward. The graph is drawn not only to the right of the origin, but also to the left, corresponding to negative times, for we may certainly suppose that the motion was going on before the instant from which we measure time. In fact fO is the time it took, or would take, for P to move from C out to Q, and Om is the time to move from Q to the end of the path.

87. Velocity in S. H. M.—The simplest considerations show that the *velocity* of the moving point P must be zero at either end of its path, and greatest when the displacement y is zero. Since P is moving sometimes upward and sometimes downward, we shall agree to call the velocity, v, positive when P is moving upward, whether it is actually above C or below it, and negative in the contrary case. Then if we draw another graph, in which the abscissa again represents time but the ordinate the corresponding velocity, it will be similar to the displacement-curve, having the same period, but it will cross the axis at points where the displacement is greatest, in a positive or a negative direction.

The velocity-curve could be constructed graphically from the displacement-curve by means of the following considerations: A velocity is the ratio of a very short distance to the very short time in which that distance is traversed, that is

$$v = \text{limit} \frac{\Delta y}{\Delta t}$$

The symbol Δ placed before a quantity means a very small increase in that quantity, so that Δt indicates the time in which the displacement of P increases by the small amount Δy. But in a graph with Cartesian coordinates, the slope of the tangent is also the limit of $\Delta y / \Delta t$. Therefore, the velocity of P at any instant t can be gotten from the displacement-curve by simply measuring the slope of the tangent, at a horizontal distance t from the origin. If the slope found be erected as another ordinate, the curve so constructed will be the required velocity-curve. The curve V in figure 90 is constructed in this way. The amplitude is not necessarily the same as for the displacement-curve, and whether it is greater or less depends upon the value of the period. It is evident from figure 90 that a shortening of the period would increase the slope of the D-curve, and also from physical considerations it is clear that, the shorter the period, the greater would be the velocity with which the body swings through its middle position.

The mathematical formula for the velocity-curve is derived as follows: Suppose the time, initially represented by t, increases by a small amount Δt. The initial displacement is

$$y = K. \cos\left(\frac{2\pi t}{\tau} - \beta\right)$$

The displacement after the change in time is

$$y + \Delta y = K. \cos\left(\frac{2\pi t}{\tau} - \beta + \frac{2\pi \Delta t}{\tau}\right)$$

The last equation may be expanded by making use of the trigonometrical relation for the cosine of the sum of two angles,

$$\cos(x + y) = \cos x \cos y - \sin x \sin y$$

letting $\frac{2\pi t}{\tau} - \beta$ take the place of x and $\frac{2\pi \Delta t}{\tau}$ that of y. We then have

$$y + \Delta y = K. \cos\left(\frac{2\pi t}{\tau} - \beta\right) . \cos\frac{2\pi \Delta t}{\tau} - K. \sin\left(\frac{2\pi t}{\tau} - \beta\right) \sin\frac{2\pi \Delta t}{\tau}$$

Since we are to find the limit of the ratio $\Delta y / \Delta t$, in which Δt approaches zero, and since the cosine of an angle approaches unity and the sine approaches the value of the angle itself when the angle becomes very small, we may make the following substitutions:

$$\cos\frac{2\pi \Delta t}{\tau} = 1 \qquad \sin\frac{2\pi \Delta t}{\tau} = \frac{2\pi \Delta t}{\tau}$$

Therefore, in the limit,

$$y + \Delta y = K. \cos\left(\frac{2\pi t}{\tau} - \beta\right) - \frac{2\pi \Delta t}{\tau} . K. \sin\left(\frac{2\pi t}{\tau} - \beta\right).$$

If we subtract from this equation the equation giving the value of y itself, we get

$$\Delta y = -\frac{2\pi\Delta t}{\tau} \cdot K. \sin\left(\frac{2\pi t}{\tau} - \beta\right)$$

and if we divide by Δt, we get

$$v = \text{limit}\ \frac{\Delta y}{\Delta t} = -\frac{2\pi K}{\tau} \sin\left(\frac{2\pi t}{\tau} - \beta\right) \qquad (2)$$

This is the equation for the velocity-curve V in figure 90, or the formula for the velocity.

88. Acceleration in S. H. M.—It is necessary also to consider the *acceleration,* or rate of change in velocity, of a body in S. H. M. It is defined as

$$a = \text{limit}\ \frac{\Delta v}{\Delta t}$$

therefore it bears the same relation to the velocity as the latter bears to the displacement, and the acceleration-curve can be drawn by plotting, for each value of the time, the corresponding value of the slope of the V-curve. The dotted curve A is obtained in this manner. From what has already been said of the relation between the V-curve and the D-curve, it follows that the A-curve will have its peak just where the V-curve is crossing the axis on the rise, and this of course entails that the peak of the A-curve and the trough of the D-curve come at the same time,—in other words, the acceleration and the displacement are exactly opposite in phase.*

The equation for the A-curve can be derived quite easily, by the same method we used for the V-curve. The velocity v at a given instant t is

*It is sometimes difficult for a student to conceive that the body P, figure 89, actually has its greatest acceleration when it is at the bottom of its path, and the velocity is momentarily zero. It must be borne in mind that just before P reaches the bottom its velocity is downward in direction, while just afterward it is upward, so that it is precisely at this point that the velocity is changing most rapidly. There is nothing incompatible in a body having a considerable acceleration when its velocity (for a single instant only) is zero. For instance, a stone thrown into the air has the acceleration 980 cm. per sec^2. even at the instant when it is at the top of its path.

$$v = -\frac{2\pi K}{\tau} . \sin\left(\frac{2\pi t}{\tau} - \beta\right)$$

let t be increased by the very small amount Δt, and let Δv be the corresponding increase in v. The new value of v is

$$v + \Delta v = -\frac{2\pi K}{\tau} . \sin\left[\frac{2\pi (t + \Delta t)}{\tau} - \beta\right]$$

$$= -\frac{2\pi K}{\tau} . \sin\left[\frac{2\pi t}{\tau} - \beta + \frac{2\pi \Delta t}{\tau}\right]$$

Remembering that $\sin (x + y) = \sin x . \cos y + \cos x . \sin y$, and substituting $\frac{2\pi t}{\tau} - \beta$ for x and $\frac{2\pi \Delta t}{\tau}$ for y, we have

$$v + \Delta v = -\frac{2\pi K}{\tau} . \sin\left(\frac{2\pi t}{\tau} - \beta\right) . \cos\frac{2\pi \Delta t}{\tau} - \frac{2\pi K}{\tau} . \cos\left(\frac{2\pi t}{\tau} - \beta\right) . \sin\frac{2\pi \Delta t}{\tau}$$

Again considering that we are to take the case of the limit, we substitute

$$\cos\frac{2\pi \Delta t}{\tau} = 1 \qquad \sin\frac{2\pi \Delta t}{\tau} = \frac{2\pi \Delta t}{\tau}$$

$$v + \Delta v = -\frac{2\pi K}{\tau} . \sin\left(\frac{2\pi t}{\tau} - \beta\right) - \frac{4\pi^2 K \Delta t}{\tau^2} . \cos\left(\frac{2\pi t}{\tau} - \beta\right)$$

Subtracting the value of v itself,

$$\Delta v = -\frac{4\pi^2 \Delta t}{\tau^2} . K . \cos\left(\frac{2\pi t}{\tau} - \beta\right)$$

and dividing by Δt,

$$a = \text{limit}\ \frac{\Delta v}{\Delta t} = -\frac{4\pi^2 K}{\tau^2} \cos\left(\frac{2\pi t}{\tau} - \beta\right) \qquad (3)$$

Another form for the equation may be obtained by substituting y instead of the factor

$$K \cos\left(\frac{2\pi t}{\tau} - \beta\right)$$

which gives

$$a = -\frac{4\pi^2}{\tau^2} y \qquad (4)$$

The minus sign in equation (4) indicates that whenever the displacement is upward the acceleration is downward, and vice versa, which means of course that the acceleration, and therefore the force acting, are always such as to drive the body back to its central position, C. The equation also shows that the acceleration is directly proportional to the displacement, the factor of proportionality being $4\pi^2$ divided by the square of the period.

89. Energy in S. H. M.—A body in S. H. M. is of course endowed with a certain amount of energy, which will be constant unless there are some disturbances. At the end of the swing the energy is all potential, while at the middle it is all kinetic, and at other positions it is partly potential and partly kinetic. The simplest way to find an expression for the total energy would be to find the potential energy at the end of the swing or the kinetic at the middle. We shall do both, and show that the same expression is found in each case.

The potential energy at the end of the path is the work that must be done to pull the body out to that point, starting with it at rest in the central position. Work is force applied times the distance moved, and the force is the body's mass, m, multiplied by the acceleration which the force, acting alone, would produce in it. If the body be pulled out slowly enough, this acceleration will be just equal and opposite to that which the body has when freely swinging, which we have found to be $-4\pi^2 y/\tau^2$. Therefore the force applied must be the mass m multiplied by the equal and opposite acceleration $+4\pi^2 y/\tau^2$. That is

$$F = \frac{4\pi^2 my}{\tau^2}$$

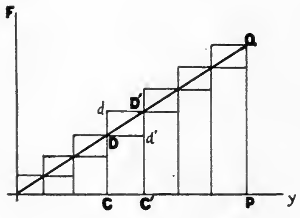

Figure 91

We cannot get the work by multiplying this expression by the whole distance moved, K, for y is a variable, and therefore F is also, having the value zero at the beginning of the motion and the value of $4\pi^2 mK/\tau^2$ at the end. We must calculate the work by infinitesimal steps, F having a different value for each step. To do this, consider figure 91, in which OQ is a graph whose abscissa is y and ordinate the corresponding value of F. In accordance with the above equation, it is a straight line. The distance OP represents the whole distance moved, or K, while PQ is the force necessary to hold the body still at the end of its path, $4\pi^2 mK/\tau^2$. We shall prove that the work done in pulling the body out to the end of its path is equal to the area of the triangle OPQ.

Recalling that the work must be calculated step by step, let us see how much work is done in moving the body from C to C′, the distance CC′ being supposed very small. At the beginning of this step, the force applied is CD, and if it continued to have this value through the step, the work would be CD × CC′, which is the area of the rectangle CDd′C′. At the end of the step, the applied force is C′D′, and if it had this value throughout the step the work would be C′D′ × CC′, the area of the rectangle CdD′C′. Evidently the actual work done during the small motion from C to C′ lies between these two values, and when we consider the work done during the whole motion from O out to P, it follows that this lies in value between the area of the stair-shaped figure extending above the line OQ and that of the similar figure all of which is below OQ. If the steps are made smaller and smaller, the area of each of these figures comes nearer and nearer to the area of the triangle OPQ, therefore we can say that, in the limit, the work done is equal to the area of that triangle, that is, the potential energy at the end of the swing, or the total energy at any time, is

$$\frac{1}{2}\,\mathrm{OP} \times \mathrm{PQ} = \frac{1}{2}\,\mathrm{K} \times \frac{4\pi^2 \mathrm{mK}}{\tau^2} = \frac{2\pi^2 \mathrm{mK}^2}{\tau^2} \qquad (5)$$

The calculation of the kinetic energy when the body is passing through its central position is easy. For kinetic energy is one half the mass times the square of the velocity, and the velocity in the central position is simply the greatest value that the body can have, which is gotten from equation (2) by putting the sine equal to 1. Therefore the kinetic energy at this point, or the total energy of the body at any time, is

$$\frac{1}{2}\,\mathrm{m} \times \left[-\frac{2\pi \mathrm{K}}{\tau}\right]^2 = \frac{2\pi^2 \mathrm{mK}^2}{\tau^2} \qquad (5)$$

the same expression.

The fact that the energy of a body in S. H. M. is proportional to the square of the amplitude of vibration is quite important, but it is, after all, what we might have expected. For, in order to set a body in vibration, first with an amplitude of 1 inch, and later with an amplitude of 2 inches, we must not only pull it out twice as far in the second case, but also exert a force whose average value is twice as great so that the work done is four times as great.

90. Two parallel S. H. M.'s.—In examples of interference, and in other problems in light or sound, it is often necessary to consider the application simultaneously of two S. H. M.'s to the same body. We shall suppose that the two motions have the same period, and are in the same direction, but they may differ to any degree in amplitude, and they may have the same phase, opposite phases, or any desired difference in phase. An excellent example of the superposition of two S. H. M.'s in this way, where the motion however is applied to an illuminated spot on a screen instead of to a material body, is shown in figure 92. H is a hole through which streams a beam of

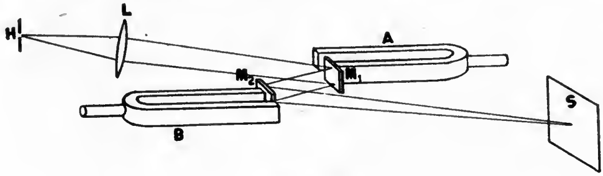

Figure 92

light from the sun or a bright artificial source, L a lens, A and B two tuning forks, S a white screen. The forks have the same period and are arranged to vibrate in the same plane. Each has a small mirror (M_1 and M_2) mounted on one prong, and the parts are arranged so that the light is reflected from M_1 to M_2, and thence to the screen, where it comes to a focus, forming a small round image of the hole H. If either fork is set in vibration while the other is held still, the illuminated spot will be set into S. H. M. If both forks are vibrating together, the spot will also have a vibratory motion. We are to find out whether this motion is simple harmonic, and if so what its amplitude and phase-constant will be.

Suppose that the motion of the spot due to the first fork alone is represented by the equation

$$y_1 = K_1 \cos\frac{2\pi t}{\tau} = r_1 \cos \omega t$$

that due to the second fork alone by

$$y_2 = K_2 \cos\left(\frac{2\pi t}{\tau} - \beta\right)$$

It is unnecessary to put a phase constant into each expression, since we may suppose that the zero-point for time measurement is so chosen that the phase-constant of the first motion is zero. Then β is simply the difference in phase, or the amount by which the second fork lags behind the first in phase.

The resultant motion, when both forks are vibrating, is gotten by adding y_1 and y_2, for evidently the movement due to one fork will be superimposed upon that due to the other. We shall discuss the problem graphically, making use of the definition of S. H. M., that it is the projection, upon a straight line, of uniform motion in a circle. In figure 93, let ba be the path of the vibration due to the first component motion. The large circle, of which the diameter is ab and the radius equal to K_1, will be called the reference-circle for the first motion, and since we have taken the phase constant for that motion as zero, it alone would for time zero put the spot of light at a, the end of the diameter, and the tracing point moving in the reference circle would also be there. That is, $y_1 = K_1 = oa$, at that instant. In order to add in y_2, the component due to the second vibration, we draw a reference-circle with a as center and K_2 as radius, and lay off the angle eac, equal to the phase-difference β. Then, for time zero, the second component of the motion gives a displacement $y_2 = ad$, the projection of ac, and the total displacement is $y_1 + y_2 = od$. Notice also that od is the projection of the single line oc.

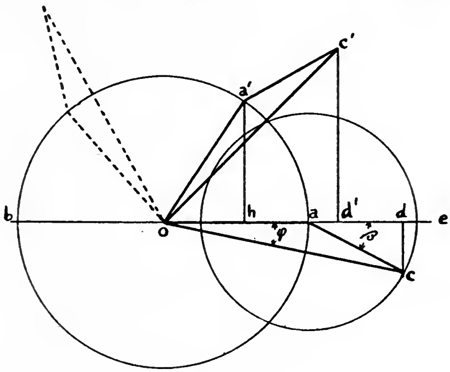

Figure 93

Now consider the state of affairs at a later instant, when the tracing point in the first reference-circle has moved to a′, through the angle aoa′. Let us suppose that, along with the radius oa, the whole triangle oac rotates about the point o as if it were rigid, taking the new position oa′c′. Whenever a rigid plane figure rotates about a point in its own plane, all lines of the figure turn through the same angle in the same time. Therefore the angle coc′, turned through by the side oc, is equal to the angle aoa′, turned through by the side oa. Also, if c′a′ and ca be produced till they meet, the angle so formed will also be equal to aoa′. (These statements can easily be proved by simple geometry.) Since the two S. H. M.'s have the same period, the radii oa and ac must necessarily turn through the same angle in the same time. Therefore a′c′ will be the direction of the tracing-point radius in the second motion at the same instant when oa′ is that for the first. For that instant, then, y_1 must be equal to oh, the projection of oa′, y_2 to hd′, the projection of a′c′, and the total displacement is

$$y_1 + y_2 = oh + hd' = od'$$

and od′ is nothing but the projection of the third side of the triangle oc′, in its new position.

The same result holds, whatever may be the angle through which oa has turned. In each case, the resulting displacement is the projection of the third side oc in its new position. Since this is the projection of a line of fixed length, rotating with the same period as the two component S. H. M.'s, the resulting motion is itself evidently simple harmonic, of the same period as y_1 and y_2, of amplitude oc, and behind the first motion y_1 in phase by the angle aoc, or ϕ.

We have proved that the following rule holds for finding the resulting motion when two S. H. M.'s of the same period and direction act upon the same body: Lay off a line oa whose length is the amplitude of the first motion. From a lay off another line ac, whose length is equal to the second amplitude. so that the angle it makes with oa produced is equal to the amount by which the second component lags in phase behind the first. Then, completing the triangle, the side oc gives the resulting amplitude, and the motion lags behind the first component in phase by the angle aoc. Since this is exactly the rule for finding the resultant of two vectors, the angle between them being the phase-difference, it may be said that S. H. M.'s are compounded as vectors, but it should be remembered that this statement refers only to a diagrammatic representation, and has nothing to do with the actual directions in which the vibrations take place.

The resultant vibration can be represented by the formula

$$y = K.\cos\left(\frac{2\pi t}{\tau} - \phi\right)$$

where K = oc, and ϕ = angle aoc. Since, from trigonometry,

$$oc^2 = oa^2 + ac^2 + 2 \times oa \times ac \times \cos(eac)$$

it follows that

$$K^2 = K_1^2 + K_2^2 + 2\,K_1\,K_2 \cos\beta \qquad (6)$$

If both the component motions have the same amplitude, so that $K_2 = K_1$

$$K^2 = 2\,K_1^2\,(1 + \cos\beta) \qquad (7)$$

If $\beta = 0$, $\cos\beta = 1$, and $K^2 = 4K_1^2$, or $K = 2K_1$, as can be seen either from equation (7) or from consideration of the above-mentioned rule for finding the resultant amplitude by a vector diagram. In this case the energy of the resultant S. H. M. is 4 times that of either component, because the energy depends upon the square of the amplitude.* If $\beta = \pi$, or 180°, then $\cos\beta = -1$, and $K = 0$, as shown by equation (7) or by the vector diagram. For other values of β, K will lie between 0 and $2K_1$.

91. Application to cases of interference.—These results can be easily applied to any case of interference where only two beams of light are concerned. Take, for example, the interference with Fresnel's mirrors, and refer to figure 20. The two slits S_1 and S_2 being nearly equally distant from any portion of the screen AB, we can regard the amplitude of the two beams as being equal. At C, where the difference in path

*In the arrangement depicted in figure 92 there is no energy associated with the movement of the illuminated spot on the screen, since this motion is not a motion of a real thing, but only a transferrence of illumination from one place to another. The light vibrations producing the illumination of course have energy, and so do the two forks.

is zero, at M_1 or M_1' where it is a wavelength, at M_2 or M_2' where it is two wavelengths, etc., the phase-difference is zero or, what amounts to the same thing, an integral multiple of 2π (360°), the resulting amplitude is twice that which one beam alone would produce, and the brightness four times as much. At the points m_1 and m_1', where the path difference is $\lambda/2$, at m_2 and m_2' where it is $3\lambda/2$, etc., the phase-difference is either π (180°) or, what amounts to the same thing, an odd multiple of π, and the cosine is -1. These points therefore have zero resulting amplitude, and they will be quite dark. At other points, the cosine of the phase-difference is neither $+1$ nor -1, but some intermediate value, and therefore the brightness will be neither so great as at C nor absolutely zero as at m_1.

The vector diagram construction for finding the resultant of two S. H. M.'s can be extended to cases of three or any number of component motions, provided all have the same period and all are in the same direction. We simply regard each amplitude as a vector, the direction of which on the diagram is given by the phase-constant, and find the resultant, the length of which gives the amplitude of the resultant vibration. Many rather complicated phenomena of interference or diffraction can be easily treated by this method.

If the student will refer back to figure 69, chapter VII, he should recall that, in the discussion of a grating with 9 openings, it was found that at the point I the rays from all the different openings arrived together in phase, while at points x and x′, close to I, the rays opposed one another in pairs by coming together in opposite phases, that is with a phase-difference which is an odd multiple of π, leaving only the ray from a single opening to exert its full effect. Therefore the amplitude of the resulting S. H. M. which produces the light at I is nine times that at x or x′, and the energy of vibration, to which the brightness is proportional, is 81 times as great at I as at x or x′.

A very interesting application of the principles we have taken up in this chapter, is in connection with the light sent out by a luminous source such as a flame. The actual centers from which the light is emitted are millions of radiating atoms or molecules, and since these are independent of one another in their vibration, we must assume that all sorts of phase-differences exist between them. It might be argued that since one of these radiating centers is as likely to have one phase-constant as any other, therefore in the whole assemblage they should annul one another's effects, with the result that no light would be emitted by the flame at all. Such a contention would be as erroneous as the statement, sometimes loosely made, that in a sufficiently large number of throws of a coin, it would show heads just as often as tails. Both are due to a misunderstanding of the laws of probability. It is true that in a great number of throws of a coin, the difference between the number of heads and the number of tails becomes less and less *as compared to the total number of throws,* but the absolute amount of the difference tends to increase. In fact, the probable excess of heads over tails or conversely of tails over heads (theory cannot predict which it will be) is directly proportional to the square-root of the number of throws. Similarly, theory predicts that in the case of a great number of radiating centers, with phases distributed at random, the probable amplitude of the resulting vibration determining the illumination at any place will be proportional to the square-root of the number of radiating centers. Therefore the energy and the brightness of illumination should be proportional to the number of centers. A flame of double size should then give double illumination, other things being equal, and this is what is actually found.

If we were able to control the phases of all the centers of radiation in a flame, so that all the rays arrived at the eye in the same phase, the brightness would be proportional to the square of the number of centers, and therefore would be enormously increased. Under these circumstances a sodium flame which is actually very feeble would appear many times brighter than the sun and would be absolutely destructive.

A similar case occurs in sound. Suppose that in an orchestra we have a large number of instruments of the same kind, say 100, all striking the same note at once. If, by any chance, the sound waves from all these instruments reached the ear in the same phase, the intensity of the sound heard, proportional to the square of the resulting amplitude, would be 10000 times as great as that heard when only one instrument is being sounded, and the result would be deafening. The chance for this to occur, however, is almost vanishingly small. In general, the phases are distributed at random, and the sound heard is proportional to the number of instruments.

92. Two S. H. M.'s at right-angles—Another kind of combination of two S. H. M.'s, the importance of which in the study of light will be shown in chapters XII and XIII, is one in which the vibrations are at right-angles. There are many examples in physics. For instance, a ball hung from a string whose upper end is clamped can be set swinging as a pendulum in either an east-west or a north-south direction, and there is no reason why both motions should not go on together. From simple reasoning, it seems probable that the resulting motion would in general be elliptical, and we shall show by analysis that this is true. Another good illustration is gotten by supposing one of the tuning-forks of figure 92 to be turned through 90° about a horizontal axis, so that its vibrations are in a vertical plane, the other still vibrating in a horizontal plane. The spot of light would then be subject to a vertical and a horizontal S. H. M., and would in general describe an ellipse.

We shall assume, as before, that the periods are the same. Let the horizontal and the vertical vibrations be represented respectively by

$$x = A.\cos\frac{2\pi t}{\tau} \qquad (8)$$

$$y = B.\cos\left(\frac{2\pi t}{\tau} - \beta\right) \qquad (9)$$

The latter equation can be rewritten in the form

$$y = B.\cos\beta.\cos\frac{2\pi t}{\tau} + B\sin\beta.\sin\frac{2\pi t}{\tau} \qquad (10)$$

To find the equation of the path, we must eliminate, between (8) and (9) or between (8) and (10), that variable which has no place in the simple equation of a line, viz., the time. This can best be done by deriving from (8)

$$\cos\frac{2\pi t}{\tau} = \frac{x}{A} \quad \text{and} \quad \sin\frac{2\pi t}{\tau} = \sqrt{1 - \frac{x^2}{A^2}}$$

and substituting these values in (10). This gives

$$y = \frac{Bx}{A} \cos\beta + B\sqrt{1 - \frac{x^2}{A^2}} \sin\beta$$

$$y - \frac{Bx}{A} \cos\beta = B\sqrt{1 - \frac{x^2}{A^2}} \sin\beta$$

By squaring both sides, we get

$$y^2 - \frac{2Bxy}{A} \cos\beta + \frac{B^2x^2}{A^2} \cos^2\beta = B^2 \sin^2\beta - \frac{B^2x^2}{A^2} \sin^2\beta$$

$$y^2 - \frac{2Bxy}{A} \cos\beta + \frac{B^2x^2}{A^2} = B^2 \sin^2\beta$$

$$\frac{x^2}{A^2} - \frac{2xy}{AB} \cos\beta + \frac{y^2}{B^2} = \sin^2\beta \qquad (11)$$

Equation (11), represents an ellipse, except when $\sin\beta = 0$, that is, when the phase difference β is either 0 or π. If $\beta = 0$, $\cos\beta = 1$, and the equation reduces to

$$\frac{x^2}{A^2} - \frac{2xy}{AB} + \frac{y^2}{B^2} = 0$$

or,

$$\frac{x}{A} - \frac{y}{B} = 0$$

the equation of a straight line whose slope is B/A. If $\beta = \pi$, $\cos\beta = -1$ and the equation reduces to

$$\frac{x}{A} + \frac{y}{B} = 0$$

a straight line whose slope is —B/A.

