CONTENTS

INTRODUCTION

§ 1. Physical optics is a special department of electrodynamics—namely, that which comprises the laws of rapidly varying fields. Its particular significance consists in the fact that it represents the branch of physics in which the most refined measurements are possible, and which consequently enables us to penetrate furthest into the details of physical phenomena. At the same time, optics presents a clearer illustration than any other branch of physics of the peculiar tendency of progressive scientific research to leave the original point of departure—namely, the specific sense-impressions—and to place physical concepts on more objective foundations. For, whereas the most important optical concepts, those of light and colour, were originally derived from the impressions on our eyes, these concepts have nothing at all to do with the immediate sensation of sight in present-day physics, but relate rather to electromagnetic waves and vibration periods—a trend of development which has justified itself in the abundant fruit which it has borne.

§ 2. We can progress most easily by linking up with the general system of Maxwell's equations for the electromagnetic field in stationary bodies, particularly if we use the special form which they assume for transparent and non-magnetic bodies. Since the transparency of a body is associated with the condition that no transformation of electromagnetic energy into heat occurs in it, all transparent bodies are electrical insulators in which the vector J of the electric flux vanishes everywhere and at all times. Besides excluding conductors, this also excludes strongly magnetizable bodies; for other bodies we may, without introducing an appreciable error, identify the magnetic induction $\boldsymbol{B}$ with the magnetic intensity of field $\boldsymbol{H}$. Then, by III (31) the field equations assume the simple form :

$$\dot{\boldsymbol{D}} = c \operatorname{curl} \boldsymbol{H}, \ \dot{\boldsymbol{H}} = -c \operatorname{curl} \boldsymbol{E} \quad . \quad . \quad . \quad (1)$$

together with the supplementary equations III (49) and (51) :

$$\operatorname{div} \boldsymbol{D} = 0, \ \operatorname{div} \boldsymbol{H} = 0 \quad . \quad . \quad . \quad . \quad (2)$$

Here $\boldsymbol{E}$ denotes the electric intensity of field, $\boldsymbol{H}$ the magnetic intensity of field, $\boldsymbol{D}$ the electric induction, c the critical velocity, all quantities being measured in the so-called Gaussian system of units (III, § 7).

The above system of equations embraces the optics of all transparent substances. But the variables that occur in them play the part only of auxiliary quantities, since they are not directly measured. There is one quantity, to determine which is the goal of all optical measurements and to calculate which is therefore the proper task of every optical theory. This quantity is the vector of the electromagnetic flux of energy :

$$\boldsymbol{S} = \frac{c}{4\pi}[\boldsymbol{E}, \boldsymbol{H}] \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

which gives the intensity and direction of the intensity of radiation [see III (26)].

For the subsequent treatment of these equations we have to take into consideration the particular relation which connects the vector of the electric intensity of field $\boldsymbol{E}$ with the vector of electric induction $\boldsymbol{D}$ and which endows a substance with its characteristic optical behaviour. Accordingly we find it appropriate to divide the material into three parts, so that we successively discuss the optics of isotropic homogeneous bodies, the optics of crystals and the optics of non-homogeneous bodies in which the phenomena of dispersion and absorption are included.

PART ONE

OPTICS OF ISOTROPIC AND HOMOGENEOUS BODIES

CHAPTER I

REFLECTION AND REFRACTION

§ 3. In the case of an isotropic and homogeneous substance the relation between electric induction and electric intensity of field is expressed by the equation III (28):

$$\boldsymbol{D} = \epsilon \,.\, \boldsymbol{E} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where ϵ denotes the dielectric constant. The field-equations (1) then become:

$$\epsilon \dot{\boldsymbol{E}} = c \operatorname{curl} \boldsymbol{H}, \ \dot{\boldsymbol{H}} = -c \operatorname{curl} \boldsymbol{E} \quad . \quad . \quad . \quad (5)$$