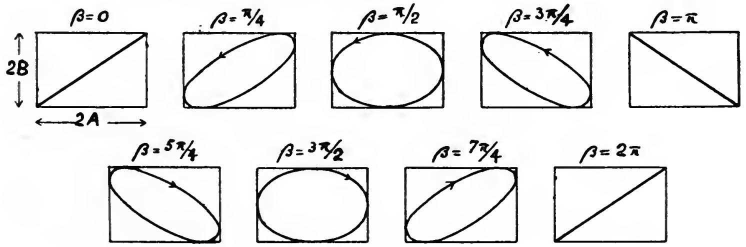

Figure 94

Figure 94 shows a number of typical shapes that the path can take according to the value of β. In each case, the whole motion must lie within the rectangle of the dimensions $2A \times 2B$, for evidently x cannot be greater than $+A$ nor less than $-A$, and y cannot be greater than $+B$ nor less than $-B$. Except in the cases where the ellipse becomes a straight line, the path is tangent to the sides of the rectangle. Arrows show the direction of motion in the path. The equation of the ellipse cannot tell us anything about the direction of motion, for the time has been eliminated, but it can be determined for each case by reasoning as in the following example: When $\beta = \pi/2$, or 90°, the vertical motion must be ¼ period behind the horizontal, that is, y must be zero but increasing when $x = +A$. A quarter period later, y will be equal to $+B$, and $x = 0$. Since, in a quarter period, the point moves from the right-hand extreme of the ellipse to the upper extreme, the rotation must be counterclockwise. When the ellipse reduces to a straight line, for β equal to zero or π, the motion is of course back and forth along a diagonal of the rectangle.

If the difference in phase is $\pi/2$ or $3\pi/2$, and if, in addition the amplitudes A and B are equal, the ellipse becomes a circle, in which the motion is counterclockwise or clockwise.

As a converse to what we have proved about the production of a harmonic elliptic motion as a result of the superposition of two linear S. H. M.'s at right-angles, it may evidently be said that any given elliptic harmonic vibration can be replaced for the purpose of mathematical analysis, by a pair of linear S. H. M.'s at right-angles which would produce this particular elliptic motion. Consideration of figure 94 shows that the analysis can be made in a number of ways. The directions of the two component vibrations that are to replace the ellipse may coincide with the major and minor elliptic axes or be inclined to them at any angle. In other words, there are many pairs of linear motions which are equivalent to a given elliptic motion, though the amplitudes and phase-difference are of course not always the same. In most cases, we wish to replace an elliptic motion by two linear motions in the directions of the elliptic axes, and then the amplitudes will be equal to the semi-axes of the ellipse, and the phase-difference will be $\pi/2$ or $3\pi/2$ according as the motion is counterclockwise or clockwise.

It is also plain, from the cases in the figure for which

$\beta = 0$ or π, that a linear S. H. M. can be resolved into two other mutually perpendicular linear S. H. M.'s either in the same or in opposite phases, and that the amplitudes of the components is gotten by exactly the same rule which holds when we resolve a force, or any other vector, into two mutually perpendicular components. For such a resolution then, amplitudes of S. H. M.'s may be treated as vectors. It is easy to see that the energy of the original vector will be the sum of the energies of the two components. For, since the amplitude of the former forms the hypotenuse of a right triangle, of which the component amplitudes are sides, the square of the former is equal to the sum of the squares of the latter, and we know that the energy is proportional to the square of the amplitude.

93. Lissajous figures.—A very curious and beautiful appearance is produced when two linear S. H. M.'s at right-angles, combining to form an elliptic motion, have periods which are nearly the same, but not quite. The difference in period causes one to gain upon the other in phase, and the motion passes through all the forms of figure 94, repeating the whole cycle again and again. The figures so formed are known as Lissajous figures. The following simple experiment, which anyone can perform, shows the cycle of changes very well in a slowly moving system. A ball R, figure 95, is hung from any support by cords in the manner shown. It is then capable of swinging as a pendulum of length RQ in the plane of the figure and as one of length RP in a perpendicular plane, and the two motions will of course have different periods. If it be started swinging in a direction inclined to both planes, it will slowly pass through all the configurations of figure 94.

Figure 95

Problems.

1. Calculate the energy of a body of mass 1000 grams in simple harmonic motion of period ½ sec. and amplitude 10 cm. What is the force acting on this body when it is 2 cm. from the center of its path?

2. Show that when two S. H. M.'s in the same straight line are applied to a body, each having the same amplitude and period, but with a difference in phase of 120°, the resultant motion has the same amplitude as either component.

3. Show that the maximum kinetic energy of a body in S. H. M. is equal to the kinetic energy of the tracing body (O of figure 89) if the latter has the same mass.

4. Show that when two S. H. M.'s at right-angles, with same amplitude and phase-difference of $\pi/2$, combine to produce a uniform circular motion, the energy of the latter is equal to twice that of either component.

5. Show that when two S. H. M.'s at right-angles, with a phase-difference of 0 or π, combine to produce a linear S. H. M. in another direction, the energy of the latter is the sum of those of the components.

6. Suppose that the two forks of figure 92 had not exactly the same period, say one had 100 complete vibrations per sec., the other 100.25. What would be the character of the vibration of the spot of light? Could the conclusions of paragraph 90 be modified to fit such a case?

7. A pendulum like that of figure 95 makes 100 complete vibrations per min. in one direction, 100.25 in the perpendicular direction. How long would it take to run through the cycle of changes portrayed in fig. 94?

CHAPTER XI.

94. Inverse square law.—Little has been said in this book so far about the brightness, or intensity, of light, except for the one point that it is proportional to the square of the amplitude. The subject is, however, of great importance, not only for practical illumination, but also for the details of experimental work.

Suppose that S, figure 96, is a mathematical point, emitting light at a steady rate uniformly in all directions. We also suppose that the medium exerts no absorption, which is absolutely true for the ether, and almost true for the air and other colorless gases, so far as visible rays are concerned. Consider two spheres, with S as center, of radii r_1 and r_2. Evidently the same amount of energy will pass in each second through each of these spherical surfaces, and therefore since the area of a sphere is directly proportional to the square of its radius, the passage of energy per second per square centimeter will be inversely proportional to the square of the radius. To put the matter in another form, suppose two white screens, each 1 square centimeter in area, are placed one 50cm. the other 100cm. from S, each being turned so that its face is perpendicular to the line joining its center to S. The nearer screen will receive four times as much light as the more distant one. and will therefore appear four times as brightly illuminated. Incidentally, the fact that the passage of energy per second varies inversely as the square of the distance shows that the *amplitude* of the light varies simply inversely as the distance.

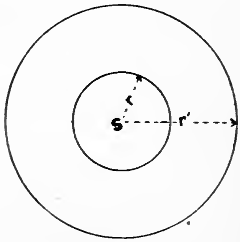

Figure 96

If, instead of a single point source, there are two points or more, or even an extended bright source like a flame, comprising a multiplicity of bright points, the inverse square law still holds true provided that the dimensions of the source are so small compared to its distance from the screen that for practical purposes we may say that all points of the former are equally distant from the latter. (This statement would not hold true if the vibrations in the different emitting centers had any special and constant phase-relations, for then interference would take place and the intensity of the radiation would be different in different directions. In an actual source such as a flame or an incandescent filament, the vibrations in different atoms or molecules are entirely independent, the phase-constants are distributed at random, and in addition the phase of any particular vibration is no doubt subject to sudden abrupt changes. Consequently interference in the ordinary sense cannot occur.)

95. Photometry.—Instead of considering the illumination of two screens at different distances, produced by the same source, let us now compare the illumination of the same screen by two different sources. By placing the brighter source farther from the screen than the fainter one, it is possible to make both sources produce the same illumination, and the relative brightness of the two sources can be expressed in terms of the two distances. The only practical difficulty lies in judging when the illumination produced by the two sources is the same. Various instruments known as photometers, a few of which will be described, have been devised for this purpose.

96. Rumford photometer.—The Rumford photometer, figure 97, is a very simple affair. S and S′ are the sources to be compared, R a cylindrical rod shown in cross-section, and ss a white screen. The part of the screen A is in the shadow from S, but is illuminated by S′, while the part B is in the shadow from S′ but is illuminated by S. The sources are moved nearer to the screen or farther away, keeping the edges of the shadows in contact, till A and B are equally bright. Let d and d′ then be the distances of S and S′ from the screen,

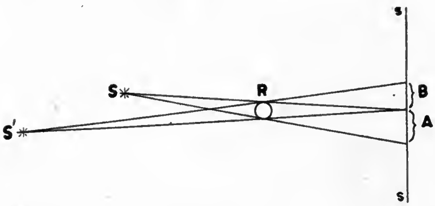

Figure 97

while L and L′ represent the brightness of the two sources. Since the illumination on the screen is the same from both

$$\frac{L}{d^2} = \frac{L'}{d'^2}$$

or

$$\frac{L}{L'} = \frac{d^2}{d'^2}$$

This photometer is unsatisfactory in practise, because it is very difficult to tell accurately when two illuminated surfaces are exactly of the same brightness unless they are exactly adjacent to one another, without any dividing area between. It is impossible to arrange the two shadows in this way unless the sources are very small. There is always at the edge of each shadow, a region of "penumbra," or half-shadow, where the screen is illuminated by part of one of the sources, but hidden from the rest, and the overlapping half-shadows from the two sources cause a region of unequal illumination between.

97. Bunsen photometer.—This is shown in figure 98. The screen ss is placed between the two sources S and S′, so that it is illuminated by them on opposite sides. A greased spot is in the center of the paper screen, and use is made of the fact that greased paper transmits more light, and reflects less, than ungreased. If no absorption occurred in either the clean paper or the greased portion, and if the latter scattered the light as effectively as the former, then the grease-spot would disappear when the illumination is the same on both sides. For, when viewed from the right, for instance, it would make up in light transmitted from S′ what deficiency there was in the light it reflected from S, and therefore would appear just as strongly illuminated as the clean paper surrounding it. Unfortunately, there is absorption, and unequal absorption, in the two portions, and the greased portion does not scatter as well as the ungreased, but acts more like a transparent medium, so that the spot may disappear when viewed at a certain angle, but

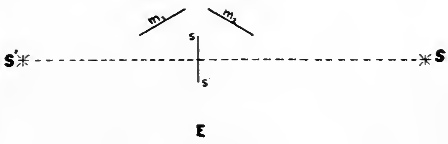

Figure 98

not when viewed at another. For this reason, it is necessary to view both sides of the paper at once, at the same angle. A pair of mirrors are set in an inclined position in such a way that the observer stationed at E sees both surfaces of the screen reflected in the mirrors. The screen and mirrors, which are rigidly connected together, are then moved to right or to left between the two sources until the spot appears equally conspicuous in the two mirrors. The brightnesses of the sources are then directly proportional to the squares of the distances from the screen. Only a small degree of accuracy is obtainable with this instrument.

98. Lummer-Brodhun photometer.—All accurate comparisons of the intensities of artificial light-sources are made with the Lummer-Brodhun photometer, shown in figure 99, or some modification of it. S and S′ are the sources to be compared, and s a two-faced screen, made of some white and efficiently diffusing material, such as fresh plaster of Paris, or a fine quality of milk-glass. A and B are two right-angled glass prisms, so set as to produce, by total reflection at the hypotenuse, reflected images of the two screen-faces. P and Q are another pair of right-angled prisms, with their hypotenuse faces cemented together by means of Canada balsam, after part of the face of P has been ground away, leaving only a circular spot in the center where actual contact occurs. T is a short-focus telescope, focussed upon the cemented faces of P and Q. That part of the light from the right-hand surface of s which, after reflection at B, strikes the cemented area where P and Q come together, passes through without reflection and never reaches the telescope, but that part which strikes the edges is totally reflected into the telescope. Consequently, if the source S were cut off and S′ alone functioning, the telescope would show an illuminated area with a completely dark circular hole in it. On the other hand, of the light coming

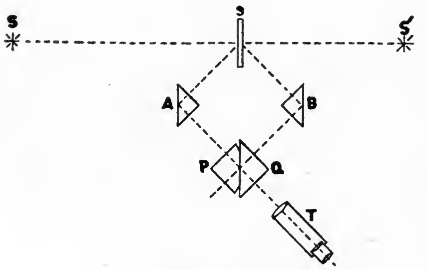

Figure 99

from the left-hand surface of s, illuminated by S, only that part which strikes the cemented area passes through to the telescope. The field of view then is a circle illuminated by S surrounded by an area illuminated by S′, and since these come accurately edge to edge it is quite easy to tell when the illumination is the same. The screen, the prisms and the telescope are mounted together in a frame which can be moved to right or left till the inner circle vanishes against the equally illuminated area surrounding it. Very accurate measurements can be made with this instrument when the lights to be compared have the same color. If they differ much in color an accurate comparison is impossible, for from the nature of things one cannot make accurate quantitative comparisons of things which differ in quality. Most artificial light-sources, however, are near enough the same color so that fairly trustworthy comparisons can be made.

99. Light-standards.—In measuring the brightness of a source, some unit is necessary. The original unit was the light from a spermaceti candle, of a specified size and burning at a specified rate, and lights are still rated in "candle-power." Actual candles, however, even when made and used according to specifications, are quite variable, and in practice a more constant standard is used, such as the Vernon-Harcourt lamp, taken as having 10 candle-power, or the Hefner lamp, .9 candle-power. Laboratory tests are usually made with an incandescent lamp whose intensity has been standardized at the Bureau of Standards or some similar laboratory by comparison with a Vernon-Harcourt or Hefner lamp.

The "foot-candle" is a unit of measurement, not for the intensity of a source, but for the degree of illumination on a surface. It is the illumination produced at a distance of one foot from a source of unit candle-power. The degree of illumination on a desk 8 feet away from a 30 candle-power lamp is $30/(8)^2 = 30/64 = .453$ foot-candles.

100. Solid angle.—In order to discuss intelligently the brightness of extended surfaces, optical images, etc., it is necessary to define a term called "solid angle." Suppose a cone of rays, of any cross-sectional shape, to emanate from a point S, figure 100. The solid-angle of this cone is the area that it cuts out on the surface of a sphere of unit radius, or the area that it cuts out on *any* sphere with S as center divided by the square of the radius of the sphere. Since the total area of a sphere is $4\pi r^2$ the total of all solid-angles from a point is 4π.

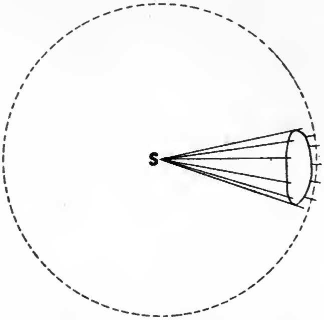

Figure 100

101. Intrinsic luminosity.—Suppose we have an extended surface emitting light, such as the surface of molten metal, or of the pole of a carbon arc. Consider the light emitted, within a solid-angle ω, from a small area a, so small that we may regard it all as lying at the apex of the solid-angle. This amount of light will evidently be proportional to the size of the solid-angle ω, and therefore may be written as equal to

$$Ia\omega$$

The factor of proportionality, 1, evidently is an indicator of the brightness of the surface, without regard to its size, and it is given the name "intrinsic luminosity." It may be defined in words as the amount of light emitted per unit area per unit solid-angle.

We shall now prove that, when a real image is formed by a lens, the intrinsic luminosity of the image is the same as that of the object, except for losses of light due to reflection, absorption, etc. In figure 101 let a be any very small area on the object O and b the corresponding area on the image I. The

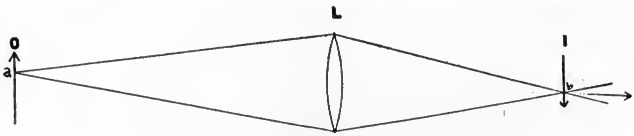

Figure 101

amount of light from a going to form b, apart from the above-mentioned losses, is the amount lying within the solid-angle of the cone which enters the lens. The value of this solid-angle is

$$\omega = \frac{A}{u^2}$$

where A stands for the area of the lens and u for the distance of the object from it. If I is the intrinsic luminosity at a, the amount of light from a entering the lens is

$$I\omega a = \frac{IAa}{u^2}$$

Let I′ be the intrinsic luminosity of the image at b, v the distance of the image from the lens, and ω′ the solid-angle of the cone which goes to form b and the equal cone which emerges from b on the right, then

$$\omega' = \frac{A}{v^2}$$

and the amount of light going to form b is

$$I'\omega' b = \frac{I'Ab}{v^2}$$

If no light were lost, we could then write

$$\frac{IAa}{u^2} = \frac{I'Ab}{v^2}$$

The dimensions of any part of an image are to those of the corresponding part of the object as v is to u, and therefore the *areas* will be in the ratio v^2 to u^2, so that

$$\frac{a}{u^2} = \frac{b}{v^2}$$

Consequently

$$I = I'$$

This conclusion is at first sight surprising, for it seems to indicate—first that the brightness of the image per unit area is independent of the distance of the object,—second that it is independent of the diameter of the lens. Distance of the object does indeed have no effect, for although the lens receives less light when the object is farther away, that light is distributed over an image whose area is smaller in the same proportion.

As to the effect of the diameter of the lens, one must remember that intrinsic luminosity is defined for unit solid-angle, as well as for unit area. A large lens does in fact send more light to form the image, but since the solid-angle is increased in the same proportion the intrinsic luminosity remains the same. If the image is cast on a photographic plate, where the photographic action depends only upon the total amount of light received per unit area of plate, entirely without regard to the size of the light-cone, it is of course advantageous to use a lens of large diameter. The same is true if the image is cast on a diffusing screen, like white paper or plaster of Paris, for such materials do not allow the light to remain in a cone of the same diameter as that in which it comes to the screen, but re-emit it in all directions. In either of the above cases, the lens should be as large in diameter as is consistent with moderate freedom from spherical aberration, astigmatism, etc., if a bright image is desired.

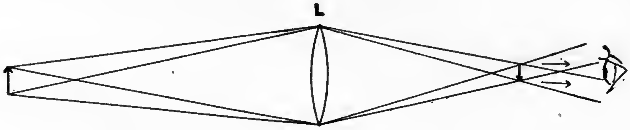

Figure 102

But when an image is viewed directly by the eye, as in figure 102, the amount of light entering the eye is limited by the size of the pupil, and all the light which fails to enter the pupil is wasted, so far as vision is concerned. Consequently, the brightness of the final image upon the retina is not affected by an increase or decrease in the diameter of the lens, so long as the cone of light remains large enough to completely fill the pupil. If the pupil were not completely filled, the intrinsic brightness of the image on the retina would still be the same, but the mental impression of brightness would be reduced. For this impression, like photographic action, depends upon the amount of light received without regard to the size of the cone of light.

In spite of all that has been said, there is a great advantage in the use of large-diameter lenses as the objectives for telescope. In the first place, a large-diameter objective permits the use of a very short-focus eyepiece, and therefore a high magnifying-power, without reducing the emerging cone of light so much that it does not fill the pupil. In the second place, we have seen that the image of a mathematical point is not a point, but a small disc surrounded by series of fainter concentric rings (see sections 72 and 73), and that the diameter of disc and rings become smaller when the diameter of the objective becomes larger. A fixed star (what follows does not apply to planets) may be many times larger than the sun, but its distance is so great that in every case what we call its "geometrical" image —the image as it would be if there were no such thing as diffraction—is hardly more than a point, being much smaller than the central disc of the diffraction pattern. We are therefore obliged to regard a star as equivalent to a point-source, whose image is the diffraction-pattern itself. Since the diameter of the central diffraction disc is inversely proportional to the diameter of the objective, it is clear that the star-image will have a much greater intrinsic luminosity with a large than with a small lens. On the other hand, when an object is large enough or close enough so that its geometrical image has perceptible size, although each point of the image is made up of diffraction-disc and rings, the only effect of these is to extend the edges of the image very slightly indeed, and the effect of diffraction upon the area of the total image is relatively negligible. Suppose that a telescope be pointed toward a star in daylight. Since the diameter of the objective exceeds greatly that of the pupil of the eye, the intrinsic luminosity of the star is much greater than when seen with the naked eye; but that of the sky, an extended area, is not increased in the least degree. Consequently, it is possible to see with a large telescope stars totally invisible to the naked eye, even in full daylight.

The principle of equality of image and object so far as intrinsic luminosity is concerned, proved above (with reservations as to loss of light by reflection and absorption) for a real image formed by a lens, can be just as easily proved for a virtual image, or for an image formed by a mirror. The only advantage in using opera-glasses, reading glasses, microscopes, etc., so far as extended objects are concerned, is the magnification in size. The brightness is never increased, but rather somewhat diminished by the unavoidable losses.

It is easy to prove also that, except for atmospheric absorption, an object will appear with the same intrinsic brightness at any distance, provided the pupil does not dilate or contract. For in figure 101 the lens L can just as well represent the lens of the eye as any other lens, with the image I formed on the retina. The proof then follows word for word as given above.

102. Spectrophotometer.—It has already been mentioned that accurate comparisons of the brilliancy, or candle-power, of light-sources can be made only when they have about the same color. To compare the intensities of a reddish and a bluish light would be much like comparing the intensities of a deep bass musical note and a shrill treble one. Precise measurements can never be made except when the objects compared, though differing in quantity, are the same in quality. It is true that a special form of instrument, known as the "flicker-photometer," has been used with some success in comparing lights somewhat divergent in color, but the indications given by it are now considered unreliable, and their interpretation is in any case somewhat doubtful.

The only completely satisfactory method for such a problem is to form the spectra of the two sources side by side and then make a comparison, color by color, for a number of different parts of the spectrum. Several devices for accomplishing this, known as "spectrophotometers" are now in use. The results of the comparison must be given in the form of a table, or better still a curve, with wavelength as abscissa and the corresponding ratio of brilliancy in the two spectra as ordinate. Thus a curve, such as that in figure 103 would indicate that the source A is relatively stronger than source B in the middle of the spectrum, about the green.

Figure 103

Problems.

1. What is the value of the solid angle included between two walls and the floor of a room?

2. Find the illumination on a desk 10 ft. from a 60 candle-power lamp. Suppose the desk were inclined in such a way that a perpendicular to it made an angle A with a line drawn from it to the lamp. What would then be the illumination?

3. Given that the illumination of the earth in full moonlight is .02 foot-candles, and that the distance of the moon is 235000 miles, find its candle-power.

CHAPTER XII.

103. Transverse and longitudinal waves.—104. Double refraction.—105. Polarization of the O and E light.—106. Wave-surface in doubly-refracting crystals.—107. The lateral displacement of the E ray.—108. Special cases of double refraction.—109. Tourmaline.—110. Biaxial crystals.—111. Polarization by reflection.

103. Transverse and longitudinal waves.—We have already spoken of the distinction between longitudinal and transverse waves, but have left unsettled the question to which of these types light-waves belong. A decision has so far been unnecessary, because everything we have said about light up to this point would apply equally well whether the waves were longitudinal or transverse. For instance, reflection, refraction, interference, diffraction, etc., could occur equally well with either.

It is manifestly impossible to *see* a light-wave, in the sense that we see water-waves, or waves in a string. We see by means of light, but we do not see light itself. Therefore the character of the waves must be determined by indirect, rather than direct, means. The most important distinction between the two types is this, that a train of longitudinal waves is completely specified when we have stated its wavelength, amplitude, phase, and direction of propagation, while a train of transverse waves is not. To explain more fully, imagine a beam of monochromatic light travelling from north to south, whose amplitude, wavelength and phase are known. If the waves are longitudinal, nothing more can be said about this beam, and any other beam, travelling in the same direction, with the same amplitude, wavelength and phase, would be indistinguishable from it. On the other hand, if the waves are transverse, there is still a possibility that the two beams should be quite different. For instance, the vibrations in the first might be up and down, while those in the second were east and west.

Evidently then, if light waves are transverse, there should exist some distinguishing characteristic of a beam of light, other than wavelength, amplitude, phase, and direction of propagation, and having to do with a direction at right angles to the direction of propagation, if we could only find the suitable means of detection. But we should hardly expect ordinary natural light, such as comes directly from the sun or a flame, to show this characteristic even if it exists; for there would be no reason why in such light one direction of vibration would predominate over another. It would be much more probable that all possible directions of vibration would be represented to about the same extent, so that the beam would not show any peculiar properties in any direction. Consequently, in seeking such a characteristic as we have mentioned as possible, which would prove light-waves to be transverse, we ought to experiment, not with light coming directly from any original source, but rather with light which has suffered reflection, refraction, or some other action which might cause one plane through the direction of propagation—for instance the plane of incidence in the case of reflection—to have a particular importance for the beam.

104. Double refraction.—In 1690 Huyghens discovered a characteristic of light, such as we have been speaking about, which we regard as a definite proof that light-waves are transverse. Huyghens himself did not draw this conclusion, for at that time it had always been assumed without question that the waves were longitudinal, like those of sound, and the possibility of their being transverse had never been suggested. Huyghens' discovery was concerned with the phenomenon known as "double refraction," which is shown by many crystals, to the most remarkable extent by the crystal known as calcite, or Iceland spar. Before Huyghens' experiment can be understood, it will be necessary to explain some of the facts about crystal structure and the nature of double refraction.

Every crystal possesses certain regularities of structure, on account of which it can be easily split in certain directions. Calcite, for instance, can be readily split into any one of the forms shown in figure 104, which differ in linear dimensions, but have exactly the same angles. The important thing about crystal structure is that the angles are fixed and invariable, while the dimensions of the faces may be anything. In any one of the three rhombohedral forms shown in the figure for calcite, each face has two angles of 101° 55′ and two of 78° 5′. In each rhombohedron, there are two opposite corners where three obtuse angles come together, as at A. A line drawn through one of these corners, equally inclined to the three faces that meet there, *or any line parallel to it,* is called the *optic axis.* Note that the optic axis is defined only by its direction, and any line having that direction may be called the optic axis.

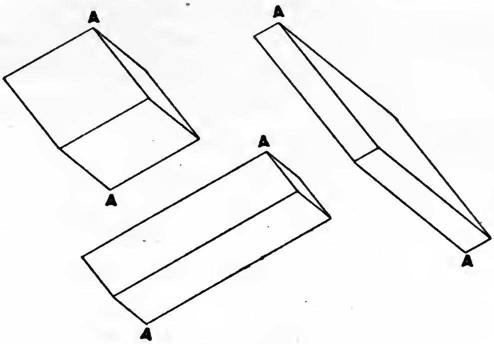

Figure 104

When a pencil of light enters such a rhomb of calcite it is separated into two parts, so that there are two refracted beams, instead of only one as in the case of glass and other non-crystalline media. One of the beams obeys the ordinary laws of refraction, the ratio of the sines of the angles of incidence and refraction being always the same, no matter what the angle of incidence may be. It is therefore called the "ordinary" ray, and we shall represent it by the letter O, for brevity. The other does not follow the ordinary laws of refraction. The ratio of the sines changes as the angle of incidence changes, showing that this beam travels through the calcite *with different velocities in different directions,* this being due, no doubt, to the regular structure of the crystal. For this reason it is called the "extraordinary" ray and we shall represent it by E. If we wish to speak of an *index of refraction* for the extraordinary ray, it must be understood that that index is not a constant for any given wavelength, but is variable with the angle of incidence. Even when the incidence is normal, when there should be no refraction according to the ordinary law, the E-ray is in general refracted, as illustrated in figure 105. ABCD is the sectional outline of a crystal of calcite. The incident ray PQ, even when it strikes the surface normally as

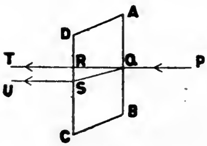

Figure 105

in the figure, is divided into two rays, one of which PQRT (the O-ray) goes straight through without bending, while the other, PQSU (the E-ray), is deflected upon entering the calcite and emerges from it parallel to the incident ray but with a lateral displacement. Consequently, if an object close to the rhomb be viewed through the latter, two images will be seen. If the object is very far away, however, only a single image will be seen; for the two emergent rays are parallel, with only a small relative displacement, and for an object very far away a small displacement does not change its apparent position by an appreciable amount. Therefore the two images coincide and appear as a single image. Another explanation is as follows: If the object is far away, the incident waves are practically plane, and therefore the two emergent beams have waves that are practically plane, and parallel to one another, and we have already found that all plane and parallel waves are focussed at the same point on the retina. Conversely, we can regard the fact that only a single image of a distant object can be seen through a block of calcite as proof that the two emergent rays coming from a single incident ray are parallel to one another.

The two images of a near object seen through calcite, formed respectively by the O-ray and the E-ray, are called the ordinary and the extraordinary images. If the crystal is turned about the incident ray as an axis, the O-image remains still and the E-image revolves about it. Let us define a plane which is parallel to the optic axis and perpendicular to the face through which the light enters as the *principal plane* of that face. A line joining the centers of the two images lies in the principal plane. The images are equally bright, and the light from them seems exactly the same to the eye. A superficial examination reveals no difference between them except for the displacement of the E-image in the principal plane, and the further fact that it seems slightly farther away than the O-image. But note that the two refracted beams have undergone a process (transmission through a crystal) which has caused a particular plane (the principal plane) to have a peculiar importance for both. Consequently, in accordance with our former discussion, if the waves are transverse we are more likely to get evidence for it in these two beams than in the original incident beam, where one direction of vibration has no reason to be preferred over another.

105. Polarization of the O and E light.—Such evidence is furnished in the experiment of Huyghens, mentioned above, which consisted in viewing an object through *two* crystals of calcite. If the crystals are equally thick and are similarly placed so that their optic axes are parallel, as in figure 106, the phenomenon observed is just the same as with only one, except that the displacement between the images is doubled. For the two are equivalent to a single crystal of double thickness. But if one crystal is turned about the incident ray as axis, while the other is held still, there are in general four images, though for certain positions of the moving crystal two of them disappear, and in one position only one is visible. The production of four images is not surprising, for one should expect that the

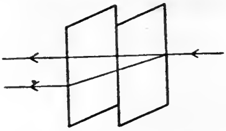

Figure 106

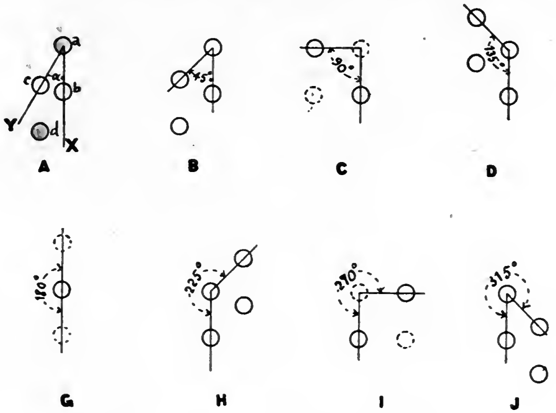

Figure 107

second crystal would divide each of the two beams that enter it into two more, but the changes in intensity that the images undergo when the second crystal is rotated, and the disappearance of some of them in certain positions, require explanation. These changes should be understood from the series of eight diagrams in figure 107. In all of them the plane of the paper is supposed perpendicular to the incident ray, and the little circles show the positions of the images. In diagram A, aX represents the intersection of the principal plane for the first, or stationary, crystal with the plane of the paper, aY that for the second, or moving, crystal, the two making in this diagram an angle a which is less than 45°. Four images are seen, of which a and d are brighter than b and c. If the angle a be increased, a and d become fainter and b and c brighter, till when $a = 45°$, as in diagram B, all four are equally bright. If a is still further increased, a and d continue to grow dimmer, b and c brighter, and when $a = 90°$ (C) a and d disappear and b and c attain maximum brightness. With a further increase in a, a and d reappear and grow brighter, while b and c diminish, till when $a = 135°$ they are all equally bright, and when $a = 180°$ a and d have reached maximum brilliancy, b and c vanished completely. Meanwhile, a and b have maintained their positions, but c has swung about a as a center, and d about b, so that in the 180° position (G) the only remaining images, a and d, fall in the same place and appear as a single image. Similar changes go on if a is increased beyond 180°. H shows the appearance when $a = 225°$ (the four images equally bright), I that when $a = 270$ (a and d vanish, b and c very bright), and J when $a = 275°$ (all equally bright). If $a = 360$ it is the same as if $a = 0$, a and d would have maximum brightness and be at the maximum distance apart, while b and c would be gone.

These phenomena can all be well explained if we assume —first, the transverse nature of light waves:—second, a certain property of crystals in regard to the velocity of light. A non-crystalline material like glass has properties which are the same in any direction, and light therefore travels through it with the same velocity, no matter what may be the direction of propagation or the plane in which the vibrations occur. Crystals, on the other hand, have decidedly different properties in different directions. Not only do they split easily into layers in certain definite planes, but also such properties as heat-conductivity and the elastic and electrical constants are different according to the direction in which they are measured. Consequently, taking the optic axis as an axis of symmetry, it is likely enough that waves whose vibrations lie in the principal plane would be transmitted with a different velocity from those with vibrations perpendicular to that plane, and this is the second assumption mentioned at the head of this paragraph. Then, if a beam of natural light, which presumably has vibrations equally in all directions, falls upon a crystal, the latter will automatically resolve it into two component beams, the vibrations of which are in one parallel, in the other perpendicular, to the principal plane. These will become separated from one another precisely because one travels faster than the other, so that when they emerge from the crystal we shall have instead of a single beam in which all directions of vibrations are equally represented, two beams in each of which the vibrations are confined to a plane, these planes for the two beams being, however, mutually perpendicular. Light restricted to a single plane of vibration is said to be *polarized,* or more specifically, *plane polarized,* so that both the O and E beams are polarized, but not in the same plane. Both beams are equally bright because the incident light (unpolarized) had presumably no excess of vibrations in any particular plane. Measurements of refraction show that the E-light travels faster in the calcite than the O-light, but there is no obvious reason to decide whether vibrations in the principal plane or perpendicular thereto should be propagated the faster. Consequently we are at a loss to know whether the O-light has its vibrations in the principal plane and the E-light perpendicular thereto, or vice versa. We avoid the dilemma by simply saying that the O-light is *polarized in the principal plane,* the E-light *polarized perpendicular to the principal plane.* These statements should be regarded as a definition of the plane of polarization, leaving as a matter for future discussion the question whether the plan of polarization coincides with the plane of the vibrations, or is perpendicular to it.

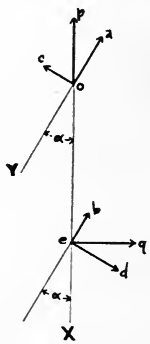

Figure 108

Now, suppose that our two equally bright polarized beams, obtained by passage through a single crystal, strike a second crystal whose principal plane makes an angle a with that of the first. In figure 108, oX represents the position of the principal plane for the first crystal, oY that for the second. Two equal vectors are laid off, op representing the amplitude of the O-ray as it comes from the first crystal, eq that of the E-ray. The vectors are drawn in the directions of the respective planes of polarization, because we are not certain as to the actual directions of vibration. Now light whose plane of polarization is either parallel or perpendicular to oX cannot as such pass through the second crystal, but only light whose plane of polarization is parallel or perpendicular to oY. We have seen, however, in section 92, that a simple harmonic motion of amplitude op is equivalent to two other simple harmonic motions of the same period and phase, having amplitudes oa and oc, gotten by the rules of vector resolution. Therefore, the O-beam from the first crystal will be automatically broken up by the second, and will pass through the latter as a beam of amplitude oa, polarized in the principal plane, and a beam of amplitude oc, polarized perpendicular to the principal plane. The former of these forms the image a of figure 107, the latter the image c. In quite the same way the E-beam from the first crystal, represented by the vector eq, will, on entering the second crystal, give rise to a beam of amplitude eb, polarized in the principal plane and responsible for the image b of figure 107, and one of amplitude ed, polarized perpendicular to the principal plane and responsible for the image d of figure 107.

The relative intensities of the four images a, b, c, and d of figure 107 can be learned from the amplitudes of the vectors oa, oc, eb and ed, figure 108. If A represents the amplitude op or oq (representing the equal O and E beams coming from the first crystal) then

$$oa = ed = A.\cos\alpha \quad \text{and} \quad oc = eb = A.\sin\alpha$$

Since the intensity of a beam of light is proportional to the square of the amplitude, the intensities of the images a, b, c, and d of figure 107 are then in the continued proportion

$$a : b : c : d = \cos^2\alpha : \sin^2\alpha : \sin^2\alpha : \cos^2\alpha$$

Therefore, when $\alpha = 90°$ or $270°$, a and d vanish while b and c have maximum brightness, and conversely when $\alpha = 0$ or $180°$, and all are equally bright when $\alpha = 45°$, $135°$, $225°$, or $315°$. These being exactly the relations of brightness which experiment shows to exist, we may regard it as definitely proved that light-waves are transverse, and that certain crystals have the power to divide up an unpolarized beam of light into two component beams polarized at right-angles to one another.

106. Wave-surface in doubly-refracting crystals.—The fact that the ordinary ray obeys the regular laws of refraction shows that light polarized in the principal plane travels with the same velocity for all directions of propagation, that is, for all directions of the ray. On the other hand, the failure of the extraordinary ray to follow these same laws shows just as surely that light polarized perpendicular to the principal plane does *not* have the same velocity for different directions of the ray. Careful measurements of the angles of refraction for various angles of incidence show that velocity varies with ray-direction in a manner that can best be explained as follows:

Suppose that, at some point in the interior of a block of glass, a disturbance of the molecules takes place, of the kind which produces light. The light would pass out from this point in wavefronts each of which would be a perfect sphere, for the velocity would be the same for all directions. If the same sort of disturbance were to occur in the interior of a block of *calcite,* each wavefront would no longer be a sphere simply, but a double surface consisting of a sphere and an ellipsoid of revolution. The optic axis through the center is the axis of the ellipsoid, and ellipsoid and sphere are tangent to one another where this axis pierces both surfaces. The spherical part is the wavefront for vibrations whose plane of polarization is a *radial* plane, a plane containing ray and optic axis. The ellipsoid is the wavefront for vibrations polarized perpendicular to the radial planes. Therefore the sphere accounts for the ordinary and the ellipsoid for the extraordinary light. In calcite, the ellipsoid is oblate, and lies outside the sphere, and therefore the E-wave travels faster than the O-wave, but in some crystals it is prolate and lies inside, and the O-wave travels the faster. In the direction of the optic axis, both waves travel with the same velocity, and double refraction fails. Only a single image can be seen from light passing through the crystal parallel to the optic axis.

107. The lateral displacement of the E-ray.—We can now explain why the extraordinary ray receives a lateral displacement in going through a block of calcite (not parallel to the optic axis) even when the angle of incidence is zero. Let AB, figure 109, represent a plane surface separating a block of calcite (left) from air (right), and let XY be a series of wavefronts advancing in the direction of the arrow, perpendicular to the surface. The optic axis is parallel to Oc, in the plane of the paper. According to Huyghen's principle, each point in the surface will, when a wavefront strikes it, become the center of wavelets that will advance into the calcite, and the tangent to these wavelets constitutes the wavefront in the crystal. Wavelets are drawn in the figure from three centers, a, b, and c, and each shows in the diagram as a semi-circle and a semi-ellipse, these being the sections of the wavesurface with the plane of the paper. If the incident wavefront consisted only of light polarized in the plane of the paper, the semi-ellipse would be omitted, for the plane of the paper is a radial plane for the wavesurface, and there would be no extraordinary wave. If the incident light were polarized perpendicular to the plane of the paper, the semi-circle would be omitted, and there would be no ordinary wave. But if the incident light is either unpolarized, or polarized in any other plane, both circle and ellipse are represented, and both rays exist. The refracted wavefront will then be double, consisting of a plane ukv, tangent to the circles and a plane slt tangent to the ellipses. These two are parallel to one another, but the E-wavefront (slt) has advanced farther than the O-wavefront (ukv) and has received a lateral shift, which increases as the wavelets progress, until the second face of the calcite is reached. Then we can again apply Huyghens' principle to find the wavefronts refracted out into the air, remembering that here all waves travel with the same velocity, whatever may be the plane of polarization. Of course, the terms "ordinary" and "extraordinary" apply only to light within a crystal or other doubly-refracting material. Evidently, the ordinary wavefronts will emerge into the air as shown at UV and the extraordinary as at ST. The latter will have gained distance and suffered a permanent lateral displacement, but neither the distance gained nor the lateral displacement increases any more.

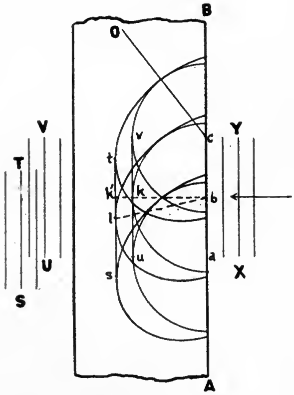

Figure 109

In the ordinary light, what we call the "ray," or the line along which the light advances, is perpendicular to the wavefront, because, the secondary wavelets being spherical, the line drawn from a center of a secondary wavelet to the point of tangency with the resulting envelope, bk, for instance, is necessarily perpendicular to the envelope. But in the extraordinary light this is not necessarily so, because the secondary wavelets are ellipsoidal. For instance, the line bl, drawn from b to the point where the ellipse from b touches the extraordinary wavefront slt, is not perpendicular to the latter, and the wavefront advances, not perpendicular to itself but in an inclined direction. We are accustomed to speak of two velocities for the extraordinary light, the ray-velocity and the wave-velocity, which are in the ratio of bl to bk′. For the ordinary light ray-velocity and wave-velocity are the same.

108. Special cases of double refraction.—If the incident light, instead of striking normally upon the surface of the calcite as in figure 109, strikes it obliquely, the two wavefronts in the crystal would be found by the method of Huyghens in a quite analogous way. Figure 110 represents such a case. The wavefront XY coming from the air side (right) strikes the calcite surface AB first at Y. While it is advancing to u through air, the point Y sends out a spherical wavefront whose radius Yv bears the same ratio to Xu as the velocity of the ordinary waves to the velocity in air, and also the corresponding ellipsoidal wavefront as shown. At later instants the points b and a take up the role of secondary centers and send out spherical and ellipsoidal wavefronts of correspondingly smaller dimensions, and so for every point between Y and u. The common tangent plane uv to all the spherical wavelets is the ordinary wavefront in the crystal, and the common tangent plane ut to all the ellipsoidal wavelets is the extraordinary.

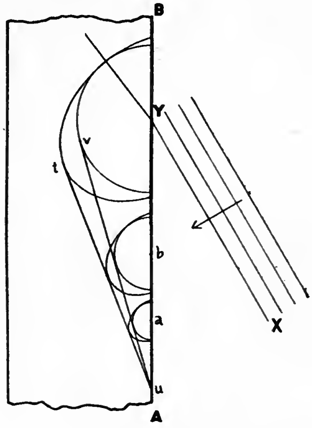

Figure 110

Here, as we see, the O-waves are not parallel to the E-waves within the crystal.

It is possible to cut or grind a piece of calcite in such a way that the optic axis is parallel to the surface, although the crystal will not naturally split in this way. The part of figure 111 to the left of AB represents calcite cut in this way, AB being the outer surface and the optic axis being in the plane of the paper parallel to AB. The Huyghens construction is shown for the case when a train of plane waves XY falls at normal incidence upon the surface. In this case there is no lateral displacement of the ordinary waves, and in a sense one may say there is no refraction, but the extraordinary waves travel faster than the ordinary, so that if they both emerge into the air through a second surface parallel to AB the E-waves will be ahead of the O-waves by an amount which depends upon the thickness of the layer of calcite. If the layer is thick enough so that one wave gets a quarter of a wavelength ahead of the other, it is called a *quarter-wave plate,* if thick enough so that one gets a half-wavelength ahead of the other it is a *half-wave* plate, etc.

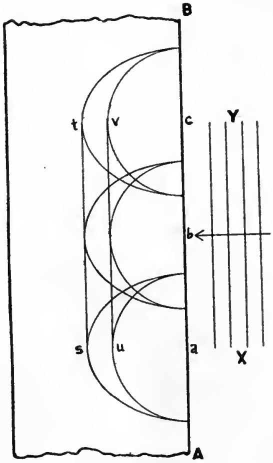

Figure 111

109. Tourmaline.—Certain doubly-refracting crystals have the peculiar property of absorbing one of the rays very much more than the other. The best-known of these is tourmaline, which absorbs the ordinary ray so strongly that two or three millimeters of the crystal practically extinguish it. The extraordinary ray is transmitted with little absorption, and therefore tourmaline is one of our means of getting a beam of completely polarized light. With two plates of tourmaline a very interesting phenomenon can be shown, whose explanation is simple in terms of transverse waves, but would otherwise probably be impossible. If the plates be held together with their principal planes parallel, and any source of light be observed through them, the light can be plainly seen, though less than half as bright as when seen directly, for of course half the light is absorbed as ordinary waves, and there is also some general absorption and loss by reflection. But if one of the plates be turned through 90° about the beam of light as axis, nothing at all can be seen through them. The light transmitted through the first plate as the extraordinary waves has its plane of polarization in such a direction that it is ordinary light for the second plate, which therefore absorbs it completely. In this position the tourmalines are said to be *crossed*.

110. Biaxial crystals.—A great many crystals show double refraction in a way different from calcite, having two directions instead of one along which light can be transmitted without showing double images. They are called *biaxial crystals*, and both rays are extraordinary, following laws which are, except for certain special directions, more complicated than the simple laws of refraction that hold for non-crystalline media. The complete wavesurface consists, not in a sphere and an ellipsoid, but in a very complex surface of two sheets. Figure 112 is a perspective view of a plaster model showing one fourth of the complete surface, the equation of which is

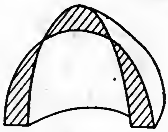

Figure 112

$$\frac{x^2}{r^2 - a^2} + \frac{y^2}{r^2 - b^2} + \frac{z^2}{r^2 - c^2} = 1$$

where $r^2 = x^2 + y^2 + z^2$, and a, b and c are certain constants, having different values for different crystals. The equation gives the wavefront emitted from a point, at a definite time, say 1 second after emission. Perhaps a better idea of the shape of the surface can be obtained if, after studying figure 112, one considers the sections made by the three coordinate planes. The section with the YZ plane consists of a circle of radius a and an ellipse of semi-axes b and c; that with the ZX plane is a circle of radius b and an ellipse of semi-axes c and a; and that with the XY plane is a circle of radius c and an ellipse of semi-axes a and b. These sections are shown in order in figure 113, which is drawn on the assumption that a is greater than b, and b than c. If it should happen that any two of these quantities are equal, the crystal would become uniaxial. If all three are equal, it would not be doubly-refracting at all. This is the case for rocksalt and some other crystals, which act, so far as the transmission of light is concerned, like glass and other isotropic media.

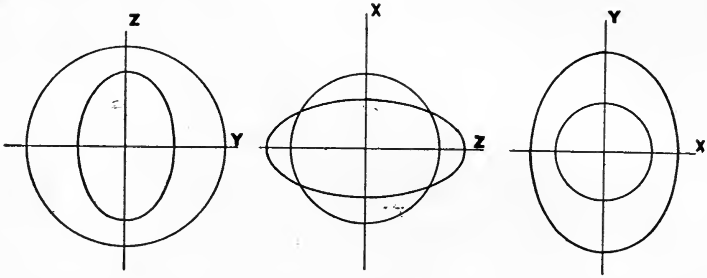

Figure 113

111. Polarization by reflection.—In 1810 Malus discovered quite by accident that the light reflected from glass is in general partly polarized. In looking through a plate of tourmaline at a glass window from which the direct light of the sun was reflected, he found that when the tourmaline was turned in a certain way the light became much dimmer. In fact, light is polarized to a greater or less degree when it is reflected from any non-metallic surface, except when the angle of incidence is zero. Observation, by means of a tourmaline or any equivalent instrument, of the light reflected from a pond of water, a slate roof, or a wet cement sidewalk, will show always a certain degree of polarization, which is an indication that the vibrations in a certain plane are stronger than in a plane at right-angles. This statement applies, however, only to light that is regularly, not diffusely reflected. Many materials, such as paper, which diffuse strongly, also reflect regularly to some extent, particularly paper which has a strong glaze. The regularly reflected light is subject to polarization by the act of reflection, but the diffusely reflected light is not. In fact, the degree of polarization of the reflected light has been used as a means of grading papers in regard to glaze.