We shall consider as the simplest particular solution of these differential equations the case of a *plane wave* which propagates itself in the direction of one of the co-ordinates, say in that of the positive x-direction. Then all the field-components are independent of y and z and we get from (5) and (2), since static fields do not come into question for optics:

$$\boldsymbol{E}_x = 0, \ \boldsymbol{H}_x = 0$$

whereas the following differential equations hold for the other components:

$$\epsilon \frac{\partial \boldsymbol{E}_y}{\partial t} = -c \frac{\partial \boldsymbol{H}_z}{\partial x}, \ \epsilon \frac{\partial \boldsymbol{E}_z}{\partial t} = c \frac{\partial \boldsymbol{H}_y}{\partial x},$$

$$\frac{\partial \boldsymbol{H}_y}{\partial t} = c \frac{\partial \boldsymbol{E}_z}{\partial x}, \ \frac{\partial \boldsymbol{H}_z}{\partial t} = -c \frac{\partial \boldsymbol{E}_y}{\partial x}$$

Thus there are two pairs of connected quantities among these four field-components; namely $\boldsymbol{E}_y$ is connected with H_z and E_z with H_y, and the same differential equation holds for each individual component, namely :

$$\frac{\partial^2 E_y}{\partial t^2} = \frac{c^2}{\epsilon} \cdot \frac{\partial^2 E_y}{\partial x^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

So if we set :

$$\frac{c^2}{\epsilon} = q^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

it follows from the general integral already derived in II, § 35, for the differential equation (6) that the most general expression for a plane wave which propagates itself in a homogeneous isotropic medium in the direction of the positive x-axis is :

$$\left.\begin{aligned} E_x &= 0 & H_x &= 0 \\ E_y &= \frac{1}{\sqrt{\epsilon}} f\left(t - \frac{x}{q}\right) & H_y &= -\,g\left(t - \frac{x}{q}\right) \\ E_z &= \frac{1}{\sqrt{\epsilon}} g\left(t - \frac{x}{q}\right) & H_z &= f\left(t - \frac{x}{q}\right) \end{aligned}\right\} \quad . \quad . \quad (8)$$

where f and g represent arbitrary functions of a single argument.

As we see, both field-strengths are perpendicular to the direction of propagation; hence the wave is called "transversal." It resolves into two components which are in general independent of one another, and which lie in the direction of the co-ordinate axes. In the case of each component the electric and the magnetic field-strengths are proportional to one another. Their signs are determined by the theorem that the directions of the electric field-strength, of the magnetic field-strength and of propagation form a right-handed system.

§ 4. If we now propose to ourselves the question as to what is to be measured in this electromagnetic wave and which of its properties can hence be ascertained objectively, we find the answer in the vector of energy-radiation (3) which in the present case reduces to its x-component :

$$S_x = \frac{c}{4\pi}(E_y H_z - E_z H_y) = \frac{q}{4\pi}(f^2 + g^2)$$

Thus in an isotropic body the direction of the energy-radiation coincides with that of the wave-normal x, and the amount of energy radiated in the time dt through a surface F which lies in a wave-plane is :

$$S_x \,.\, F \,.\, dt = \frac{q}{4\pi}(f^2 + g^2)\,.\,Fdt \quad . \quad . \quad . \quad (9)$$

Since, however, appreciable effects of radiation always require a finite time, we never measure the radiation vector S_x itself, but rather only its time-integral or its mean value in time taken over a sufficiently great interval of time T. Hence if we use the following abbreviations for the mean values :

$$\frac{1}{T}\int_0^T f^2 dt = \overline{f^2}, \quad \frac{1}{T}\int_0^T g^2 dt = \overline{g^2} \quad . \quad . \quad . \quad (10)$$

then the amount of energy radiated through the surface F in unit time is :

$$\frac{q}{4\pi}(\overline{f^2} + \overline{g^2})\,.\,F \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

which can be recorded by any instrument that takes up the radiant energy completely, and provided that it is sufficiently sensitive (bolometer, radiometer, thermopile).