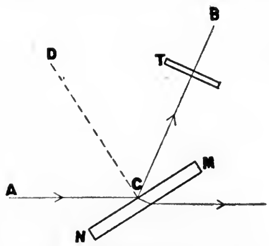

Figure 114

Figure 114 illustrates the polarization by reflection from glass. MN is a glass slab, T a plate of tourmaline, in such position that its principal plane is perpendicular to the plane of incidence of the light on MN, i. e., to the plane of the paper in the figure. Under these circumstances, the reflected light passes through the tourmaline, but if the latter is turned about the reflected ray CB as axis, till its principal plane is parallel to the plane of incidence, it cuts off most or all of the reflected light, depending upon the value of the angle of incidence ACD. There is a certain value of this angle, about 57° for the common varieties of glass, for which practically all the light is cut out by the tourmaline. Since this light is transmitted by the tourmaline when the plane of incidence is perpendicular to the principal plane, and since tourmaline transmits its extraordinary light, whose plane of polarization is perpendicular to the principal plane, therefore the light reflected from MN is polarized in the plane of incidence. The angle of incidence for which polarization is most nearly complete is called the *angle of polarization.* If the reflecting surface is thoroughly and freshly polished, the polarization is almost perfect at the angle of polarization, but an old or soiled surface polarizes incompletely.

The mechanism of polarization by reflection can best be explained as follows: The incident light coming from A, being unpolarized, has its vibrations as much in one plane through the ray AC as in any other, but every vibration can be resolved into a component vibration in the plane of incidence and one at right-angles. Therefore, we may regard the beam AC as composed of two parts, one having its plane of polarization in the plane of incidence, the other at right-angles to it. These two parts are unequally reflected, and for a certain angle of incidence, if the surface is fresh and clean, the second is not reflected at all. Experiment shows that if A represents the amplitude of that part of the incident light which is polarized in the plane of incidence, i the angle of incidence, and r the angle of refraction, the amplitude of the reflected ray to which it gives rise is

$$A' = -A\frac{\sin(i-r)}{\sin(i+r)}$$

and that of the refracted ray is

$$A'' = A\frac{2\sin r.\cos i}{\sin(i+r)}$$

But if B represents the amplitude of that part of the incident light whose plane of polarization is perpendicular to the plane of incidence, B′ that of the reflected beam, and B″ that of the refracted beam, to which it gives rise,

$$B' = B \frac{\tan (i - r)}{\tan (i + r)}$$

$$B'' = B \frac{2 \sin r \cos i}{\sin (i + r) \cos (i - r)}$$

Theses formulæ become indeterminate when $i = 0$, but hold good for any angle other than this. None of the numerators can vanish (except when $i = 0$), because for no other value of i can $i - r = 0$, provided there is any change in the medium at all. But one denominator, that in the expression for B′, does become infinite, if $i + r = 90°$. Therefore, when the angle of incidence becomes large enough—the angle of refraction becoming larger along with it—so that the two of them together amount to 90°, B′ vanishes. This means that the angle of polarization has been reached, for then none of the light polarized perpendicular to the plane of incidence is reflected at all, all of it being refracted. The only reflected light is then polarized in the plane of incidence.

Since the angle of polarization is the angle of incidence when the reflected light is all polarized in the plane of incidence, that is the angle such that $i + r = 90°$, we can derive a simple relation between this angle and the index of refraction. For then

$$\sin r = \cos i$$

and since

$$n = \sin i / \sin r$$

we have

$$n = \sin i / \cos i = \tan i$$

That is, the index of refraction is the tangent of the angle of polarization. It is easy to prove, by simple geometry, that when the angle of incidence is the polarizing angle, the refracted ray and the reflected ray make an angle of 90°.

When the plate on which the light falls at the polarizing angle has plane and parallel sides, as indicated in figure 114, the refracted light strikes the second surface at what is the polarizing angle for reflection inside the plate. Consequently, not only the light reflected from the first surface, but also that which emerges through the first surface after any odd number of reflections inside the plate is polarized. On the other hand, the transmitted light, although it shows some trace of polarization, is by no means strongly polarized. For, although all of the light whose plane of polarization is perpendicular to the plane of incidence is refracted, the major part of that polarized in the plane of incidence is also, so that the transmitted light has only an excess of vibrations in one plane.

Since a glass plate, by reflection, gives polarized light, a second glass plate may be used instead of a tourmaline to detect the polarization, as shown in figure 115. If the second glass plate, XY, be held parallel to the first, MN, it can reflect light of the same kind that MN reflects, for their planes of incidence are parallel. But if XY be turned about the ray CB as an axis through 90°, to the position shown at X′Y′, their planes of incidence become perpendicular, and the light reflected by MN cannot be reflected at X′Y′. Similarly a glass plate can be used to test any beam of light to see whether it is plane polarized or not. It is only necessary to set the plate so that it receives the beam at the angle of polarization, and then turn it about the incident beam as an axis, so as to keep the angle of incidence always equal to the polarizing angle. If, for any position of the plate, the reflected light vanishes, the beam in question is polarized in a plane perpendicular to the plane of incidence.

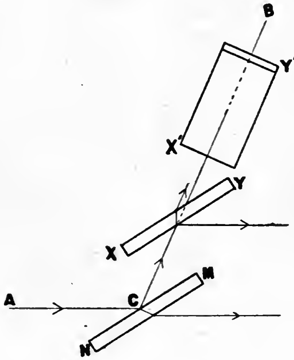

Figure 115

Problems.

1. What must be the angle a, figure 108, in order that the images b and c shall be twice as intense as the images a and d?

2. Prove that, when light strikes a glass plate at the polarizing angle, the reflected and refracted rays are at right-angles.

3. Find the angle of polarization for glass whose index of refraction is 1.71.

4. Show that light that has gone through a prism spectroscope must be partly polarized. In what plane will be the maximum polarization?

CHAPTER XIII.

112. Methods of polarizing light.—For many purposes in experimental optics, as well as in its industrial applications, it is desirable to obtain a strong beam of plane polarized light. A simple block of calcite does not answer the suppose, for it transmits two beams with mutually perpendicular planes of vibration, and these are parallel in direction of propagation, with only a lateral displacement which is too small to separate them completely. Tourmaline is better, since the ordinary beam is removed by absorption, but unfortunately tourmaline exerts a fairly strong absorption on the extraordinary also, particularly in certain wavelengths, and is usually strongly colored, so that the transmitted beam is weak. Reflection from a glass plate at the polarizing angle can furnish a wide clear beam of polarized light, free from absorption, but the following calculation will show that this device utilizes only a small portion of the incident light, and therefore the polarized beam obtained is weak.

If we take the index of refraction of the glass plate as 1.54, and remember that this is the tangent of the polarizing angle, this angle is found to be 57°. With an angle of incidence of 57° and index 1.54, the angle of refraction is 33°. We saw in the last chapter that if light of amplitude A, polarized in the plane of incidence, strikes the surface, the amplitude of the reflected light is

$$A' = -A\frac{\sin(i - r)}{\sin(i + r)}$$

If we substitute $i = 57$, $r = 33$, we get $A' = -.407A$, that is the amplitude of the reflected ray is only about .4 that of the incident, and consequently the energy of the reflected light is only $(.4)^2 = .16$ that of the incident. That is, only 16 percent. of the incident energy is reflected, 84 percent transmitted. All this is on the supposition that the incident light is already polarized in the plane suitable for reflection. In reality, we are always provided with a completely unpolarized incident beam, only half the energy of which is polarized in the plane of incidence, and therefore only 8 percent of the incident energy is available for the production of a polarized beam. If the reflecting plate were a perfectly efficient polarizer, the reflected polarized beam would contain 50 percent of the incident energy. It is true that the efficiency of the plate is raised somewhat if we consider the light reflected from the second surface, but this contribution is small, and the presence of extra images by internal reflection is usually undesirable. Glass plates used for polarizing by reflection are usually made of black opaque glass, so as to avoid these extra images.

113. The Nicol prism.—Figure 116 shows an end and a side view of the *Nicol prism,* the best-known device for obtaining polarized light. It is made of calcite worked in such a way as to get rid of the ordinary light and allow the extraordinary to pass through. To make one, a rather long and narrow crystal of calcite is cut in two by a plane through the obtuse corners A and B. The two halves, after being ground and polished, are

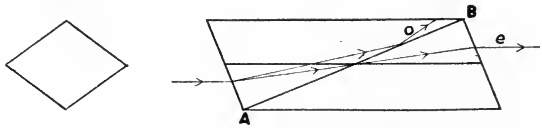

Figure 116

cemented together again with a very thin layer of Canada balsam. The inclination of the end faces is also slightly altered from the natural cleavage surfaces. Canada balsam is a mucilaginous substance which has an index of refraction less than that of calcite for the ordinary light, but greater than the effective index of the calcite for extraordinary light, at the angle at which it comes in the ordinary use of the prism. For this reason, it is possible for the ordinary light to be totally reflected, if the angle of incidence on the plane AB is great enough, while the extraordinary cannot be totally reflected, no matter what this angle may be. The slant of the end faces and that of the cut AB are so calculated that for light incident parallel to the prism axis, or for a few degrees each side of it, the ordinary light strikes AB at greater than the critical angle. Consequently the ordinary light is totally reflected off to the side of the prism, where it is lost, and the extraordinary passes on alone. The Nicol prism therefore gives a clear colorless beam of plane polarized light, and its efficiency is high, although of course the emergent light is to some extent weakened by reflection where the beam enters and leaves the prism.

114. Double-image prisms.—There are also several devices in which, although neither the ordinary nor the extraordinary light is eliminated, they emerge, not with parallel rays as from a simple calcite rhomb, but with an appreciable angle between them, forming a diverging pair of beams. One of the oldest is the Wollaston prism, figure 117. It is made of two wedges of calcite, ABC and ACD, so cut that in each the optic axis is perpendicular to the entering ray of light, but the axis in one wedge is perpendicular to that in the other. For instance, it is perpendicular to the plane of the paper in ABC, but parallel to the line CD in ACD. A beam of unpolarized light, on entering the face AB, is divided into an ordinary and an extraordinary, neither of which is bent when the incidence is normal, because the optic axis is parallel to the face. But on striking the diagonal face AC, the ray which had been the ordinary in the first wedge becomes the extraordinary in the second, and vice versa, because the optic axis of the second is perpendicular to that of the first. Therefore one beam passes from a medium where it has less velocity to one where it has greater and is therefore bent away from the normal, while the other passes from a medium where it has greater velocity to one where it has less, and is therefore bent toward the normal. Both beams, on striking the air at the surface CD, pass into a medium of greater velocity and are therefore bent away from the normal. Thus they emerge from the prism with a considerable angle of divergence, and become more and more separated as they recede from the prism.

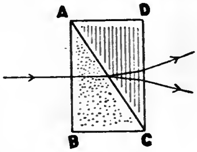

Figure 117

The Rochon prism, figure 118, differs from the Wollaston in that one wedge has its optic axis parallel to the entering ray, so that both rays travel through it with the same velocity. One ray, the ordinary for the second wedge, suffers no change in speed on striking the diagonal AC, but the other undergoes an increase in speed which bends it away from the normal. The net result is two polarized beams, one of which goes straight through the prism without any bending whatever, while the other suffers a change of direction.

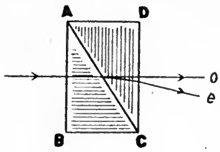

Figure 118

Devices like the Wollaston and Rochon prisms, which give two divergent beams of polarized light, are called *double-image prisms.* Prisms are now made which differ from the Rochon type only in that a wedge of plain glass is substituted for the wedge whose optic axis is perpendicular to the entering face, and they work very well. Double image prisms are useful for many optical purposes, but where only a single beam of polarized light is desired it is more usual to employ a Nicol prism.

115. Crossed Nicols and crystal plate.—Two Nicol prisms are said to be *crossed* when they are held so that their transmission planes are at right-angles, that is, so that the polarized beam transmitted by the first Nicol is refused transmission by the second. The first one is then called the *polarizer,* for obvious reasons, the second the *analyzer.* No light, of course, can be seen through a pair of Nicols so held, but if a thin slip of mica, or of any other doubly-refracting substance, is inserted between the polarizer and the analyzer, a considerable amount of light will, in general, pass through the combination. If the slip of crystal be rotated about the beam of light as an axis, certain positions will be found where this light disappears, leaving things as they were before the slip was inserted. On the other hand, if the slip be left in such a position that light passes through, and the analyzer be rotated, the brightness of of the transmitted light may show fluctuations, but it does not vanish completely for any position of the analyzer. The last fact shows that the action of the crystal slip is not to rotate the plane of polarization, but either to depolarize the light or to change it to a new kind of polarization which cannot be shut out by a Nicol, however the latter be held.

116. Elliptic polarization.—It is in fact the latter alternative which holds here, that is the experiment introduces us to a new kind of polarization, known as *elliptic* polarization, which can be explained by reference to figure 119. Let OA represent in direction and length the amplitude of the light which the polarizer transmits. Then, since the Nicols are crossed, XY is the direction of vibrations which alone can be transmitted through the analyzer. Mica, being a doubly-refracting crystal, transmits vibrations in two mutually perpendicular planes, but with different velocities. Suppose that OB and OC are these two directions for transmission through the mica. The amplitude OA will then be resolved into two vibrations of amplitude

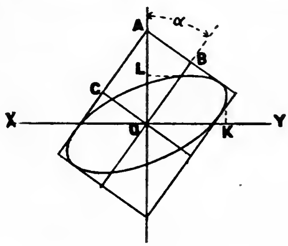

Figure 119

$$OB = OA. \cos \alpha$$
$$OC = OA. \sin \alpha$$

These two are in the same phase when they enter the mica, but since one travels faster than the other it will gain in phase till they both emerge again. If the gain in phase does not amount to an exact multiple of π, the two will, on emergence, no longer be equivalent to the linear vibration OA, but to some elliptical vibration, the elliptical form being inscribable within the rectangle of dimensions 2OB $\times$ 2OC, as fully explained in section 92. The beam of light coming from the slip of mica is therefore neither unpolarized nor plane-polarized, but is clearly a very particular kind of vibration, and is appropriately known as elliptically polarized light. The elliptical vibration of figure 119, which is equivalent to the linear vibrations of amplitude OB and OC with a certain phase-difference, is also equivalent to vibrations of amplitudes OK and OL, with another phase-difference. The vibration OL cannot pass through the analyzer, but the vibration OK can. Consequently light is seen through the combination when the mica slip is inserted. If the plane of transmission of the analyzer, XY is rotated, the amplitude of the transmitted component will vary, its greatest and least values being the major and minor semi-axes of the ellipse, but it will never vanish. On the other hand, suppose that the analyzer is held in the position shown in the figure, but the crystal slip is rotated, so that the angle α changes. When α is zero or 180°, OB is parallel to OA and equal to it in absolute amount, while OC is zero. Then only a single beam goes through the mica and it is in such a direction as to fail to pass through the analyzer. If a is 90° or 270°, OC is parallel to OA and equal to it in absolute amount, while OB is zero. Again only a single beam goes through the mica and it is in such a direction as to fail to pass the analyzer.

117. Circular polarization.—If, in figure 119, $a = 45°$, OB and OC are equal. If, in addition the difference in phase of the two beams through the mica is $\pi/2$, or 90°, the ellipse becomes a circle, and we have emerging from the mica what we call *circularly-polarized* light. Evidently circularly polarized light may be right-handed or left-handed, according to which of the two component beams passes through the mica with the greater velocity. (The same is true of elliptically-polarized light.) Now a difference of phase of $\pi/2$ evidently means that one of the beams passing through the mica gains a quarter of a wavelength over the other, and as we saw in the preceding chapter a plate thick enough for one component to gain just a quarter of a wavelength over the other is what is known as a quarter-wave plate. Accordingly, we have the following rule for the production of circularly polarized light: Place a quarter-wave plate in the path of a beam of plane-polarized light, so that its two planes of transmission make an angle of 45° with the plane of polarization of the incident light. The transmitted light is then circularly polarized. Quarter-wave plates are usually made of mica because it splits into layers so readily. It is peeled down, a thin layer at a time, to the required thickness.

A quarter-wave plate made of calcite would have to be exceedingly thin, on account of the great difference, in this crystal, between the ordinary and the extraordinary velocities. It is true, that, theoretically, a plate so thick that one wave gains over the other a quarter of a wavelength plus any whole number of wavelengths would act as a simple quarter-wave plate, and it would so act in practice if the light used were monochromatic. But doubly-refracting substances, like isotropic materials, are subject to *dispersion,* or differences in velocity for different wavelengths, and the dispersion for the ordinary and for the extraordinary light is not the same. For example, a plate thick enough for the extraordinary ray to gain 10¼ wavelengths over the ordinary in the yellow might show a difference of 10¾ in the red, 9¾ in the green, 9 in the blue, and 8½ in the violet, these figures being merely illustrative, and not actual statements of what any particular plate would do. Under these circumstances, certain colors would be cut out completely, all those in fact for which the gain was an exact whole number of wavelengths. For an examination of the figure will show that for these colors the rays of amplitude represented by OB and OC would emerge from the plate in the same phase, and would therefore recombine to produce the original plane polarized light of amplitude OA. Of the other colors, some would be represented by right-handed, some by left-handed circularly polarized light, and others by elliptically polarized light of various configurations of the ellipse. Consequently, of the light passing through the analyzer, parts of the spectrum would be completely eliminated, other parts much weakened, and still other parts quite strong. The result would be brilliant color effects, which would change when either the analyzer or the crystal plane was rotated. On the other hand, if the gain of the extraordinary over the ordinary ray were only ¼ wavelength for any particular color, it would not differ a great deal from that for any visible wavelength, and there would hardly be a suggestion of color in whatever light got through the analyzer.

The phenomena explained above afford a very convenient and delicate test for double refraction. It is only necessary to set up a pair of crossed Nicols and insert between them a slice of the material to be investigated. If, when this is rotated into various positions, no light passes through the analyzer, the material is free from double refraction. Glass is of this character if carefully annealed and not strained, but a piece of commercial glass-ware, if experimented with between crossed Nicols, is almost sure to show streaks of light which indicate an irregular double refraction, and even well annealed glass, if bent between the fingers, or otherwise slightly strained, shows the same effect, though the degree of double refraction is far too small to show double images to the unaided eye.

118. Rotation of the plane of polarization.—Quartz is a doubly-refracting crystal, and shows phenomena similar to those of calcite, though to a far less degree, the difference in the velocities of the two rays being always small. But quartz shows in addition a very remarkable property which calcite and most other doubly-refracting crystals do not share, what is called *optical rotation.* When a beam of light is passed through quartz along the optic axis, just the condition under which double refraction does not occur, the plane of polarization is turned without changing the direction of the ray, and the number of degrees of turning is directly proportional to the thickness of quartz passed through. Some samples of quartz twist the plane to the right, others to the left, and in all samples the amount of turning is very different for different wavelengths.

If a layer of quartz crystal, cut so that the optic axis is perpendicular to the faces, is inserted between a pair of crossed Nicols, light reappears, just as when a slip of mica is so inserted, but it is very easy to distinguish between the two effects. With the slip of mica, as we have already seen, a rotation of the slip shows certain positions where the light vanishes when the Nicols remain crossed, but a rotation of the quartz has no effect, so long as the optic axis remains parallel to the beam. On the other hand, with the mica fixed in position so as to show light through the crossed Nicols, a rotation of the analyzer may cause fluctuations in the transmitted light, but does not quench it entirely, while with the quartz a rotation of the analyzer will enable the light to be completely cut out. These facts show that the effect of the quartz is not to produce elliptically polarized light like the mica, but to leave the light plane polarized but with a changed azimuth of the plane of polarization. Since quartz is itself a doubly-refracting crystal, it too would produce elliptical polarization if the light went through it perpendicular to the optic axis, and when the ray is inclined to the optic axis the effect is a combination of double refraction and rotation which is complicated and difficult to describe.

The power to rotate the plane of polarization is shown by a number of other crystals besides quartz, and also by certain solutions, notably the sugars, and it forms the basis of an elaborate technical method of analyzing and testing commercial sugars. Incidentally, since in solutions there can be no question of a regular arrangement of molecules such as occurs in crystals, it seems certain that solutions must owe their rotating power to something in the interior structure of the molecule.

Fresnel has offered a very ingenious kinematical explanation of optical rotation. Taking a suggestion from the fact that in ordinary double refraction linear vibrations in two perpendicular planes are transmitted through the crystal with different velocities, he made the supposition that along the optic axis of quartz *circular* vibrations in two directions, viz., right-handed and left-handed, are transmitted with different velocities, and from this supposition the rotation of the plane of *plane*-polarized light follows as a logical deduction. We have shown in section 92 that a circular vibration is equivalent to two linear vibrations, at right-angles in direction and with a phase-difference of $\pi/2$, and we can now show that conversely a linear vibration is equivalent to a right-handed and a left-handed circular vibration, each having half the amplitude of the linear vibration. This can be done analytically, by the use of equations, but the following graphical method is perhaps better.

A right-handed circular motion can be represented by a vector, whose length is equal to the radius of the circle, and which swings at a uniform rate about the origin, like the vector OA in figure 120, that is, the end of the vector always gives the position of the body undergoing the circular motion. A left-handed circular motion would be similarly represented by a vector swinging in the opposite direction. The rule for compounding two vectors is to place the origin of one at the terminus of the other. Therefore, to compound a left-handed with a right-handed circular motion, we take the terminus, A, of the latter for the center about which the former, AB, turns. Now if OA swings at a uniform rate to the right while AB swings at a uniform rate to the left from the end of

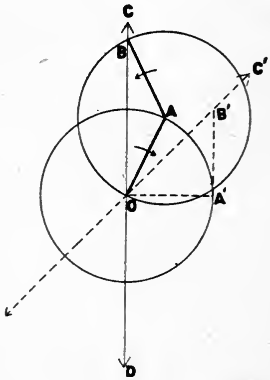

Figure 120

OA, the point B will describe the path of a point whose motion is a combination of the two circular motions. If OA and AB are equal in length, and if they both rotate at the same rate, this path is the straight line CD, as can be easily proved by simple geometry. It is also easy to prove that the motion is simple-harmonic, with amplitude twice the radius of either circle. Any plane-polarized beam of light can therefore be regarded as composed of two circularly-polarized components of the same period but half the amplitude, one right-handed, the other left-handed. According to Fresnel's hypothesis, these, on entering the quartz along the optic axis, will travel through it with different velocities. On emerging into the air again, one of these components will have gained on the other in phase, because of its greater velocity, and this gain in phase causes the plane-polarized beam to which the two on emergence are equivalent to be turned somewhat from the plane of polarization of the light as it entered the quartz. In order to explain this, suppose that the right-handed vibration in the figure is ¼ revolution ahead of the left-handed. Then, when the former had turned 90° from the vertical, into the position OA′, the latter would be just passing through the vertical position, as A′B′. Consequently, the path of the tracing point would be along the diagonal OC′ instead of OC, and the plane of the vibration would be rotated through 45°. Evidently, the angle through which the plane of polarization is rotated is half the gain in phase of one component over the other, and therefore it will be proportional to the length of path traversed in the quartz.

Fresnel subjected his theory to a further interesting test. If the two kinds of circularly polarized light travel with different velocities within the quartz, they will have different indices of refraction. Now suppose that plane polarized light be sent through a prism cut out of quartz in such a way that the optic axis is parallel to the base of the prism. The two circularly-polarized components to which the plane-polarized beam is equivalent, having different indices of refraction, should separate, just as two different wavelengths separate in going through a glass prism, emerge with different directions, and be brought to a focus at different points. This actually proves to be the case. Light of a single wavelength, when sent through a spectroscope with such a prism gives two spectral lines instead of one, one line composed of right-handed, the other of left-handed circularly polarized light, but both having the original wave-length. When Fresnel tried this experiment, he actually used a row of prisms instead of one, a prism of right-handed quartz followed by one of left-handed quartz with its base in the opposite direction, this in turn followed by a prism like the first, and so on. By this arrangement, the effect of a single prism in separating the two beams is greatly increased, but such an elaborate arrangement is not necessary. A single 60° prism of ordinary size is sufficient to show the effect, and indeed this doubling of the spectrum lines is so troublesome that whenever quartz prisms are used for studying ultraviolet light, they must be made double, half a prism being of right-handed, half of left-handed quartz, the two halves having their bases in line and their apexes together. By this contrivance the effect is neutralized, one prism undoing the rotation produced by the other, and single spectral lines result.

It should be noted that the doubling of spectral lines by a quartz prism can be explained without making use of Fresnel's hypothesis, as a simple result of rotation and diffraction, but there can be no doubt that right-handed and left-handed light does pass along the optic axis of quartz with different velocities.

119. Magnetic rotation.—Michael Faraday found that when certain substances are placed within a strong magnetic field and plane-polarized light is sent through them along the magnetic lines of force, the plane of polarization is rotated just as it is in a sugar solution, or in quartz along the optic axis. This discovery was the first intimation of any connection between optical and electromagnetic phenomena, and is one of the facts that stimulated the electromagnetic theory of light, which we are to take up later. A similar phenomenon was discovered by Kerr. He found that when a beam of light, polarized in or at right-angles to the plane of incidence, is reflected from the polished pole piece of a strong magnet, the plane of polarization is slightly rotated in the reflected light.

There is one important difference between the rotation in quartz and other crystals or solutions, and the rotation due to magnetic action discovered by Faraday. In the former, if the rotation is right-handed, for instance, going in one direction, it will also be right-handed going in the opposite direction. Stating the matter in another form, if a beam be sent along the optic axis of a piece of right-handed quartz, so that the plane of polarization is turned to the right, and then reflected back over its path by means of a mirror, the plane will again be rotated to the right, bringing it back exactly to the azimuth which it originally had, for a right-handed rotation with reversed path undoes the effect of the original right-handed rotation. On the other hand, in the case of magnetic rotation, the turning is to the right when the light goes out in the direction of the magnetic lines of force, but to the left when it goes against the latter, so that a beam sent out along the lines of force and reflected back suffers twice as much rotation as if it traversed the distance only once. In magnetic rotation, the direction of rotation depends upon the direction of the lines of force of the magnetic field, but in rotation produced by crystals and solutions it depends upon the direction of the ray.

120. The rings-and-brushes phenomenon.—Some very remarkable and beautiful effects are produced when a beam of plane-polarized light is passed through a thin layer of doubly-refracting crystal cut with its faces perpendicular to the optic axis, and the light is received through an analyzing Nicol prism and a telescope. Thus, let AB, figure 121, represent a slip of some uniaxial crystal which is free from the rotating power of quartz. CD, perpendicular to the faces of the slip, shows the direction of the optic axis, and the polarized light comes through from below, in a cone whose axis is parallel to this line. The axial ray CD obviously suffers no double refraction, but a ray ED, inclined to the optic axis, is resolved into an ordinary and an extraordinary ray, and these emerge from the crystal parallel to one another and with a phase-difference, unless the plane of polarization of the incident light happens to be either parallel or perpendicular to the plane through ED and the optic axis, the plane of the paper

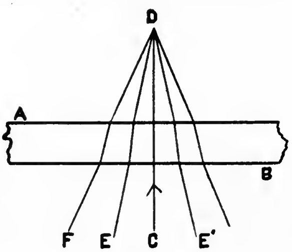

Figure 121

for this case. The phase-difference between ordinary and extraordinary will depend upon the inclination and length of the path within the crystal, for instance it will evidently be greater for the ray FD than for ED, since FD is more inclined to the axis and has also a longer path within the crystal.

The square area in figure 122 represents the crystal slip as seen from above, the optic axis now being perpendicular to the paper. Let O be the point of emergence of the axial ray Cd of figure 121, and AB the plane of polarization of the incident light. Referring back to figure 121, it is clear that the phase-difference that exists between the components of the ray

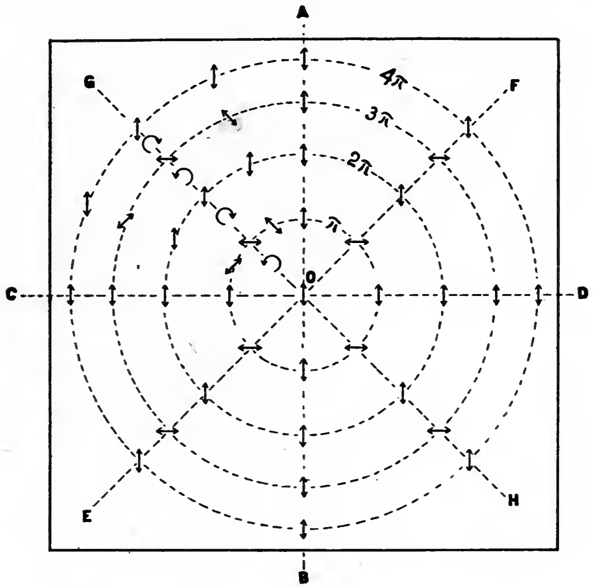

Figure 122

ED will be the same as for any other ray such as E′D which emerges from the crystal at the same distance from the point O of figure 122. Therefore, if in the latter figure we draw circles about O, the phase-difference between the ordinary and the extraordinary rays on emergence will be the same for all points on any one circle, but different for points on different circles. Circles are drawn for which this phase difference is π, 2π, 3π, 4π, respectively.

Consider first the light that comes through at points along the line AB. For these points the principal plane is the same as the plane of polarization, therefore there is no extraordinary ray and all the light passes through as ordinary light. Next take points along CD. For these the principal plane is perpendicular to the plane of polarization of the incident light, and therefore there is no ordinary ray, and all the light comes through as extraordinary ray with its plane of polarization unaltered. Points along AB and CD would then be represented by plane-polarized light with the same plane, as indicated by the arrows along these two lines. Now take points along EF or GH. Here the principal plane is at 45° with the plane of polarization of the incident light, there will be both an ordinary and an extraordinary ray, and their amplitudes will be equal, but the phase-difference between them will depend, as we have seen, upon the distance of the point from the center. At O itself, and also where EF and GH cross the circles marked 2π, 4π, etc., the two rays are equivalent to the original polarized incident ray, for a phase-difference of 2π, or any even multiple of π, is equivalent to no phase-difference at all. Consequently we mark arrows at these points parallel to the plane of polarization of the incident light and to the arrows along AB and CD. But where EF and GH cross the circles marked π, 3π, etc., (any odd multiple of π is equivalent to π when we are speaking of phase-differences) the two rays are equivalent to a ray polarized at *right-angles* to the incident light, as can be seen by comparing the diagram of figure 94 for $\beta = \pi$, with that for $\beta = 0$ or $\beta = 2\pi$. These points are marked with arrows at right-angles to those along AB and CD. The reader should now be able to foresee what happens at points where the circles are cut by lines making any other angle, say 22.5°, with AB or CD. At all these points there will be both an ordinary and an extraordinary ray, of unequal amplitudes. At the circles of phase-difference 2π, or 4π, the original plane of polarization is reproduced, but at the π or 3π circles the result will be plane polarized light, but with an inclined direction. Arrows are marked only in the upper left-hand quadrant of the figure to indicate the azimuth of the plane-polarized light at points on these circles.

Other circles might be drawn, for which the phase-difference is $\pi/2$, $3\pi/2$, $5\pi/2$, etc., and it is not difficult to show that at various points on these circles the resulting light is circularly or elliptically polarized, except where they are crossed by the lines AB and CD, where the ellipse becomes a straight line.

We have shown that the light coming through at various places on the plate would have diverse states of polarization, but nothing of all this would show to the eye without the use of a second Nicol prism or other analyzer, for the eye cannot detect polarization. If an analyzer is inserted, parallel to the polarizer which furnishes the incident light, brightness will show at all those points of the diagram marked with arrows like those along AB or CD, complete darkness at points marked with arrows at right-angles to these, and partial illumination at points where the polarization is elliptical or circular, or plane but inclined to the line AB. There will be a bright cross AB and CD, cutting across a series of bright and dark rings. If the analyzer is crossed with the polarizer, the pattern will be exactly reversed, consisting of a black cross cutting across a ring-system. This is the phenomenon of "rings and brushes," the name being suggested by the appearance.

121. The nature of elliptic and circular polarization.—Students often find difficulty in forming a satisfactory physical conception of circularly-polarized waves, or of unpolarized waves, although plane-polarization may offer no particular difficulties. Probably this may be overcome most easily by considering simple mechanical waves in an elastic jelly, as was done in chapter III to explain plane wavefronts. Figure 17 of that chapter represents a block of jelly with a stiff board attached to it so that any movement of the board in its own plane sends a train of transverse plane waves through the jelly. It is clear enough that a movement of the board back and forward in the direction AB, or along any line in the plane of the board inclined to AB at any angle, would cause the waves produced to be plane-polarized. In order to make circularly-polarized waves, the motion of the board would have to be circular in its own plane, but the motion must be one of *translation,* not of *rotation.* The board must not turn about a fixed axis, like a wheel on its shaft, but must move so that its edges remain parallel to their initial directions and so that every point in the board moves in a circle of the same diameter. Then one plane after another in the jelly would take up the motion, and circularly-polarized wavefronts would advance through it. If the path of every point in the board were an ellipse, the same would be true of each point in the jelly when the wavefront reached it, and the waves would be elliptically polarized. In order to produce unpolarized waves, the board must move so that, although its edges keep parallel to themselves, the path of each point in it would be a very irregular and random sort of curve. Each point in the jelly would in its turn go through the same path, and no particular direction of vibration would predominate.

Whether light vibrations actually consist of real mechanical displacements in the ether, like the mechanical displacements in the jelly, is a question which up to the present we have ignored. But it seems certain at any rate that whatever may be the character of the ether-disturbances which produce light, they must be of the nature of vectors, since the phenomena of polarization clearly indicate that they have direction. Therefore it is convenient and permissible to represent them as mechanical displacements, remembering that they may prove to be something else.

Problems.

1. A device that has been used for producing polarized light is a pile of plates set so that the light passing through strikes each at the polarizing angle. If each reflection (two for every plate) reduces by 16% the energy of that part of the beam incident upon it which is polarized in the plane of incidence, what would be the total percentage reduction by 10 plates?

2. Light passes through one Nicol prism and then through another whose plane of transmission makes an angle a with that of the first. Neglecting losses by reflection from the end faces, what must be the angle a in order that the second Nicol cut down the intensity to ¼? to ½? to ¾?

3. Suppose a beam of light, whose origin is unknown, is passed into a room. What instruments would be used, and how would one proceed, to find whether it is unpolarized, plane-polarized, circularly polarized, elliptically polarized, or partly plane-polarized?

4. Mica transmits, perpendicularly to its natural cleavage planes, two beams whose indices are 1.5609 and 1.5941, their planes of polarization being of course mutually perpendicular. These indices are for light of wavelength .00005893. What will be the thickness of a quarter-wave plate of mica for this wavelength?

CHAPTER XIV.

122. Plane of polarization and plane of vibration.—It has been shown that, in the ordinary and the extraordinary waves produced by double refraction, the directions of the vibrations —or, more precisely, the directions of the light-disturbance vectors—are at right-angles to one another. But no evidence was shown to indicate whether this vector for the ordinary or for the extraordinary lies in the principal plane. The difficulty was temporarily avoided by agreeing to call the principal plane the "plane of polarization" of the ordinary waves; and consistently therewith, the plane of polarization for the extraordinary waves must be perpendicular to the principal plane. This is a pure definition, rather than a statement of physical fact, and leaves it an open question whether the vibrations in polarized light lie in or perpendicular to the plane of polarization.

123. Elastic-solid theories.—Such a question could not be permanently ignored by physicists, particularly as it was found to be of decisive importance for certain theories which were worked out mathematically during the nineteenth century. These theories all started from the assumption that the ether behaves like an elastic jelly, and that light consists of real mechanical transverse waves in it, the velocity of which depends on the density and the elastic coefficients of the ether. Since light travels slower through glass and other material media than through the free ether, it was assumed that the presence of material molecules alters either the density or the elastic properties of the ether. In doubly-refracting substances, the velocity is different for different directions of the vibrations, and this fact would lead us to infer that it is the elastic constants, rather than the density, which is changed, for density, as we ordinarily know it, is not a vector quantity, and has nothing to do with direction. Nevertheless, some theorists assumed that it is the density that is changed, others that it is the elastic coefficients, for the development of the science had reached such a point that further progress could be made only by making such assumptions and seeing whether all the conclusions to which they led were in accord with experimental facts. Each of these assumptions accounts for some of the facts of reflection and double refraction, but neither is satisfactory at all points. One leads to the conclusion that in polarized light the vibrations are parallel to the plane of polarization, the other that they are perpendicular to it.

A fundamental weakness in these "elastic solid" theories is that they fail to explain why we never find indications of *longitudinal* ether waves. For a disturbance inside an elastic solid like a jelly will send out not only transverse but also longitudinal waves, which will travel with a different velocity, greater or less than the transverse, according to the relative values of the elastic coefficients, the compressibility and the rigidity. Moreover, longitudinal waves striking a reflecting surface should set up transverse waves, and vice versa. Attempts were made to solve this difficulty by assuming that the longitudinal waves failed to make their presence known to us because their velocity was enormously great or enormously small, but the results were by no means satisfactory. It is also difficult to see how a medium which exerts an elastic reaction against a displacement can allow bodies like the planets to move through it without any retardation.

124. Electromagnetic theory.—Although text-books written in comparatively recent years have given a prominent place to the elastic solid theories, these are now practically obsolete, supplanted by the electromagnetic theory. Therefore we shall in this book make no further reference to them, but confine our attention to the electromagnetic theory and its consequences. A complete mathematical treatment would be out of place in a text of this kind, but an attempt is made in the following pages to give, first a physical conception of the character of electromagnetic waves, second a statement of the electromagnetic laws that form their basis, with a little of the history of the development of the theory.

We are no longer to think of the ether as a material medium, with a certain density and a certain degree of rigidity, but simply as the seat of electrical and magnetic forces, and the laws of these forces are all that we need to know about it. Figure 123 is designed to explain the electrical part of what

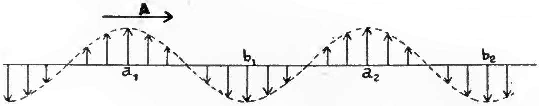

Figure 123

we mean by an electromagnetic wave. At a_1, a_2, etc., and at points in their immediate neighborhood, what we call the *electric force,* or the *intensity of the electric field,* is directed upward in the figure, while at b_1, b_2, etc., and in their neighborhood, it is directed downward. The meaning of the above statement is this: if a body with a positive electric charge were placed at a_1 or a_2 it would be acted upon by an upward force, while at b_1 or b_2 it would be acted upon by a downward force. The small vertical arrows indicate the direction and the magnitude of the electric force at the different points. Now imagine this whole condition of affairs to be in process of transferrence to the right as indicated by the arrow A, so that after a certain time the a's will be places of downward and the b's of upward electric force. We should then have an electric wave travelling in the direction A.

Such a wave, however, cannot exist alone, for the changes in electric force are necessarily accompanied by changes in *magnetic* force, in a plane at right-angles to the electric force, which means at right-angles to the paper in figure 123. To show properly both the electric and the magnetic parts of the complete wave, would require a three-dimensional model, but in figure 124 an attempt is made to represent the whole thing in

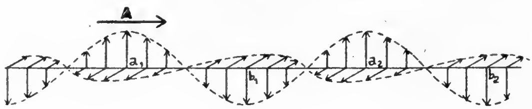

Figure 124

a perspective drawing. A positive charge placed at a_1 or a_2 would experience a force upward, while the north-pointing pole of a magnet placed there would experience a force outward toward the reader. At b_1 or b_2 both forces would be reversed. The reader should not think of the electric and the magnetic parts as two separate waves, but simply as two different aspects of the same wave, for, from the nature of electrical and magnetic phenomena, neither one can exist without the other.

Such a wave as is depicted in figure 124 would undoubtedly be a polarized wave, for we have considered the electric vibrations to be always in a vertical, the magnetic in a horizontal plane. Either set of vibrations might, however, be in any plane parallel to the direction of propagation, provided the two parts, electric and magnetic, are perpendicular to one another. A train of waves in which the plane of the electric vibrations is continually changing at random would be an unpolarized train. Polarized electromagnetic waves of the type of figure 124 are exactly the kind sent out by the transmitting apparatus of a wireless telegraphy outfit. The electric vibrations are perpendicular to the earth's surface, the magnetic parallel to it.

The electromagnetic theory of light, although historically it arose long before the development of wireless telegraphy, really amounts to the belief that polarized light-waves are exactly identical in everything except wavelength and frequency with the waves of wireless telegraphy, so that if a wireless transmitter could be made to vibrate rapidly enough, and thus give out short enough waves, the usual receiving apparatus could be dispensed with and the signals received by the eye. Wireless telegraphy would then become identical with signaling by flashes from a lantern, and some of its advantages, which depend upon great length of wave, would be lost.

125. Direction of the vibrations.—The introduction of the electromagnetic theory, with its two kinds of vibrations at right-angles to one another, makes a fundamental change in the old question about the plane of polarization. Instead of asking, "Are the vibrations parallel or perpendicular to the plane of polarization?", we must now ask, "Is it the electric or the magnetic vibrations that lie in the plane of polarization?" The answer to this question has been found by applying rigorous mathematical methods to the reflection of electromagnetic waves from glass at the polarizing angle, first when the electric vibrations, second when the magnetic, lie in the plane of incidence. The theory indicates that in the first case there is no reflection, all the energy being transmitted, while in the second part is reflected and part transmitted. If the incident light has both its electric and its magnetic vibrations inclined to the plane of incidence, we can resolve it into two components. In one component, the electric vibrations are in the plane of incidence and the magnetic perpendicular to it, and no part of this component is reflected. The other component has its magnetic vibrations in the plane of incidence and its electric at right-angles, and some of this energy is reflected. Therefore, in general, electromagnetic waves reflected at the polarizing angle have their magnetic vibrations in the plane of incidence, and their electric vibrations at right-angles. Since the plane of polarization has been defined so that, for light reflected at the polarizing angle, it is identical with the plane of incidence, we can now draw the general conclusion that, if the electromagnetic theory is correct, magnetic vibrations lie in the plane of polarization, electric at right-angles to it. Since it has become customary to understand when one speaks of the "vibrations" of electromagnetic waves, without any qualifying adjective, that it is the electric vibrations that are meant, it is often stated somewhat loosely, that the vibrations are perpendicular to the plane of polarization.

The theory of electromagnetic waves in space was worked out on a strictly mathematical basis by Clerk Maxwell. He was able to prove that such waves are possible, and that their velocity of propagation would be 3×10^{10} centimeters per second in the free ether. Since this is the velocity found experimentally for light, he announced the belief that light consists of electromagnetic waves of very short length.

A complete verification of Maxwell's theory, however, would call for an experimental demonstration that waves of the character contemplated in this theory actually exist, under such circumstances that, from the manner of their production and the methods of detecting them, there could be no doubt that they are electromagnetic, and this was not done until after Maxwell's death. The difficulty lay not simply in producing the waves,—for according to Maxwell's theory any electric or magnetic phenomenon such as the discharge of a condenser or the mere movement of a charged body or magnet would send out such a wave—but in getting them of sufficient intensity and devising suitable means of detecting them. These experimental difficulties were first solved by Heinrich Hertz, whose work formed the foundation from which Marconi, Deforest and others developed modern wireless telegraphy.

The complete mathematical theory of electromagnetic waves would be too difficult for a book such as this, but the following treatment should give some physical idea of why these waves are possible and how they propagate themselves through the ether. We begin by recalling two fundamental laws, found by experiment, for electromagnetic phenomena.

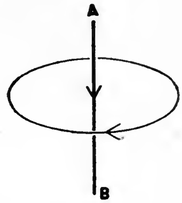

Figure 125

126. Fundamental electromagnetic laws.—If the wire AB in figure 125 carries an electric current in the direction shown by the arrow, it is surrounded by a "field of magnetic force," that is, magnetic lines of force encircle the wire in the direction shown. If the wire is long and straight, and the return part of the circuit far away, the lines of force are circles, and in fact any circle coaxial with the wire is a line of force. Otherwise, the lines are not true circles, but they still form closed curves surrounding the wire. The direction of the magnetic lines bears the same relation to the direction of the current as the direction of rotation of a right-handed screw to the direction in which the screw advances. This is our first fundamental empirical law.

The second law relates to the induction of electric currents. Suppose we have a plain loop of wire, in which there is originally no current, and no battery, dynamo, or other device for producing a current. Then if, by any such means as the movement of a magnet, the starting or stopping of a current in the neighborhood, etc., lines of magnetic force are caused to thread the loop, or lines previously threading it are withdrawn, a momentary current will traverse the circuit.

127. Faraday's displacement-currents.—Now consider the kind of circuit which we call incomplete, shown in figure 126. The two straight lines at C represent the plates of a condenser, such as the inner and outer coatings of a Leyden jar, and B is an ordinary battery cell. A *steady* current cannot flow in such a circuit, but when the poles of the cell are first joined by connecting wires to the plates of the condenser, there is a real current which lasts only a very short time. The current is clockwise, that is, we say that positive electricity flows from the right-hand plate of the condenser around through the wires and the cell to the left-hand plate, leaving the right-hand plate negatively charged and charging the left-hand plate positively. The current ceases when the difference in potential between the condenser plates, due to their charges, equals that between the poles of the cell.

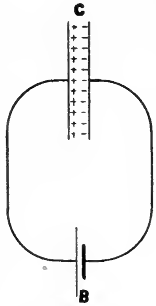

Figure 126

Such a circuit is called incomplete because there is no conducting path through the condenser. But Michael Faraday took the view that every electrical circuit is in a certain sense a complete circuit. He regarded electricity as being capable of motion within a non-conductor (or, dielectric) as well as within a conductor, but with this difference: in a conductor, its motion is analogous to the flow of water through a pipe, which retards the water by a sort of frictional resistance but has no tendency to reverse its motion; while in a non-conductor there is a sort of elastic reaction opposing the flow, a reaction which becomes greater as the displacement becomes greater and which tends to reverse it. A rather good analogy with Faraday's ideas is gotten by considering the wires of figure 126 replaced by pipes, the cell by a pump, positive electricity by water, and the condenser by an elastic membrane which the water cannot penetrate, but which can be stretched when water is pumped away from one side of it to the other side. When the pump is started water flows through the pipes and the membrane is bulged out toward one side, and the flow stops when the membrane maintains a difference of pressure equal to that produced by the pump. If another pipe-route, not containing a pump, is opened between the two sides of the membrane, the latter will flatten out, discharging the water through this route. If the resistance of this pipe is not too great, the inertia of the water will cause the membrane to overshoot the mark, it will bulge out on the other side, and there will be a series of oscillations in the water, till it is brought to rest through loss of energy by friction in the pipes. Analogous electrical oscillations occur when a condenser is suddenly discharged through a wire circuit of small electrical resistance. Notice that the force which the pipes themselves exert upon the water is in the nature of a drag, or resistance. It increases when the velocity of flow increases, and becomes zero when the velocity is zero. On the other hand, the reaction of the membrane depends not at all upon the velocity of flow, but only upon how much water has been pumped out from one side into the other.

In the wires of the electrical circuit there is a real current, or flow of electricity, when the condenser is being charged and when it is being discharged, just as there is a real flow of water in the pipes when the membrane is being bulged out and when it is flattening itself again. Faraday regarded the condenser also as the seat of a flow of electricity, opposed by an elastic reaction instead of a mere resistance. Such a flow has been called a *displacement current,* as distinguished from the *conduction current* that takes place in the metal wires.

128. Maxwell's assumptions.—So far, Faraday's idea is merely a theory that affords a convenient way of thinking of electrical phenomena, but it becomes of great importance when we enquire whether displacement currents have the same relation to magnetic phenomena as conduction currents. Are displacement currents surrounded by lines of magnetic force?—Are displacement currents induced when there is a change in the lines of force threading through a circuit composed in whole or in part of non-conductors? An answer by direct experiment would be difficult, but Maxwell started out with the assumption that the answer to each of these questions is *yes,* and tried to see what conclusions this assumption would lead him to. The principal conclusion is the existence of electromagnetic waves in the ether of the kind we have already described.

129. Hertz's experiments.—The mechanism by which these waves propagate themselves can be seen by a consideration of the apparatus with which Hertz first studied them. The two ends of the secondary of an induction coil (S, figure 127) are led to the two halves of what is called the *oscillator.* This consists of two rectangles of sheet metal in the same plane, to each of which is attached a metal rod ending in a ball. The two balls are placed close enough together so that when the induction coil is operated a series of sparks pass between them. One may regard the rectangles of sheet metal as being the two plates of a condenser, the plates being opened and separated from one another. They still constitute a sort of condenser, but one of small capacity. So long as a spark is in existence between the balls, the spark, the balls and the rods form a short-circuit for the condenser, but before the spark is formed the plates are connected only through the coil S in the induction coil. The first action then, when the induction coil starts to work, will be to charge up the plates just as if the induction coil were a battery cell, except that they are charged to much higher potential. Let us say, for the sake of concreteness, that the upper plate is charged positively, the lower negatively. Evidently there will be what we have called displacement-currents in the space between the balls and also all around the two plates. When the difference in potential between the two plates becomes great enough, a spark starts between the balls, immediately establishing a short-circuit. The plates then discharge, and rapid electrical oscillations occur in the system, much more rapid, in fact, than the operation of the induction coil which starts them. The induction coil, for example, might make one or two hundred sparks per second, while the oscillations in the oscillator are at the rate of some hundred million per second, so that each single spark would include a large number of oscillations, though in each set of oscillations there is no doubt that the amplitude dies down rapidly, or, as we say, they are strongly *damped.* Each spark starts a series of oscillations which die out completely before the next spark occurs.