After the total radiation of the wave has been measured, its further analysis presents a two-fold problem; firstly, we must separate the two summands $\overline{f^2}$ and $\overline{g^2}$ from each other; secondly, we must pass from the mean time values to the functions themselves; that is, we must investigate the exact form of the wave-functions f and g. For this purpose we require special optical contrivances the theory and action of which we must derive in the sequel. At this stage nothing at all can be stated about them. In particular there is no reason for assigning any sort of periodicity to the functions f and g. Actually there are in optics no waves which have a sharply definite period in the mathematical sense, such as we have, say, in acoustics. We therefore do best by leaving the question of the form of the waves completely aside for the present, taking it into consideration only when it becomes really necessary. There is only one assumption which we may make from the very outset, namely, that the mean time values of f and g vanish, that is:

$$\bar{f} = 0 \text{ and } \bar{g} = 0 . \quad . \quad . \quad . \quad . \quad (12)$$

For if a wave-function has a mean value different from zero, we can imagine the wave in question to be replaced by another wave for which the conditions (12) are fulfilled, with a statical field superposed on it, the field being characterized by the mean value, which is not equal to zero. The presence of this field can be made manifest by its ponderomotive action on a charged test-body (electron) and can so be separated from the true optical wave.

§ 5. A plane-wave of unlimited cross-section cannot, of course, be realized in nature. Nevertheless we can produce waves which approximate appreciably to the character of plane-waves. For let us imagine a point-like source of light which begins to emit light at a definite moment of time—say when $t = 0$. Then, since the surrounding medium has been assumed to be homogeneous and isotropic, the light will propagate itself uniformly in all directions. The bounding surface which has been reached by the light after a definite interval of time is called a *wave-front*. There is thus a wave-front corresponding to every moment of time, and the whole of the surrounding space is hence filled by the system of successive wave-fronts which enclose one another. In the present case these wave-fronts are obviously spherical surfaces which surround the source of light concentrically, and so a small portion of a sufficiently great spherical surface can be regarded to a sufficient degree of approximation as a plane wave-front or a wave-plane. Its normal is the corresponding radius of the sphere, and the radiation vector S points in the same direction.

§ 6. Let us now investigate the phenomena that occur when the plane-wave (8) falls on the plane bounding surface of a second isotropic body. We shall take the normal of this bounding plane, the so-called incident normal, as the ξ-axis of a new co-ordinate system directed towards the interior of the second body, whereas the origin 0 of the xyz-system is to be coincident with the origin of the $\xi\eta\zeta$-system. Without reducing the generality of the case we can then make the y-axis and also the η-axis lie in the plane defined by x and ξ, the so-called incident plane, and take this as the plane of Fig. 1. Here all points for which $\xi < 0$ denote the first body (on the left), from which the wave (8) comes, and all points for which $\xi > 0$ denote the second body (on the right); the points for which $\xi = 0$ (the η-axis) constitute the boundary plane. The x-axis is the direction of the ray which comes from the first body—that is, from the left-hand side; it makes the angle θ with the incident normal ξ. The y-axis denotes the wave-plane of the incident ray; this wave-plane is perpendicular to the plane of the figure and also makes the angle θ with the boundary plane. It has been omitted in the figure so as not to multiply the directions to be shown unnecessarily. The z-axis coincides with the ζ-axis and points from the plane of the figure towards the observer.

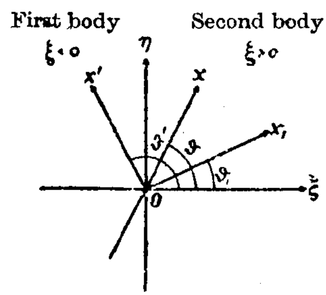

Fig. 1.

We base the solution of the problem before us on the reflection that every system of waves which satisfies the differential equations in the interior of the two bodies and also the boundary conditions, represents a process which is possible in nature.