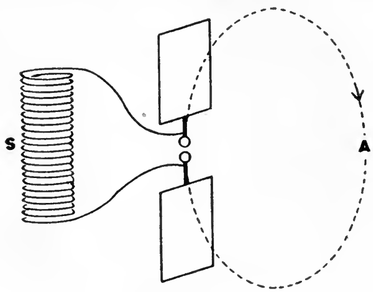

Figure 127

130. Propagation of electromagnetic waves.—Now, consider what happens in the surrounding space, supposing for convenience that the axis of the oscillator is vertical. When the first discharge occurs, there is a downward current through the spark which causes lines of magnetic force, clockwise as seen from above, to circulate in a horizontal plane about the oscillator. According to our second law of electromagnetic phenomena, the introduction of magnetic lines through any *conducting* circuit will induce a momentary current in the latter; and according to Maxwell's theory the same thing holds true also for a *non-conducting* circuit. Therefore, any circuit drawn at will so as to enclose some of these lines of magnetic force, such as the dotted line in the figure, which is drawn to represent a circle in a vertical plane, would be the seat of an induced current, whose direction is given by the arrow. This means that there would be an induced current *upward* at the spark (where its effect would be to diminish somewhat the downward current already there, and so account in part for the dying down of the amplitude of the oscillations) and *downward* at A. This downward current at A would in its turn produce lines of magnetic force in a horizontal plane, clockwise as seen from above. Their effect would be to neutralize the magnetic field between A and the oscillator, and to produce beyond A other induced displacement currents, which would in turn produce a magnetic field, and so on. Thus, the effect of the downward current in the oscillator is to cause downward displacement currents and lines of magnetic force in a horizontal plane, which progress farther and farther from the oscillator. The above explanation might lead the reader to draw the incorrect conclusion that the progression out from the oscillator comes in a series of steps, for an explanation of such a complicated phenomenon given without the assistance of mathematical analysis is necessarily crude and incomplete. In reality the progression is perfectly continuous and goes on at a velocity of which we shall speak later.

The downward discharge of the condenser is followed after a very small fraction of a second by a reverse discharge upward. This of course will also send out vertical displacement-currents and horizontal magnetic field, both of which will be in the reverse direction from the corresponding vectors caused by the initial discharge, but they will follow them in their progression out from the oscillator. The phenomena are repeated by the succeeding discharges, although the amplitude of the oscillations must become less and less since energy is constantly being sent out from the oscillator. Thus there will be a series of damped electromagnetic waves, for each time that the induction coil charges up the plates of the oscillator. It is evident that the waves are transverse and polarized, since electrical vibrations occur only in a vertical, magnetic only in a horizontal plane.

We have traced the causes of the wave propagation only for a direction at right angles to the axis of the oscillator, but waves are also sent out in other directions, though with less intensity. No waves at all are sent out in a direction parallel to the axis.

131. Velocity of the waves.—The general method which we have here used to show how electromagnetic waves are propagated does not enable us to predict the velocity of propagation. Evidently this will depend upon the electrical and magnetic properties of the free ether, or of whatever medium the waves are passing through. The mathematical treatment shows that the velocity is

$$V = \frac{1}{\sqrt{\mu k}}$$

where μ is the magnetic permeability of the medium and k is its dielectric constant (specific inductive capacity). In order to calculate the velocity from this formula, it is of course necessary that both μ and k be expressed in the same system of units, for two systems are used in the theory of electromagnetic phenomena, the *electrostatic* system and the *electromagnetic*. Since we must choose one or the other, suppose the latter is chosen, so that μ and k are both expressed in electromagnetic units. Now the electromagnetic system is based on the assumption that the *permeability* of the ether is 1, the electrostatic system on the assumption that the *dielectric constant* of the ether is 1; therefore in the above equation for the velocity we may substitute 1 for μ but we may not substitute 1 for k. Instead, we must find the value of k in electromagnetic units. When two charges e and e′ are placed x centimeters apart, the force between them is

$$F = \frac{ee'}{kx^2}$$

Since k is expressed in electromagnetic units, e and e′ must be also. If E and E′ be the values of these same charges in electrostatic units,

$$F = \frac{EE'}{Kx^2}$$

where K is now the dielectric constant in the electrostatic system. F and x remain as they were, for the unit of force and the unit of distance do not change from one system to the other. Therefore, we may say

$$\frac{ee'}{k} = \frac{EE'}{K}$$

Now let c be the number of electrostatic units of charge which are equal to one electromagnetic unit of charge. Then

$$E = ce \qquad E' = ce'$$

$$\frac{ee'}{k} = \frac{c^2ee'}{K}$$

$$\frac{1}{k} = \frac{c^2}{K}$$

$$k = \frac{K}{c^2}$$

Since K = 1 for the ether, we may write $k = 1/c^2$. Substituting this value for k and 1 for μ in the equation for the velocity, we get

$$V = c$$

The velocity of electromagnetic waves appears then to be equal to the ratio of the units of charge on the two systems of electrical units. This ratio has been measured with great care by a number of different experimental methods, and is found to be 3×10^{10}, the same value as the velocity of light in centimeters per second. This is also the value found experimentally by Hertz for the velocity of his electromagnetic waves.

From this close accord between Hertz's experimental results and Maxwell's theory on the one hand, and the known properties of light on the other, the electromagnetic theory of light seems fairly surely grounded. One is naturally led to enquire what sort of a mechanism sends out the extremely short waves that constitute visible light. Light always starts from some material particles, never, we believe, from an empty spot in the ether. Moreover, we have seen that the different chemical elements emit characteristic wavelengths. Consequently, the mechanism of emission for ordinary light, corresponding to the oscillator of Hertz or the elaborate transmission apparatus of wireless telegraphy, must be contained within the atom. Many phenomena not directly connected with light indicate that an atom contains positive and negative charges, and we are led to infer that it is the vibrations of one or more of these charges which start the electromagnetic waves. A mass of material becomes luminous whenever, as a result of high temperature or any other cause, the charges within some of the atoms are set into vibration with sufficient amplitude to produce, in the surrounding ether, waves strong enough to affect the eye. Waves coming from a single atom would be polarized, but from a great mass of luminous atoms, as in a flame, all planes of vibrations would be represented. If the waves strike against any material body, part of their energy is reflected and part enters. In some materials the waves that enter are absorbed before they have gone more than a short distance, their energy being converted into heat. Such materials are said to be *opaque* to the waves. Other materials, called *transparent,* allow the waves to pass through with little absorption. The velocity is always, with a few exceptional cases, slower in the material than in the free ether; and if the material is doubly-refracting there are two velocities—one for waves having the electrical vibrations in the principal place (extraordinary), and one for those having the electrical vibrations perpendicular to the principal plane (ordinary).

132. Refractive index and dielectric constant.—Of course, since $V = 1/\sqrt{\mu k}$, a change in velocity means a change in μ or in k. The permeability for all materials except the magnetic metals, iron, nickel, cobalt, etc., is practically the same as for the free ether, so that it must be a change in k, the dielectric constant, which causes the change in velocity. In fact, the index of refraction should bear a direct relation to the dielectric constant. For if V is the velocity of the light in the material in question, V′ that in the ether, k the dielectric constant for the material and k′ that for the ether, the index of refraction will be

$$n = \frac{V'}{V} = \frac{1}{\sqrt{\mu k'}} \div \frac{1}{\sqrt{\mu k}} = \sqrt{\frac{k}{k'}}$$

Since we have to do only with the ratio of two dielectric constants, it is immaterial whether we use the electromagnetic or the electrostatic units. If we use the electrostatic, $k' = 1$, $n = \sqrt{k}$, or $n^2 = k$. That is, the square of the refractive index is equal to the dielectric constant in electrostatic units. The following table shows how well this prediction of theory agrees with experimentally measured values for a few common materials. The refractive indices are given for yellow light of wavelength .00005893 cm.

	n	n^2	k
Carbon bisulphide	1.64	2.68	2.62
Turpentine oil	1.47	2.16	2.23
Benzene	1.50	2.25	2.29
Ice	1.31	1.71	2.85
Alcohol	1.37	1.88	25.8
Water	1.33	1.78	81

It will be noted that n^2 and k are nearly the same in value for carbon bisulphide, turpentine oil, and benzene, but quite different for ice and absurdly different for alcohol and water. This failure of agreement causes a suspicion that there is some defect in the theory, a suspicion that becomes all the greater when one reflects that even for a single substance the index varies with wavelength, while the dielectric constant, as we have defined it, has nothing to do with wavelength. The defect actually lies, not in the electromagnetic theory of light, but in the conception of the dielectric constant for material substances. This quantity, which is defined for steady electrostatic fields, requires some modifications when applied to such rapidly alternating fields as we are concerned with in light waves. These modifications will be considered in the next chapter.

CHAPTER XV.

133. Dispersion.—134. Electron theory of matter.—135. Electromagnetic dispersion formula.—136. Anomalous dispersion.—137. Reststrahlen.

133. Dispersion.—The preceding chapter ended with the citation of a case where the electromagnetic theory, in its simple form, disagrees decidedly with experiment. According to the theory, the index of refraction of water should be equal to the square-root of its dielectric constant, i. e., about 9, whereas it is in fact about 1.33 for yellow light and variable for different wavelengths. Similar wide divergences are shown by some other materials, and the index varies with the wavelength in all. Obviously, before the electromagnetic theory can be made worthy of acceptance, it must be supplemented in some way to account for dispersion, or the variation of index with wavelength. So far as the passage of light through the ether is concerned, the theory may stand as it is, for it is only in ponderable matter, made up of molecules and atoms, that the velocity is different for different colors, or wavelengths.

No two materials behave exactly alike in regard to dispersion, but so far as the visible spectrum is concerned most of them agree in this, that both the index and the rate at which the index changes are greater for short than for long waves. Cauchy proposed the formula

$$n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \qquad (1)$$

where A, B, and C are constants for one material, but have different values for different materials. The formula is rather useful when applied to prisms of glass, etc., for if we know the indices of a particular glass for three different wavelengths, then by substituting the values of λ and the corresponding values of n we get three separate equations in which A, B, and C occur as unknowns. If we solve, and substitute the numerical values in (1), we have an equation from which n can be found for other wavelengths. But the equation works satisfactorily only over a limited range of wavelengths, and has no sound theoretical basis.

134. Electron theory of matter.—The dispersion of materials can be accounted for on the electromagnetic theory, and the nature of the dielectric constant at the same time explained, if we take into account *the electron theory of matter,* brought into being as a result of researches partly in optics, partly in radioactivity and other branches of physics. According to it, every atom consists of a nucleus of positive electricity surrounded by a number of very small negative charges called *electrons.* All the negative electrons are believed to be exactly alike, even in different elements, and the elements differ merely in the number and arrangement of electrons and the corresponding magnitude of the positive nucleus. In a complete atom, the sum of all the negative charges equals the positive charge of the nucleus, so that the atom is electrically neutral, but under some circumstances one or more electrons may become detached from the atom and either remain free and unattached in space or become temporarily attached to some other atom or group of atoms. Such a free electron, or one attached to an otherwise neutral atom or group of atoms, is called a *negative ion;* while the atom from which it came, which is left with an excess of positive electricity, and which possibly also gathers neutral atoms about it, is a *positive ion.* Experiments on the leakage of electricity through gases indicate that there are always a number of ions present in any gas, though exceedingly few in comparison with the number of neutral atoms, and that certain agencies such as X-rays and radioactive substances cause a considerable increase in the number of ions. In metals the electrons are believed to be particularly free to become detached from their atoms and susceptible of great freedom of motion within the metal. This accounts for the fact that metals are invariably good conductors of electricity, for the electrons are supposed to carry the current. In nonconductors, on the other hand, the electrons are not easily detached from the atoms.

135. Electromagnetic dispersion formula.—The dielectric constant, or specific inductive capacity, is defined as follows: Imagine two small metal balls, one charged positively, the other negatively, e and e' being the respective charges and r the distance apart. Since the force of attraction is proportional to the product of the charges and inversely as the square of the distance, we can write it

$$F = \frac{ee'}{kr^2}$$

The factor of proportionality, k, is the dielectric constant, and its value depends upon the nature of the medium between and immediately surrounding the charges. If the medium is water, F has quite a different value from that for the ether, and therefore k is different. According to the electron theory, this decided alteration in the force, caused by interposing material atoms, is due to the production of what is called *electrical polarization.* In each atom the positive nucleus is pulled slightly toward the negatively charged ball, the negative electrons toward the positive ball. When the atoms are of such a nature that this polarization is great, the result will be a value of k differing greatly from that for the ether.

If the amount of polarization is proportional to the intensity of the electrical field producing it, k should be constant, and we find that it is so for steady electrostatic fields. But when light waves pass through a material the latter is, as we have seen, the seat of electrical fields which alternate very rapidly, and in such a case we should expect the so-called dielectric constant not to be constant at all, but highly variable. Particularly should this be the case if the waves happened to have a period somewhere near a natural period for the vibration of the electrons within the atoms; that is, if the wavelength happened to be nearly the same as that which the atoms could absorb very strongly, or could themselves emit if excited to luminescence. In such a case the wavelength is said to lie close to an *absorption-band* of the material. The electrons would be set into violent vibration by resonance, and the electrical polarization would oscillate through a very wide range.

In the theory of such a case, it is necessary to calculate a sort of *average*, or *effective,* value for the dielectric constant, which will of course be different for different wavelengths. The calculation is complicated, because we must take into account not only the amplitude of the electronic vibrations, but also the relation in phase between these vibrations and the light vibrations. For instance, if the period of the light-waves is shorter than the natural period of the electronic vibrations, the latter will be behind the former in phase, and conversely in the converse case. The full mathematical treatment is beyond the scope of this text, but the above considerations are sufficient to show that the effective value of k, and therefore the value of n, should be abnormally large or small for wavelengths close to an absorption-band. When the atoms have only one such band, the final formula is

$$n^2 = 1 + \frac{M_1\,(\lambda^2 - \lambda_1{}^2)}{(\lambda^2 - \lambda_1{}^2)^4 + b^2} \qquad (2)$$

Here n is the index for light of wavelength λ, λ_1 is the wavelength which the material absorbs most strongly, and b and M_1 are constants whose values depend upon conditions within the atom and the density of the material. If the atom has several absorption bands, the formula becomes

$$n^2 = 1 + \frac{M_1\,(\lambda^2 - \lambda_1{}^2)}{(\lambda^2 - \lambda_1{}^2)^4 + b_1{}^2} + \frac{M_2\,(\lambda^2 - \lambda_2{}^2)}{(\lambda_2 - \lambda_2{}^2)^4 + b_2{}^2} + \frac{M_3\,(\lambda^2 - \lambda_3{}^2)}{(\lambda^2 - \lambda_3{}^2)^4 + b_3{}^2} \qquad (3)$$

λ_1, λ_2, λ_3, etc., being the wavelengths of the several absorption bands.

136. Anomalous dispersion.—If, taking equation (2), we plot n as ordinate against λ as abscissa, we get a curve like figure 128, indicating very low values of n for waves slightly

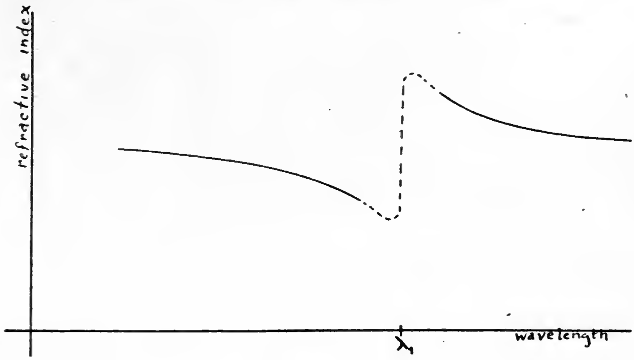

Figure 128

shorter than λ_1, very large values for waves slightly longer. There are certain substances, particularly the aniline dyes, which have very strong absorption bands within the visible region, and for these the dispersion curves are found by experiment to be of the form of this figure, except that the exceedingly strong absorption which these substances exert upon light whose wavelength comes anywhere close to the center of the absorption band (i. e., close to λ_1 in the formula) prevents us from taking measurements there. For instance, all that part of the graph which is indicated in dotted lines might be missing from the experimental curve. Such substances were said to have "anomalous dispersion," because the curve, with two branches, differs so much from the single curve obtained for colorless transparent materials in the visible region. We now know, however, that the complete dispersion curve of any material, taken over the complete range of wavelengths, would show an inflection like that of figure 128 at each strong absorption band. The curve for the visible region, in the case of different kinds of glass, quartz, etc., is similar to the right-hand branch of the figure, showing that these materials must have absorption-bands in the ultraviolet.

There are some cases of absorption which do not seem to influence dispersion in this way, for example the absorption of solutions of copper sulphate, potassium bichromate, etc. Such absorption is much weaker, however, than the kind shown by the aniline dyes.

A very interesting case of absorption as influencing dispersion is that of sodium vapor, which absorbs very strongly two wavelengths quite close together in the yellow. (See sections 53 and 56). It is possible to adjust a Bunsen flame fed with sodium so that it takes the form of a prism, and thus investigate its index of refraction for light of different wavelengths. Since the density of the vapor is quite small, the index is practically unity except for wavelengths very close to one or the other of these two absorption lines, that is light passes through it with practically the same speed as through air. An ingenious experiment due to Becquerel causes the vapor to draw its own dispersion curve. Suppose the sodium flame to be made to take a prismatic form, with the refracting edge of the prism below and horizontal, and placed in front of the slit of a grating spectroscope. Light from some source giving a continuous spectrum is sent through this flame, enters the slit of the spectroscope, and is spread out by the grating into a spectrum in a horizontal plane. Those wavelengths for which the sodium vapor has a high index of refraction will be bent upward by the flame-prism, and so will be raised above the general level of the spectrum, while those for which the index is low will be sent downward and brought below the general level. Figure 129 shows the actual appearance under such circumstances. Each of the absorption lines produces a break in the dispersion-curve similar to that shown in figure 128, affording a very satisfactory verification of the theory.

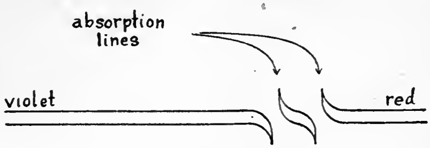

Figure 129

137. Reststrahlen.—There are a large number of cases where a material has the power to resonate to a certain period of vibration so that we should expect it to absorb very strongly light of the corresponding wavelength, but instead it *reflects* this wavelength with extraordinary power, in some instances as strongly as a silvered mirror reflects ordinary visible light. The reflection in such cases is a true surface reflection, not the sort of diffuse reflection from within the material that occurs in the case of paper and many other more or less translucent materials, and which, coupled with some absorption, is responsible for the colors of most natural objects. This selective surface reflection, due to resonance, is particularly common among substances having strong resonance bands far out in the infrared, such as quartz, rocksalt, sylvite, potassium bromide, potassium iodide, and other crystals or fused salts. This property enables us to isolate and study some very long waves which would otherwise be difficult to detect. The method employed is to take the radiation from some source such as a Welsbach mantle burner and reflect it again and again from polished surfaces of the material being investigated. The light obtained from a single reflection contains, beside the wavelength selectively reflected, light of other wavelengths which are reflected to a lesser degree, but the latter are almost completely eliminated after a number of reflections, while the former remain almost as strong as at first. The Germans have called waves isolated in this manner "Reststrahlen," and this name has been pretty generally adopted into other languages, though the English equivalent "residual rays" is also used. By this means light-waves as long as .00965 cm. (nearly 1/10 mm., more than 100 times as long as the deep red) have been isolated from the radiation of a Welsbach burner by using potassium iodide for the reflecting material.

The student may wonder why, if these wavelengths are already present in the light from a Welsbach burner, they could not be detected easily and simply by sending the whole beam through a prism, or dispersing it with a grating, as we would for shorter wavelengths, without resorting to the residual ray method for first isolating them. Prisms are useless for such a purpose because many prisms would absorb such long waves, and even if that did not occur we could hardly hope to determine wavelengths from the deviation by the prism because the course of the prism's dispersion curve so far in the infrared would not be accurately known. The trouble with using a grating lies in the fact that a grating gives many spectra which overlap, and a wavelength such as we are concerned with would come in the same place as many other shorter waves. For example, a wavelength of .00965 in the first spectrum would coincide with .004825 in the second, .003217 in the third, .002412 in the fourth, and many others, and it would be impossible from such a complex to tell what wavelengths were actually present. After the method of reflections from a selectively reflecting material have been used to isolate the waves, a coarse grating or a specially constructed interferometer can be used to measure the wavelengths. Of course the actual detecting device must be a bolometer, thermopile, or some similar absorbing and heat-recording instrument.

It was found that quartz, although opaque to some shorter waves, is rather transparent to the residual rays from potassium iodide, and has for such waves the very large refractive index 2.2. A quartz lens then would have for such long waves a focal length much less than for shorter waves. Rubens and Wood made use of this fact for isolating these waves and even longer ones, their method being in principle as shown in figure 130. A is a Wels-

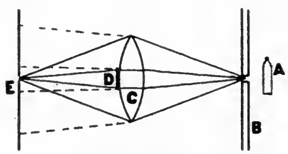

Figure 130

bach mantle, emitting the radiations, B a double metal screen with a small hole, C a quartz lens, E another screen with a small hole. E is placed at the proper focus for the very long waves to be investigated. For visible and the shorter infrared waves, the focal length of the lens is so long that the hole in B comes within the principal focus. Therefore such waves are diverged by the lens, and none of them would get through the hole in E except such as came through close to the center of the lens, and these were cut out by fastening there a small opaque disc D. It is clear that by this method not only the particular waves strongly reflected by potassium iodide would get through, but also other waves having about the same index of refraction, which would be lost in the reststrahlen method. In this way, they were able to obtain waves as long as .0107 cm. from a Welsbach mantle, and by using a mercury arc in a quartz tube instead of the Welsbach burner Rubens and von Baeyer detected waves of .0343 cm., the longest waves yet found in the radiation from a light-source. Waves from electrical oscillations in constructed apparatus (i. e., waves of the Hertz type) have been obtained as short as .2 cm., leaving only a very short gap for undetected wavelengths. Doubtless waves of length suitable to fill in this gap exist, and it only remains to detect and measure them.

Problems.

1. Calculate the constants A, B, and C, of equation (1) of the preceding chapter, for glass whose indices are 1.7774 for λ .00004713, 1.7582 for λ .00005600, and 1.7444 for λ .00006708. Then calculate the indices for wavelengths .00005016 and .00005893.

2. Show that the middle point in a spectrum as given by a prism of the glass of problem 1 would correspond to a much shorter wavelength than that in a spectrum of the same source as given by a grating.

3. From a consideration of the above two problems show why, in general, spectra produced by prisms are, relatively speaking, abnormally bright in the longer wavelengths.

CHAPTER XVI.

138. Production of X-rays.—139. Their properties.—140. Are X-rays ether waves?—141. Crystal reflection of X-rays.—142. Measurement of wavelengths. Crystal structure.—143. X-ray spectra.—144. The K and L series.—145. Quantum theory applied to X-rays.—146. Secondary X-rays. Absorption.—147. Total range of ether waves.

138. Production of X-rays.—As soon as the discovery of X-rays was announced by Roentgen in 1897, speculation began as to whether, like light, they consisted of ether-waves; but all the evidence acquired for a long time was conflicting on this point. Before discussing it, a summary of what was known of the properties of the rays a few years ago will be given.

They arise when an electric discharge is sent through a glass tube which has been exhausted to a very low pressure, or as it is sometimes stated, to a "high vacuum." At such pressures the discharge in a gas is carried largely by the "cathode rays," which had been shown before Roentgen's discovery to consist of a stream of electrons shot out with very high velocity from the cathode, or negative terminal of the tube. The green color, like a fluorescent glow, which is seen during the discharge, is caused by the impact of the cathode-ray electrons upon the glass walls of the tube. It was soon found that the particular place where the electrons strike is the source of the X-rays, and the tubes are arranged so that they impinge upon a sheet of metal called the "anticathode," or "target," which may or may not be metalically connected with the positive terminal of the tube. The X-rays then stream out in all directions from the anticathode as shown in figure 131, where K is the cathode, T the anticathode, A the anode. The student is warned against thinking that X-rays are reflected cathode rays. This cannot be, since cathode rays are negatively charged particles, and X-rays are something entirely different, whatever they may be.

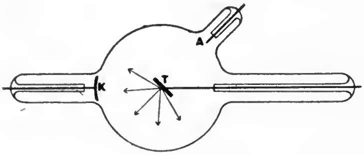

Figure 131

139. Their properties.—From the standpoint of the general public, the most striking characteristic of the rays is their ability to pass readily through flesh, somewhat less freely through bone, and to some extent even through metals, all of which are opaque to light. But to a person of analytical mind this fact should be no more surprising than that light will pass through glass. Glass is transparent to light and flesh is transparent to X-rays, and one of these facts calls for explanation exactly as much as the other.

A characteristic that is more interesting to the physicist is the *ionizing* power of the rays. Any gas through which they pass becomes for a short while a relatively good conductor of electricity, showing that the passage of the rays causes many molecules to separate into positive and negative ions.

The rays affect a photographic plate very much as light does, and they cause certain mineral salts to fluoresce strongly, though they themselves do not stimulate vision in the eye.

X-rays are not subject to refraction, but pass straight through a prism or lens without any bending. Neither are they regularly reflected from a polished surface as light is. However, they are to some extent scattered or diffused in going through matter, very much as light is diffused in going through an attenuated fog. The scattering occurs not only at the surface, but also throughout the material, even when the latter is a dense solid.

A grating has no effect upon them. They pass right through it as they would through any material, with some scattering and absorption. An attempt was made to produce diffraction by passing X-rays through a very narrow aperture, or slit, the jaws of which were made of lead, which is very opaque. Under similar treatment light, as we have seen in section 72, spreads out after passing the aperture, and forms a set of bright and dark bands, whose distance apart depends upon the width of the aperture and the wavelength of the light. The narrower the aperture and the longer the wavelength, the wider apart are the bands. It was hoped by this means to show that X-rays also consist of waves, and to measure their length, but it was found that if they spread out at all it was to such a slight extent that the wavelength, if there is any such thing, must be something like 1/5000 that of yellow light, or less.

X-rays are not all alike, for some will penetrate a material like aluminum far more readily than others. Rays of great penetrating power are said to be *hard,* those of slight penetrating power *soft.* The hardness of the rays depends mainly upon two factors, the material of the anticathode and the degree of exhaustion of the bulb.

When a stream of X-rays from a vacuum-tube like figure 131 falls upon a metal, under certain circumstances the latter gives off electrons and also new X-rays, different from the scattered rays, called secondary *X-rays.* It is very remarkable that the velocity with which these electrons are sent off is nearly the same as that of the electrons of the cathode ray stream which produced the original X-rays, and is entirely independent of whether the incident rays be strong or weak.

After the discovery of radioactive substances, with their three types of rays known as α, β, and γ, it was soon found that the γ-rays are similar in all respects to X-rays, except that they are more penetrating than the hardest rays obtained from a vacuum-tube.

Professor Marx, of Leipsic, carried out a very ingenious experiment by which he claimed to prove that X-rays travel with the same velocity as light; but his method was necessarily very complicated, and he was not able to convince his fellow investigators that his conclusion was justified. No accepted measurement of the velocity of X-rays has yet been made.

140. Are X-rays ether waves?—The fact that X-rays are not refracted does not necessarily prove that they are not ether-waves, for it is conceivable that exceedingly short waves would travel through matter with the same velocity as through empty space. Neither is the failure of regular reflection an objection, when we know that the distance between the atoms of a solid body is something of the general order of 10^{-8} to 10^{-7} cm., that the wavelength of visible light is in the neighborhood of 5×10^{-5}, and that if X-rays have a wavelength it is probably less than 10^{-8}. A long string of atoms could lie within a single wavelength of yellow light, while on the other hand a number of wavelengths of X-rays might lie between two adjacent atoms. A surface that we may regard as finely polished for visible light, would therefore be an exceedingly rough structure for the very short waves. In the discussion of reflection in chapter IV, it was stated that *each point* in the straight line MN of figure 23 representing the reflecting surface becomes the center for a secondary wavelet, but it would no doubt have been more in accordance with facts if it had been said that each *atom* of the surface became such a center. Figure 132 represents what

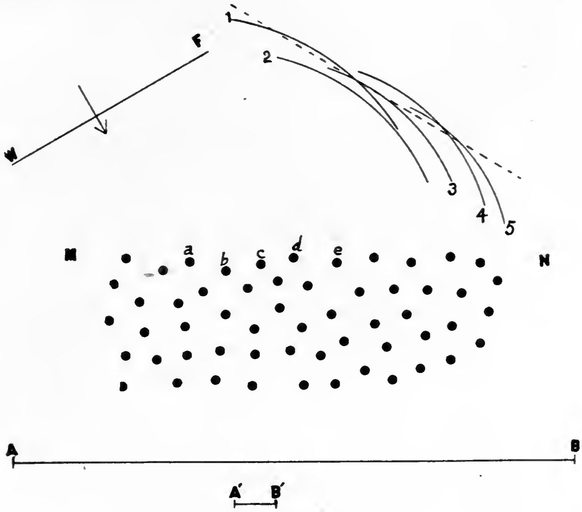

Figure 132

a reflecting surface might look like if it were magnified till individual atoms became clearly perceptible. The continuous straight line MN of figure 23 has become a somewhat irregular row of atoms forming the top layer of a body composed of many such atoms. When the wavefront WF strikes any one atom it becomes a center for a secondary wavelet. 1, 2, 3, 4 and 5 are the respective wavelets from the atoms a, b, c, d, and e. Since these atoms are not on a straight line, the secondary wavelets do not lie tangent to a straight line, but if the wavelength of the light is as long as AB the amount by which each wavelet falls off a common tangent such as the dotted line is a very small fraction of a wavelength, and therefore the secondary waves are nearly in phase along this line. Therefore, for visible light we are justified in treating the polished surface of a solid as if it were a continuous surface. But if the wavelength were as short as A′B′ the conditions would be very different, for then any one wavelet might miss the hypothetical reflected wavefront by half a wavelength or more. Therefore there would be no general agreement in phase of the secondary wavelets along any line which could act as a reflected wavefront; on the contrary the secondary wavelets would have a general tendency to annul one another's effects, leaving only a weak resulting effect in any direction, which might account for the diffuse scattering of X-rays.

141. Crystal reflection of X-rays.—But there is one circumstance under which we might expect a sort of regular re-

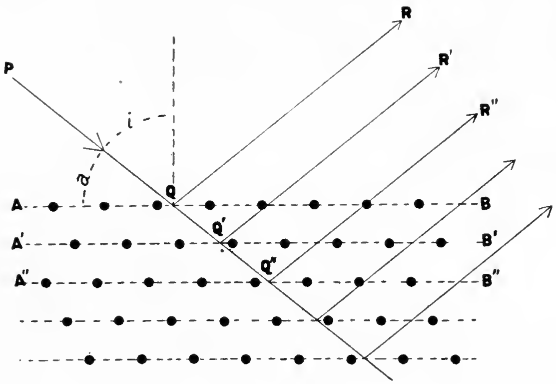

Figure 133

flection even for exceedingly short waves, namely, when the atoms are arranged in regular plane layers, as in a crystal. This idea occurred to Professor Laue, of Zurich, and at his suggestion Friedrich and Knipping tried the effect of several different crystals in the path of a beam of X-rays. They found that reflection does occur, not necessarily from the surface of the crystal, but from every plane within the latter along which atoms are distributed in a regular pattern.

Laue's explanation of the phenomena observed by Friedrich and Knipping is unnecessarily involved, and is in fact less useful than the following method, which is due to Messrs. W. H. and W. L. Bragg, who have done a great deal of experimental work with X-rays. Their book, entitled "X-rays and Crystal Structure," gives the best account of the subject yet published.

In figure 133 let AB, A′B′, A″B″, etc., represent the rows, or rather the contours of the planes, in which atoms are regularly arranged. Let PQ be a ray of incident waves. When the wavefront strikes each point in the layer AB, secondary wavelets will be sent out, and since the atoms lie in a plane, the secondary wavelets will all be tangent to a plane, and there wili be a reflected wavefront for which the ray QR is drawn, the angles of incidence and reflection being equal. But this is not all for the incident waves will pass on and strike the next layer of atoms A′B′, which will also give rise to a reflected wavefront represented by the ray Q′R′. In the same way one layer after another would take up the act of reflection, until, after penetrating to a sufficient depth within the crystal, the incident waves are too much weakened to produce sensible reflection. Here we have to do with reflection from many parallel surfaces, spaced at equal intervals. The condition is really quite analogous to the reflection of light from a material film of uniform thickness (see section 76 and figure 80), although there are certain important differences. Here we have to do with a series of reflected beams produced by reflection from many different planes, rather than between only two planes. All the reflections in the present case are of the same kind, so that there is no occasion for a difference in phase due to reflection alone. Moreover, the index of refraction is to be regarded as having the value 1, and the angles of incidence and refraction are equal. Bearing these facts in mind, the difference in path between rays PQR and PQ′R′, or between PQ′R′ and PQ″R″, etc., can be gotten from the analogous case of figure 80. It is simply

$$2t. \cos r, \text{ or } 2t. \cos i.$$

t being the perpendicular distance between successive layers of atoms. X-ray workers have found it more convenient to express their results in terms of what they call the "glancing-angle," a, than in terms of i, the angle of incidence, and therefore the difference in path is usually written

$$2t. \sin a$$

If this quantity is equal to a wavelength, or to any whole number of wavelengths, the reflections from the different layers will reinforce one another, and the reflection will be strong. The formula for strong reflection

$$2t. \sin a = n\lambda$$

reminds one of the formula for the grating, and indeed a crystal does provide a natural grating for exceedingly short wavelengths; but it differs from the ordinary grating in that the regular spacing extends in three different directions instead of only in one, and this fact makes important differences. With the ordinary grating, the angle of incidence of the light may be zero, and we look for various wavelengths at various angles; but with the three-dimensional grating, or "space-lattice," the angle of incidence as well as the angle of reflection must have a particular value, and only one wavelength can be shown at a time.

142. Measurement of wavelengths. Crystal structure.— The procedure in investigating the wavelengths of X-rays is, in crude outline, as follows. The X-ray tube, such as that shown in figure 131, is enclosed in a lead box with a narrow slit at a certain place, so that only a very fine beam of the rays gets out through the slit. A crystal is held in the path of this beam, and slowly turned so that the angle a is changed. At the same time, a photographic plate, or a device for measuring the ionization produced by the rays, is turned about the same axis at twice the rate of motion, so as to always be in position to receive the reflected rays if there are any. At certain positions a blackening of the plate or a functioning of the ionization apparatus shows a reflection, and the angle a is measured. In addition to this angle a we must of course know the distance t between the layers of atoms. The manner in which this is found can be best understood from the following example, the special crystal being rocksalt. It is composed of the atoms of sodium and chlorine in equal numbers (formula NaCl) and its density is 2.17. From considerations in regard to its crystal form, which are well supported by the experiments with X-rays, we are convinced that the atoms in this crystal are arranged in a cubical pattern as illustrated in figure 134, where

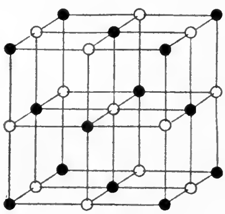

Figure 134

the white circles represent sodium atoms, the black chlorine atoms. To each small cube of the structure one atom must be assigned, as can be readily understood if the reader notes that cubes of the same size can be drawn with an atom at the center of each, the cubes filling the entire space. Therefore, since t is the length of each side of a little cube, each atom commands a space of volume t^3, and in one cubic centimeter of crystal there are $1/t^3$ atoms. Since one cubic centimeter has a mass of 2.17 grams, the average mass of the atoms must be

$$2.17 \div 1/t^3 = 2.17t^3$$

Now the atomic weight of sodium is approximately 23, and that of chlorine 35, so that the average atomic weight is 29, that is the average mass of the atoms making up the crystal is 29 times that of the hydrogen atom. From a number of physical and chemical lines of attack, we know the mass of the hydrogen atom to be 1.64×10^{-24}, therefore the average atom of rocksalt has mass

$$29 \times 1.64 \times 10^{-24} = 47.56 \times 10^{-24}$$

We can then put

$$2.17\ t^3 = 47.56 \times 10^{-24}$$

$$t = 2.80 \times 10^{-8}$$

One might be inclined to wonder whether it is not a *molecule,* rather than an atom, which is located at each corner of a cube of the crystal lattice. The way in which it was proved that it is a matter of atoms rather than molecules will not be taken up here. It is fully explained in Bragg's book, from which most of the facts of this chapter are taken.

Having the value of t, it is possible to find the wavelength of the X-rays emitted by any anticathode. A palladium anticathode is found to emit rays of wavelength $.576 \times 10^{-8}$ as well as certain other wavelengths. One wavelength being known, it is now possible to reverse the procedure outlined above, and find the structure of crystals less simple than rocksalt. Thus we are provided with a new method, by which we can determine the inner structure of crystals as well as measure the wavelengths of X-rays, although it is only in the latter that we are at present interested.

143. X-ray spectra.—We carry over into the discussion of X-rays a number of the terms familiar in connection with visible light. Thus we speak of the X-ray "spectrum" of a material, meaning the total array of wavelengths which it emits when acting as the anticathode of an X-ray tube. When a spectrum extends over a considerable range of wavelengths, without the absence of any wavelength within that range, we say it is "continuous." On the other hand, a spectrum consisting of several distinct wavelengths only is said to be made up of so many "lines." Any metal serving as anticathode emits a continuous spectrum, but also, if the vacuum be good enough and the consequent speed of the cathode rays great enough, certain characteristic lines, much more intense than the continuous spectrum. Each element emits a group of lines similar to those characteristic of other elements, but the higher the atomic weight the shorter the wavelengths in the group. If we pick out corresponding lines for different elements, calculate their frequencies from the general relation that frequency equals velocity of light divided by wavelength, and then plot the square-roots of these frequencies as abscissæ against the "atomic numbers"* as ordinates, the points lie very accurately upon a curve which is nearly straight, that is, the square roots of the frequencies are proportional to the atomic numbers.

144. The K and L series.—Certain elements of neither very high nor very low atomic number yield two groups of lines, the K series and the L series, of which the former is constituted of shorter wavelengths. For elements of lower atomic numbers only the K series has been found, for members of higher numbers only the L series. The rule of proportionality between square-root of frequency and atomic number holds for each series separately. If this rule be carried down to the case of hydrogen, for which naturally X-rays cannot be obtained in the usual way, it indicates that the K series for hydro-

*The atomic number is the ordinal number of the element when the whole list of elements is arranged in the order of increasing atomic weight. For elements of smaller atomic weights, it is nearly equal to half the atomic weight. Recent investigations show that the atomic number is far more indicative of an element's physical and chemical properties than the atomic weight.

gen should come in what we have hitherto regarded as the extreme ultraviolet region, where part of a series of lines has actually been discovered by Lyman; also, that the L series should about correspond to the hydrogen series in the visible spectrum shown in figure 62, for which Balmer's series-formula applies. That is, hydrogen being the lightest known element, its X-ray spectrum seems to be composed of wavelengths so long as to bring the lines in the visible and the extreme ultraviolet regions. Possibly the ordinary visible series spectra of the heavier elements could legitimately be regarded as representing X-ray series of longer wavelength than the L-series.

145. Quantum theory applied to X-rays.—Mention has been made in section 59 of the quantum theory, introduced by Planck to explain the radiation of an absolutely black body. There appears in this theory a constant h, numerically equal to 6.55×10^{-27}, whose significance is that a radiating center which sends out light of frequency v can do so only in energy-amounts equal to a multiple of hv. Curiously enough, this same constant appears in X-ray phenomena, for it is found that an X-ray tube can emit the radiations characteristic of its anticathode metal, of frequency v, only when the velocity of the electrons constituting the cathode-ray stream is great enough so that their energy is at least equal to hv. Of course we have no rational explanation for this fact, or for any other case in which the constant h appears.

146. Secondary X-rays. Absorption.—It was stated earlier in this chapter that characteristic secondary X-rays are produced under certain circumstances when primary rays from a vacuum-tube fall upon a metal outside the tube. These secondary rays have the wavelengths of the K or the L series appropriate to the metal which emits them. These rays are produced only when the primary rays have a somewhat higher frequency (shorter wavelength).

The absorption of X-rays by a given material is different according to the wavelength of the rays, just as we should expect. But the wavelength most strongly absorbed is not the same as that which the material could under proper stimulus emit, as we should infer from analogy with phenomena in the visible region (absorption by sodium vapor, or absorption by the sun's atmospheric gases of those wavelengths which they can emit).

For X-rays, the emitted wavelengths are always longer than those most readily absorbed, as in the phenomenon of fluorescence.

147. Total range of ether waves.—The facts related in this chapter show beyond reasonable doubt that X-rays are ether-waves of the same general character as visible light, though of exceedingly short length. They may be regarded as ultraviolet carried to the extreme, though the term ultraviolet is usually meant to include only waves stimulated by the same general methods used to produce visible light, such as high temperature, ordinary electric spark discharges between metal points close together, discharge through gases at pressures not excessively low, etc. There is still a gap between the shortest ultraviolet waves produced by such means and the longest X-rays produced either directly or indirectly by cathode-ray discharge, but we are now acquainted with an extraordinarily large range of ether waves, from the exceedingly short γ-ray waves emitted by radioactive substances at one end of the scale to the waves familiar under the head of wireless telegraphy at the other; and such gaps as exist in the range are relatively small. The following table gives the salient points in the range of ether-waves, according to length, as we now know them:

	wavelength, in cms.
Shortest known waves (γ-rays from radium)	10^{-9}
Longest X-rays from cathode-ray discharge	1.2×10^{-7}
Shortest ultraviolet from spark discharge	2.7×10^{-6}
Shortest visible	3.5×10^{-5}
Longest visible	7.0×10^{-5}
Longest infrared (from mercury-arc lamp)	3.4×10^{-2}
Shortest waves from electrical oscillations in constructed apparatus	2.0×10^{-1}
Approximate length used in wireless	10^{5}
Longest waves attainable	no limit.

CHAPTER XVII.

148. Review of the development of light-theory.—149. Lines of modern investigation. 150. The Zeeman effect.—151. Lorentz's theory. 152. The Stark effect.—153. The photo-electric effect.—154. Atom-models. —155. Bohr's theory of the hydrogen atom.

148. Review of the development of light-theory.—The development of the theory of light given in the preceding chapters follows roughly the chronological order of the history of the subject. It is of interest to notice that there are a number of stages in the development, with rather clearly marked dividing points.

Let us turn our attention first to the old corpuscular theory. It could never have held its place so long, but for the reverence in which the authority of Sir Isaac Newton, its chief advocate, was held, long after his death, particularly in England. Considerable progress was made under this theory, particularly in regard to colors and the phenomena of reflection and refraction.

Nevertheless, the final adoption of the wave theory was a distinct break, for it offered satisfactory explanations of such phenomena as interference and diffraction, which were serious stumbling-blocks to the corpuscular theory. Also, it explained reflection, refraction, and color differences in a more rational manner, brought about the invention of gratings and interferometers, and thus led to the detailed and accurate study of spectra.

The discovery of polarization may be said to have introduced a third stage, for it proved that light waves, of whose nature nothing had previously been known except their length, were not longitudinal but transverse. A period of active speculation as to the nature of the ether and ether waves followed, leading to the "elastic-solid" theories mentioned in chapter XIV. These very ably worked out theories are excellent examples of the application of mathematical analysis to physical phenomena.

Meanwhile, the mathematical theory of electrical phenomena was also being perfected, and this, as we have seen in chapter XIV, enabled Maxwell to predict the existence of electromagnetic waves and announce the electromagnetic theory of light, which fully supplanted the older theories after the experiments of Hertz. This brought the subject to the fourth, and so far the final, stage.

It must be admitted that, at the present time, the electromagnetic ether-wave theory of light does not stand on an absolutely secure foundation, for certain experimental facts cast some doubt upon it,—namely, those facts that gave rise to the relativity theory and the quantum theory. Strange as it may seem, the old corpuscular theory of light could account very nicely indeed for the Michelson-Morley experiment (chapter IX) and for some of those phenomena in which the constant h of the quantum theory appears. In fact, one might regard the relativity theory and the quantum theory as hypotheses introduced to make the wave theory square with these facts, and the necessity for such additional hypotheses may be regarded as a weakness in the structure. On the other hand, the corpuscular theory is condemned by interference, diffraction, and polarization, which seem impossible to explain except in terms of waves.

149. Lines of modern investigation.—Since the introduction of the electromagnetic-theory, progress has been very rapid, and has followed several general lines. First, our knowledge of the range of ether waves has been greatly extended, not only by thrusting farther out into the ultraviolet and infrared, but also by an abrupt jump into the region of very short waves (X-rays) and another into that of very long waves (Hertz waves).

Another line of attack has been the study of the radiation from an absolutely black body (see chapter VII). This has involved a great deal of very careful experimental work, as well as difficult theoretical study, leading, as we have seen, to the formulation of the quantum theory.

Third, a number of previously unknown phenomena have been discovered and investigated, partly optical and partly electrical in nature. These have to do with the emission and absorption of light, that is with the relation between radiated energy and matter, and they are interpreted in terms of the electron theory of matter. The most striking of these phenomena is X-rays. Since this subject has been dealt with at some length as the special subject of chapter XVI, we shall not discuss it further here, but proceed to a brief discussion of a few of the others.

150. The Zeeman effect.—Long before Maxwell formulated the electromagnetic theory of light, Michael Faraday, who had discovered that a beam of light passing through a magnetic field has its plane of polarization rotated, conceived the idea of placing a source of light directly in the magnetic field, to see whether the latter produced any effect upon the spectrum lines. Although we now know there *is* such an effect, Faraday was unable to detect it, because the spectroscopes then available were not efficient enough. Rowland also tried the experiment without success. In 1897 it was tried again by Zeeman, using a strong magnet and a good grating spectroscope. The effect which was discovered has been named the "Zeeman effect."

In figure 135 let MM be the two parts of a strong electromagnet, each terminating in a conical pole-piece, leaving a small space between, where the magnetic field is very strong. The source of light, S, is placed in this strong field. The slit of the spectroscope may be placed either at s_1 (broadside position), or at s_2 (end position). In the latter case a hole must be bored through one pole-piece and magnet core, to let the light through. In either case, an image of the source is focussed upon the slit by means of a lens. The results found in the earlier experiments of Zeeman, and of others who took up this study, may be stated as follows: In the broadside position, each spectrum line is changed by the presence of the magnetic field into a group of three lines very close together (triplet). The middle line of the triplet is polarized in a plane perpendicular to the plane of the figure, and the two outer

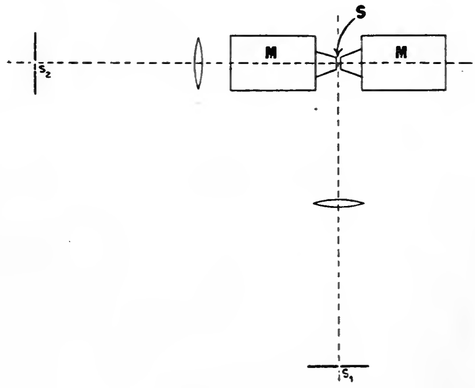

Figure 135

lines are polarized in the plane of the figure. In the end position, the magnetic field causes each spectrum line to become a close *doublet,* consisting of two lines, one of which is circularly polarized in the right-hand direction, the other in the left-hand direction.

151. Lorentz's theory.—These results were given a very clear explanation in terms of the electron theory by H. A. Lorentz. He supposed the light to be sent out from an atom by the vibrations of electrons within the atom. Each electron was conceived to be bound to the center of its path by a sort of elastic force, that is a force that varies directly as the distance. Suppose OX, OY, and OZ, figure 136, to be three directions perpendicular to one another. We take one of these, OX, to represent the direction of the magnetic field, and suppose O to be the center about which an electron vibrates. The direction of vibration might be anything, so we choose a direction OR at random, the length OR representing the amplitude of the vibration. We want to find how such a vibration would be affected by the presence of a magnetic field in the direction OX. To do this, it is best to resolve the vibration into certain components. In the first place, a vibration of amplitude OR is equivalent to one of amplitude OP parallel to OX and one of amplitude OQ in the plane of OY and OZ. (OQ and OR lie in a plane through OX perpendicular to the plane of OY and OZ, and the figure OPRQ is a rectangle.) The vibration of amplitude OQ can again be resolved into a right-handed and a left-handed circular motion about OX as axis, as was done by Fresnel in explaining the rotation of the plane of polarization by quartz. (See figure 120, chapter XIII,) Thus we may think of the single electron as replaced by three different electrons, of which one vibrates back and forth in the direction OX, one rotates about OX in one direction, and the third in the opposite direction. Since an electron is a negatively charged body, a moving electron constitutes a current. A magnetic field has no effect upon a current in the same direction as the field, and therefore the vibration along OX goes on just

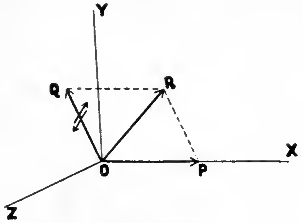

Figure 136

as if the field were absent. The two rotary motions, however, constitute currents across the field, and it is well known that the effect of the field upon such a current is to exert a force on the latter, perpendicular both to the field and to the current, that is, either toward the center O or away from it, according to the direction of the rotation. That is, one of the electrons rotating about OX as axis will have the elastic force pulling it to the center somewhat *strengthened* by the addition of a magnetic pull in the same direction. This electron will therefore have its angular speed increased, its period diminished. The electron rotating in the opposite direction will have its pull toward the center somewhat *weakened* by the magnetic action. It will therefore rotate somewhat slower, that is with a longer period.

The vibration along OX sends out waves of the same period, and the same wavelength, as the vibration in absence of a magnetic field. These will have their electric vibrations parallel to OX, so that their plane of polarization will be perpendicular to OX, i. e., in the plane of OY and OZ. They will pass out in all directions except the positive or negative direction of the X-axis, the axis of the magnet, but most strongly in the YZ plane, so that light of this wavelength would be observed in the broadside position, but not in the end position. The circular vibration would send out circularly polarized light in both directions along the X-axis, and elliptically polarized light in every other direction except directions in the YZ plane. In this plane, since the circles in which the electrons rotate are seen edge-on, plane-polarized light would be the result. The direction of the electric vibrations would be perpendicular to the magnetic field, so that the plane of polarization would be parallel to the field. The wavelength would be slightly shorter in one case, and slightly longer in the other, than the natural wavelength. Thus, the three different wavelengths seen in the broadside position, and the two in the end position, are fully accounted for, even to the state of polarization. Moreover, the theory indicates that the change in wavelength of the two outer lines of the triplet depends upon the ratio of the charge upon the electron to its mass, and the measured change gives about the same value to this ratio that was known to be right from electrical experiments with cathode rays.

Unfortunately for Lorentz's theory, it was soon found, when stronger magnetic fields were available, that many lines split up into four, five, and even more different component lines in the field. In fact, the simple triplet seems rather the exception than the rule. Lorentz's theory was incapable of explaining these more complicated cases of the Zeeman effect; and the only theory which has been applied to them with any degree of success is the theory of Ritz, which will be mentioned further on in this chapter.

152. The Stark effect. A phenomenon somewhat analogous to the Zeeman effect was discovered by Stark. It is a similar splitting up of the spectrum lines when the source of light is placed in a strong *electrostatic* field. It has not been studied to the same extent as the Zeeman effect.

153. The photo-electric effect.—Hallwachs discovered in 1888 that a negatively charged zinc plate, if illuminated with ultraviolet light, would lose its charge. This phenomenon, the photo-electric effect, is also shown by other metals, and with some even visible light is effective. The action of the light is to cause electrons to be shot out from the plate. The phenomenon is difficult to work with, because it is necessary for the surface of the metal to be perfectly clean and free from tarnish, and most metals tarnish under the action of the air fast enough to make the effectiveness decrease very rapidly. In some experiments, the metal and other parts of the apparatus were enclosed within a highly exhausted vacuum-tube, which also contained a device whereby a thin shaving could be cut off the metal at any time, leaving a fresh clear surface.

Suppose that the illuminated metal, which we shall call A, is placed face to face with another metal plate B, which is shielded from the ultraviolet light, so that B may receive the electrons shot off from A. If B is kept at zero potential by being connected with the earth, while A is kept at a positive potential V, then, provided V be high enough, the electrons will be kept from reaching B by the attraction of the positive charge upon A, which will draw them back. The magnitude of the potential V which is just sufficient to accomplish this is very simply related to the speed with which the electrons are emitted. If v represents this speed, and m the mass of an electron, the kinetic energy with which they leave A is $mv^2/2$. If e represents the charge upon the electron, the expenditure of work necessary to take it from a potential V to potential O is equal to the product of charge and potential-difference, or eV. If they are just stopped before reaching B, the kinetic energy $mv^2/2$ is just lost in doing this work, therefore

$$eV = mv^2/2$$

Since the values of m and of e are well known from electrical experiments, this equation enables us to find v by measuring the potential V just sufficient to prevent any current passing between A and B, for the transfer of electrons constitutes a current.

It has been found by such experiments that the velocity of the emitted electrons is exactly the same, whether the beam of ultraviolet light be strong or weak, although of course the number emitted per second is greater with stronger illumination. (Notice that a similar relation was found to hold for the emission of electrons and secondary X-rays, under the stimulus of another beam of X-rays, chapter XVI). The velocity does depend, however, upon the wavelength of the light. The shorter the wavelength (i. e., the higher the frequency), the greater is the velocity, and the relation between the electron velocity and the frequency of the exciting light involves the constant h of the quantum theory.

$$mv^2/2 = h\nu - a$$

where ν is the frequency and a is a constant characteristic of the particular metal employed. The equation may be interpreted in the following manner: As soon as an atom has absorbed from the incident light a quantum of energy, i. e., the amount $h\nu$, an electron is emitted. Part of the absorbed energy, the amount a, is used up in getting free from the metal, and the remainder, $h\nu - a$, is retained by the electron as kinetic energy. If the frequency is so small (wavelength so great) that $h\nu$ is less than a, the electron cannot escape, and therefore short wavelengths are necessary for the photoelectric effect.

154. Atom-models.—One of the aims of all physical research, as has already been mentioned in this book, is to explain the structure and behavior of atoms. If this aim is ever to be accomplished, we must pay due attention to what information the chemists have accumulated, as well as to such physical phenomena as spectral series, X-rays, the Zeeman effect, the photoelectric effect, the behavior of gases under electric discharge, etc. It now seems fairly certain that an atom is made up of a positive charge, the *nucleus,* and one or more electrons. If the atoms are arranged in the order of increasing atomic weights, hydrogen, helium, lithium, beryllium, boron, carbon, nitrogen, oxygen, etc., the first, hydrogen, is believed to have a small nucleus and a single electron,—helium a nucleus of twice the charge of the hydrogen nucleus, with two electrons, —lithium a nucleus of three times the charge of the hydrogen nucleus, with three electrons,—and so on, an element of high atomic weight having a high nuclear charge and a correspondingly large number of electrons. The "atomic number" of an element is the number of its electrons, or the ratio of its positive nuclear charge to that of the hydrogen atom.

So far, physicists and chemists are fairly in agreement, but they differ in their ideas as to the arrangement of the electrons. It is supposed that the outer electrons are responsible for the chemical valence of the element; and certain facts, particularly those concerning the compounds of carbon, which has a valence four (that is, a carbon can hold four hydrogen atoms in combination) seem to indicate that these valence-electrons have definite relations to one another in space. For this reason chemists are inclined to regard them as stationary in the atom, each electron having a fixed position. On the other hand, the physicists, having more in mind that the equilibrium of the electrons must be explained, look upon them as being in revolution about the nucleus. Perhaps some means of reconciling these apparently diverging points of view may be found.

We shall ignore any further consideration of stationary electrons, and confine our attention to atom-theories in which the electrons are in motion. Lorentz's theory to account for the Zeeman effect, which is really an atom-theory, has already been mentioned. It is formulated on the assumption that an electron is attracted toward the center of its orbit by a force directly proportional to the distance. Now if this force is the attraction beweeen the nucleus and the electron, it should vary *inversely as the square* of the distance, like any other electrostatic force, provided the electron is outside the nucleus. It would vary directly as the distance only if the nucleus were in the form of a sphere of positive electricity with the charge uniformly distributed throughout its volume and the electron moving inside this sphere. The atom would then be a globe of positive electricity with enough electrons moving inside to neutralize the positive charge, each electron attracted to the center by a force proportional to its distance from it.

There is a vital difference between a central force varying as the distance and one varying inversely as the distance squared, as will be shown by some simple equations. We know that when a body moves at uniform rate in a circle, the force toward the center must be equal to mv^2/r, m being the mass, v the speed, r the radius. If we let η represent the frequency of revolution, $v = 2\pi r\eta$ and the force may be written $4\pi^2 mr\eta^2$. If we put this force proportional to the distance,

$$4\pi^2 mr\eta^2 = Cr \quad \text{or} \quad \eta = \sqrt{C/4\pi^2 m} \quad (1)$$

showing that the frequency is independent of the distance. That is the electron would make the same number of revolutions per second whether it were far from or close to the nucleus, and if these revolutions sent out light waves, the latter would have the same length whether the revolutions of the electron were violent (large radius of orbit) or weak. On the other hand, suppose that the central force is the usual inverse square force of electrostatic attractions. Then

$$4\pi^2 mr\eta^2 = \frac{Ee}{r^2} \quad \text{or} \quad \eta = \sqrt{Ee/4\pi^2 mr^3} \quad (2)$$

E, and e being respectively the charges of the nucleus and the electron. Here the frequency depends on the radius r, and smaller orbits would be traversed with higher frequencies than larger ones, and therefore send out shorter waves.

If an electron gives out waves as a result of its rotation, it is bound to lose energy, and therefore to draw closer to the nucleus. If the central force is directly proportional to the distance, this would make no difference in the wavelength emitted, and sharp spectrum lines would be the result. On the other hand, if the force is inversely as the distance squared, the electron would give out shorter and shorter waves as it lost energy, so that sharp spectrum lines would not result. Since gases emit spectrum lines which are usually more or less sharp, often so sharp that only the most refined spectrum apparatus can show that they are not absolutely so, this evidence is strongly in favor of a central force proportional to the distance, and therefore of a spherical nucleus with electrons inside. This is the basis of the atom theory proposed at one time by Sir J. J. Thomson.

But certain experiments in radioactivity, and some other considerations, indicate that the nucleus, at least in some elements, is much too small to contain the electrons, and this fact has made the Thomson atom somewhat obsolete.

Ritz's theory assumes that the rotating electrons are under the control of a magnetic field produced by the nucleus, which is assumed to have magnetic properties. By introducing certain vibrations of the nucleus itself, Ritz was able to explain the variations of the Zeeman effect. This theory, however, has not been given much attention by physicists, perhaps because it does not seem to agree very well with the results of radioactivity experiments.

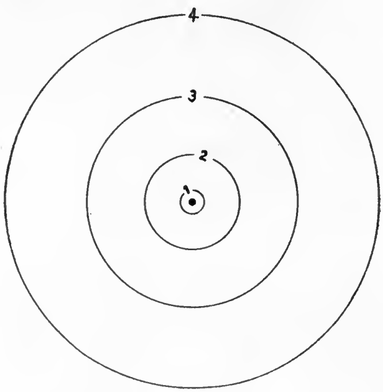

Figure 137

155. Bohr's theory of the hydrogen atom.—Within recent years an atom theory has been formulated by Bohr, in which the electron is considered as being under an electrostatic force of attraction toward the center of its orbit, varying inversely as the square of the distance. We shall confine our attention to Bohr's application of the theory to the simplest element, hydrogen. He conceives the single electron of the hydrogen atom as capable of revolving in any one of a large number of orbits, as indicated by the circles 1, 2, 3, etc., in figure 137, and that during such rotation it *neither radiates nor absorbs energy.* The frequencies of revolution in these orbits are governed by the equation (2) above, which for this case may be written

$$\eta = \frac{e}{2\pi\sqrt{mr^3}} \qquad (2)$$

since for hydrogen $E = e$. The electron would have a different energy as well as a different revolution frequency in each orbit. The potential energy of a negative charge e at a distance r from a positive charge e is $C - e^2/r$, where C is a constant, the potential energy when the electron is removed to an infinite distance from the nucleus. The kinetic energy is of course $mv^2/2$, or $2\pi^2mr^2\eta^2$, since $v = 2\pi r\eta$. If we substitute the value of η from (2), the kinetic energy becomes $e^2/2r$. This makes the total energy

$$W = C - \frac{e^2}{2r} \qquad (3)$$

Thus orbits of smaller radius have smaller energy, and higher revolution frequency.

Now it has long been suspected that an atom cannot absorb or emit light when in the normal condition, but that absorption occurs when it is being ionized, and radiation when the electrons are in the act of recombining. Accordingly, Bohr supposes that the hydrogen electron, in jumping from an orbit of larger to one of smaller radius, to do which it must get rid of a certain amount of energy, emits this energy in one, two, or an integral number of quanta. If the electron jumps from what is practically an infinite distance (complete ionization) to the innermost ring, it emits one quantum, if to ring 2 two quanta, to ring 3 three quanta, etc. The frequency of the emitted light, which we shall call v, is not the same as the revolution frequency η of the orbit at which it chances to stop. Bohr assumes that it is half of η and this assumption fixes the radii of the orbits. For, if we call r_n the radius of the n^{th} orbit, the energy in that orbit is $C - e^2/2r_n$ while the energy at an infinite distance is simply C. Therefore the radiated energy is $C - [C - e^2/r_n] = e^2/2r_n$. Since this is equal to n quanta, or n times hv, and v is $\eta/2$, we have

$$e^2/2r_n = nhv = nh\eta/2$$

Recalling the value of η given in equation (2), we get, by simple substitution and solution,

$$r_n = n^2h^2/4\pi^2e^2m \qquad (4)$$

If we substitute this value of r_n for r in equation (3), we get for the energy in the n^{th} orbit

$$W_n = C - 2\pi^2e^4m/n^2h^2 \qquad (5)$$

Now suppose that the electron, instead of being at the start completely removed from the neighborhood of its nucleus, is only in one of the outer orbits, and it jumps to an inner one. Let us say that it goes from the n^{th} to the n'^{th} orbit. The energy radiated will be

$$W_n - W_{n1} = C - 2\pi^2e^4m/n^2h^2 - [C - 2\pi^2e^4m/n'^2h^2]$$

$$= \frac{2\pi^2e^4m}{h^2}\left(\frac{1}{n'^2} - \frac{1}{n^2}\right)$$

Bohr assumes that in this case the energy is emitted in a single quantum, so that the frequency of the emitted radiation can be gotten by dividing this radiated energy by h, giving

$$v = \frac{2\pi^2e^4m}{h^3}\left(\frac{1}{n'^2} - \frac{1}{n^2}\right) \qquad (6)$$

This can be put into terms of the wavelength λ, instead of the frequency v, by substituting $\lambda = c/v$, c being the velocity of light. This gives

$$\lambda = \frac{ch^3}{2\pi^2e^4m}\left(\frac{n^2n'^2}{n^2 - n'^2}\right) \qquad (7)$$

If we calculate the numerical value of the fraction outside the parenthesis, we must put $c = 3 \times 10^{10}$, $h = 6.55 \times 10^{-27}$, $e = 4.78 \times 10^{-10}$, $m = 8.9 \times 10^{-28}$. This gives

$$\lambda = 909 \times 10^{-8}\frac{n^2n'^2}{n^2 - n'^2} \qquad (8)$$

Now consider the array of wavelengths that would be obtained when an electron jumps to the first orbit from the second, then from the third, then from the fourth, etc. For this purpose, we put $n' = 1$ throughout, and let n take the successive values 2, 3, 4, 5, etc. This will give a spectral series, of formula

$$\lambda = 909 \times 10^{-8} \frac{n^2}{n^2 - 1}$$

This formula about corresponds to a series the beginnings of which were found by Lyman in the far ultraviolet. It also comes at about the right place for the K-series of X-rays for hydrogen.

If we consider the wavelengths gotten when an electron jumps to the *second* orbit, from the third, then from the fourth, etc., the series formula will be, putting $n' = 2$,

$$\lambda = 909 \times 10^{-8} \frac{4n^2}{n^2 - 4}$$

$$= 3636 \times 10^{-8} \frac{n^2}{n^2 - 4}$$

Notice that this is the form of the Balmer formula, for the visible series of hydrogen lines, found by actually considering the wavelengths of the lines as found experimentally (see chapter VII), even the value of the numerical constant being as nearly the same as we could expect, considering uncertainties in the values of h, e, and m.

If we let n' equal 3, or 4, etc., while n takes all the integral values larger than n', we get other possible series, according to Bohr's theory. Traces of one such series, that for $n' = 3$, have been found in the infrared. Wavelengths in the other series would come very far in the infrared, and are not known experimentally.

Bohr does not attempt to explain why the electron does not radiate when revolving in the fixed orbits, as it should do according to the known laws of electromagnetic phenomena, nor in what manner the radiation occurs in the jump from one orbit to another. Still, it is almost inevitable that a theory which makes use of the quantum theory should do violence to the laws of mechanics or electromagnetics, and it is certainly remarkable that he could predict so closely the wavelengths of the hydrogen series by making use of values of h, e, and m found from entirely distinct experiments. At any rate, the theory is regarded seriously by physicists, as a beginning toward an atom-theory that admits of quantitative predictions which can be compared with experimental data.

APPENDIX I.

It is often necessary, in optical experiments, to have at command some means of producing bright-line spectra of known wavelengths. A number of different sources that yield such spectra are mentioned below, and the principal lines given. The wavelengths are expressed, not in centimeters, but in the unit most often employed for this purpose, the angstrom, which is equal to 10^{-8} cm.

The easiest source of all to work with is the sodium flame spectrum. The two bright yellow lines, 5890 and 5896, are close enough together so that for many purposes they answer well enough for truly monochromatic light. They are obtained with considerable brightness by soaking a thin strip of asbestos paper in a solution of common salt, and tying it about the mouth of a Bunsen burner, so that the flame burns from the top edge of the asbestos. If lithium chloride be used instead of common salt, a bright red line is given, of wavelength 6708, but usually there is enough sodium present as an impurity to show its lines also. The lithium line is not as bright as the sodium lines, nor so persistent. Quarter-wave plates are usually adjusted for the sodium lines, and this light is also used in measuring the rotating power of sugar solutions.

A very bright and convenient source is the mercury arc. It is made in various forms, but each consists of a glass or quartz tube partly filled with mercury, the air being pumped out. When a current is once started through this, it continues to run through mercury vapor, giving a number of bright lines. The principal wavelengths are 5790 and 5768, both yellow and bright,—5461, yellow-green and very bright,—4916, blue-green, weak,—4359, blue, strong,—4047, violet, strong. The line 5461 is probably the strongest line available.

Hydrogen, at a few millimeters pressure in a glass tube, and excited by a high-tension electrical discharge, shows a great number of not very strong lines, but also the following series lines: 6563 red, 4861 blue-green, 4341 blue. The first of these is the strongest, the last the weakest. The next line of the series, 4102, does not show visibly, because it takes a fairly strong line to show to the eye in the violet.

Under the same conditions, helium gives a number of clear bright lines, including the following wavelengths: 6678 red, 5876 yellow, 5048, 5016, and 4922, green, 4713 greenish blue, 4472 blue, 4388 violet. The yellow line is very strong, the lines 5016, 4472, and 4388 rather strong, the rest weaker.

A good, bright, and practically monochromatic beam can be obtained by passing the light from the mercury arc through a lens and a prism, so as to form the spectrum, and allowing one of the brightest lines alone to pass through a slit in the plane in which the spectrum is focussed. The interference rings of figure 84 A were photographed with light obtained in this manner.

APPENDIX II.

The following may serve as a proof that, for the special case of plane monochromatic waves, the velocity of propagation is $1/\sqrt{k\mu}$.

According to Faraday's theory of dielectrics, if a charge $+e$ be given to any conducting sphere, (figure 138) an equal amount of electricity is pushed outward through every surface surrounding it. If we take the surface of a sphere of radius r, the amount pushed through each square centimeter will be $e/4\pi^2$, and this amount is defined as the "displacement" and is represented by D. The intensity of the electric force at the surface of this same sphere is e/kr^2. Hence, for this case,

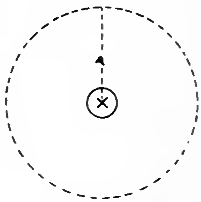

Figure 138

$$D = kF/4\pi$$

and it can be shown that this relation holds in all cases. Faraday regarded D as analogous to a mechanical "strain," F to a mechanical "stress," and therefore $4\pi/k$ is analogous to an elastic coefficient.

Consider a system of three orthogonal axes, as in figure 139, and suppose that a train of plane monochromatic waves are advancing along the positive direction of the X-axis. We shall suppose that they are polarized so that the electrical vibrations are along the Y-axis, the magnetic along the Z-axis. If we represent the former by F, the latter by H, each must follow an equation of the type of (4) in chapter III.

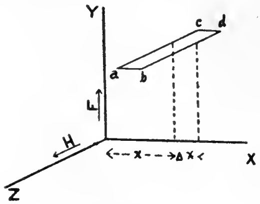

Figure 139

$$F = K. \cos \frac{2\pi}{\lambda} (x - Vt)$$

$$H = M. \cos \frac{2\pi}{\lambda} (x - Vt)$$

where K and M are the amplitudes respectively of the electrical and the magnetic vibrations.

In order to find the relation between V, μ, and k, we must apply the two fundamental laws of electromagnetics.

A. First Law. Consider a very narrow rectangle abdc, perpendicular to the Y-axis, which extends in width from x to x $+ \Delta$x, but may have any length, say L. The amount of electricity which, at any instant t, has been pushed through the area, is gotten by multiplying the area $L\Delta x$ by the value of D in this neighborhood. This gives

$$E = DL\Delta x = kFL\Delta x/4\pi = \frac{kL}{4\pi} \Delta x K. \cos \frac{2\pi}{\lambda} (x - Vt)$$

Since E is changing with the time, this constitutes a current, whose value is the limit of $\Delta E/\Delta t$. The value of E at time $t + \Delta t$ is

$$E + \Delta E = \frac{kL}{4\pi} \Delta x K. \cos \frac{2\pi}{\lambda} (x - Vt - V\Delta t)$$

$$= \frac{kL}{4\pi} \Delta x K \left[\cos \frac{2\pi V\Delta t}{\lambda} \cos \frac{2\pi}{\lambda} (x - Vt) + \sin \frac{2\pi V \Delta t}{\lambda} \sin \frac{2\pi}{\lambda} (x - Vt) \right.$$

$$= \frac{kL}{4\pi} \Delta x K \left[\cos \frac{2\pi}{\lambda} (x - Vt) + \frac{2\pi V\Delta t}{\lambda} \sin \frac{2\pi}{\lambda} (x - Vt) \right]$$

The last equation is gotten by putting

$$\cos \frac{2\pi V\Delta t}{\lambda} = 1, \ \sin \frac{2\pi V\Delta t}{\lambda} = \frac{2\pi V\Delta t}{\lambda}$$

which is permissible since Δt becomes vanishingly small in the limit.

If we now subtract the value of E for time t, and divide by Δt, we get the current equal to

$$\frac{kL\Delta x}{2\lambda} KV. \sin\frac{2\pi}{\lambda}(x - Vt)$$

According to the laws and definitions of electromagnetic phenomena, 4π times the current is equal to the work done in carrying a unit north pole about this area, in the direction abdca. The work done in going from a to b is exactly equal and opposite to that from d to c, but that from b to d is not necessarily balanced by that from c to a, since the magnetic force at distance $x + \Delta x$ from the origin is in general different from that at distance x. Evidently the net work done in the path is the distance bd = ac = L, multiplied by the excess of the value of H at distance x from the origin over that at distance $x + \Delta x$, that is by $-\Delta H$. At distance x,

$$H = M. \cos\frac{2\pi}{\lambda}(x - Vt)$$

and at distance $x + \Delta x$

$$H + \Delta H = M. \cos\frac{2\pi}{\lambda}(x - Vt + \Delta x)$$

$$= M\left[\cos\frac{2\pi}{\lambda}(x - Vt) - \frac{2\pi\Delta x}{\lambda}\sin\frac{2\pi}{\lambda}(x - Vt)\right]$$

as can be proved by trigonometrical transformations similar to those carried out above. This gives

$$\Delta H = -\frac{2\pi\Delta x}{\lambda} M. \sin\frac{2\pi}{\lambda}(x - Vt)$$

and the work done in carrying the unit pole about the rectangle is

$$\frac{2\pi\Delta x}{\lambda} ML. \sin\frac{2\pi}{\lambda}(x - Vt)$$

Putting this equal to 4π times the current, we get

$$\frac{2\pi\Delta x}{\lambda} ML. \sin\frac{2\pi}{\lambda}(x - Vt) = 4\pi\,\frac{kL\Delta x}{2\lambda} KV. \sin\frac{2\pi}{\lambda}(x - Vt)$$

$$M = kKV$$

$$M/K = kV \qquad (1)$$

B. Second Law. The second law states that the electromotive force induced in any circuit is equal to the rate of change in the number of lines of force through that circuit, and is in such a direction as to oppose that change.

Take this time a long narrow rectangle perpendicular to the Z-axis. The number of lines of force through it is μ times H times the area. If the length of the rectangle is again L and its width Δx, this number is

$$\mu HL\Delta x = \mu L\Delta xM. \cos \frac{2\pi}{\lambda} (x - Vt)$$

The rate at which this changes with the time is easily found, by the methods used above, to be

$$\mu L\Delta xMV \frac{2\pi}{\lambda} \sin \frac{2\pi}{\lambda} (x - Vt)$$

The electromotive force is, by definition, the work done in carrying a unit + charge about the circuit. By the same process already used in finding the work done in carrying a unit pole about the other circuit, we find the electromotive force to be

$$-\frac{2\pi\Delta x}{\lambda} KL. \sin \frac{2\pi}{\lambda} (x - Vt)$$

Since this opposes the increase in the lines of force, we have

$$\frac{2\pi\Delta x}{\lambda} KL. \sin \frac{2\pi}{\lambda} (x - Vt) = \mu L\Delta xMV \frac{2\pi}{\lambda} \sin \frac{2\pi}{\lambda} (x - Vt)$$

$$K = \mu MV$$

$$M/K = 1/\mu V \qquad (2)$$

Now, if we eliminate M/K between equations (1) and (2), we get the required relation

$$V^2 = 1/k\mu$$

INDEX.

References are to pages.

4