In order to have the differential equations satisfied in the second body we imagine a plane-wave in it also, after the model of equations (8), which has the ray-direction x_1 (see Fig. 1), inclined at an angle θ_1 to the ξ-axis, and the wave-plane y_1z_1, where we shall again suppose z_1 to coincide with z and ζ. Then the equations (8) hold for the six field-components E_{x_1}, E_{y_1}, E_{z_1}, H_{x_1}, H_{y_1}, H_{z_1}, except that the co-ordinate x_1 now occurs in place of x on the right-hand side of these equations, the functions f_1 and g_1 replace the wave functions f and g, while the constants ϵ and q are supplanted by the dielectric constant ϵ_1 and, by (7), the velocity of propagation :

$$q_1 = \frac{c}{\sqrt{\epsilon_1}} = q\sqrt{\frac{\epsilon}{\epsilon_1}} \quad . \quad . \quad . \quad . \quad (13)$$

in the second body.

But this assumption does not suffice. For by III, § 6, the boundary conditions require that for $\xi = 0$ the values of the tangential field-components—that is, the quantities E_η, E_ζ, H_η, H_ζ—are coincident in both bodies. This gives four equations connecting the wave-functions; to satisfy them, however, we have only the two functions f_1 and g_1 available, since the functions f and g are initially given. To generalize our initial assumption still further, therefore, we assume a second wave in the first body; this wave is, of course, also represented by the equations (8), except that it has a different ray-direction x', which we shall assume to make an angle θ' with the ξ-axis (see Fig. 1), and has the wave-plane $y'z'$, where again $z' = z$. The six field-components $E_{x'}$, $E_{y'}$, $E_{z'}$, $H_{x'}$, $H_{y'}$, $H_{z'}$ are given by the equations (8), if we substitute in them the wave-functions f' and g' and the co-ordinate x', the constants ϵ and q remaining the same.

We have now approximately generalized our assumption for the interior of the two bodies, and can proceed to set up the boundary conditions. In the first body there is an electromagnetic field which results from the superposition of the two plane waves that we have assumed. Hence, remembering that the field-components E_z, H_z,

$\boldsymbol{E}_{x'}$, $\boldsymbol{H}_{x'}$ vanish, we get for the field-components which interest us in the first body:

$$\boldsymbol{E}_\eta = \boldsymbol{E}_y \cos\theta + \boldsymbol{E}_{y'} \cos\theta' = \frac{\cos\theta}{\sqrt{\epsilon}}\cdot f + \frac{\cos\theta'}{\sqrt{\epsilon}}\cdot f'$$

$$\boldsymbol{E}_\zeta = \boldsymbol{E}_z + \boldsymbol{E}_{z'} = \frac{1}{\sqrt{\epsilon}}\cdot g + \frac{1}{\sqrt{\epsilon}}\cdot g'$$

$$\boldsymbol{H}_\eta = \boldsymbol{H}_y \cos\theta + \boldsymbol{H}_{y'} \cos\theta' = -\cos\theta \,.\, g - \cos\theta' \,.\, g'$$

$$\boldsymbol{H}_\zeta = \boldsymbol{H}_z + \boldsymbol{H}_{z'} = f + f'$$

On the other hand, we obtain for the second body ($\xi > 0$), remembering that the field-components $\boldsymbol{E}_{x_1}$ and $\boldsymbol{H}_{x_1}$ vanish:

$$\boldsymbol{E}_\eta = \boldsymbol{E}_{y1} \cos\theta_1 = \frac{\cos\theta_1}{\sqrt{\epsilon_1}} f_1$$

$$\boldsymbol{E}_\zeta = \boldsymbol{E}_{z1} = \frac{1}{\sqrt{\epsilon_1}} g_1$$

$$\boldsymbol{H}_\eta = \boldsymbol{H}_{y1} \,.\, \cos\theta_1 = -\cos\theta_1 \cdot g_1$$

$$\boldsymbol{H}_\zeta = \boldsymbol{H}_{z1} = f_1.$$

Hence if we use the abbreviation:

$$\sqrt{\frac{\epsilon_1}{\epsilon}} = \frac{q}{q_1} = n \quad . \quad . \quad . \quad . \quad . \quad (14)$$

we must have for the boundary plane $\xi = 0$:

$$\cos\theta \,.\, f + \cos\theta' \,.\, f' = \frac{\cos\theta_1}{n}\cdot f_1$$

$$g + g' = \frac{g_1}{n}$$

$$\cos\theta \,.\, g + \cos\theta' \,.\, g' = \cos\theta_1 \cdot g_1$$

$$f + f' = f_1$$

These four equations comprehend all the details of the theory of reflection and refraction. As we see, they fall into two groups, one of which contains only the f-waves, and the other only the g-waves. Thus these two kinds of waves behave quite independently of one another; each obeys its own laws.

§ 7. By means of the last four equations we first calculate the unknown wave-functions f', f_1, g', g_1 from the given wave-functions f and g. We get:

$$f' = \frac{n\cos\theta - \cos\theta_1}{\cos\theta_1 - n\cos\theta'}\cdot f = \mu \cdot f \quad . \quad . \quad (15)$$

$$f_1 = \frac{n(\cos\theta - \cos\theta')}{\cos\theta_1 - n\cos\theta'}\cdot f = \mu_1 \cdot f \quad . \quad . \quad (16)$$

$$g' = \frac{\cos\theta - n\cos\theta_1}{n\cos\theta_1 - \cos\theta'}\cdot g = \sigma \cdot g \quad . \quad . \quad (17)$$

$$g_1 = \frac{n(\cos\theta - \cos\theta')}{n\cos\theta_1 - \cos\theta'}\cdot g = \sigma_1 \cdot g \quad . \quad . \quad (18)$$

As for the arguments of these functions, we have:

$$t - \frac{x}{q} \text{ in } f \text{ and } g$$

$$t - \frac{x_1}{q_1} \text{ in } f_1 \text{ and } g_1$$

$$t - \frac{x'}{q} \text{ in } f' \text{ and } g'$$

And $\xi = 0$ everywhere, so that in transforming to the co-ordinates ξ, η, ζ we have:

$$x = \eta\sin\theta,\ x_1 = \eta\sin\theta_1,\ x' = \eta\sin\theta'$$

which makes the arguments assume the values:

$$t - \frac{\eta\sin\theta}{q},\ t - \frac{\eta\sin\theta_1}{q_1},\ t - \frac{\eta\sin\theta'}{q}.$$

Since the functional equations (15) to (18) must be satisfied for all times t, and for all points η of the boundary surface, it follows that these three arguments must be equal to one another—as can also be seen directly if we differentiate one of the functional equations by parts first with respect to t and then with respect to η, and divide the resulting equations by one another. We get:

$$\frac{\sin\theta}{q} = \frac{\sin\theta_1}{q_1} = \frac{\sin\theta'}{q} \quad . \quad . \quad . \quad . \quad (19)$$

and hence arrive at the *law of refraction*:

$$\frac{\sin\theta}{\sin\theta_1} = \frac{q}{q_1} = n = \sqrt{\frac{\epsilon_1}{\epsilon}}. \quad . \quad . \quad . \quad (20)$$

and the *law of reflection*:

$$\theta' = \pi - \theta \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

If we call the angle which the reflected ray makes with the reversed incident normal the angle of reflection, then the angle of reflection is equal to the angle of incidence.

§ 8. Snell's law of refraction (20), which states that the ratio of the sine of the angle of incidence θ to the sine of the angle of refraction θ_1 is equal to the refractive index n of the second body with respect to the first or to the ratio of the velocities of propagation q and q_1, has been accurately confirmed by innumerable measurements. The refractive index of a substance is usually referred to air as the first substance. Thus the refractive index of water is equal to 1·3, that of glass to 1·5. We then obtain the refractive index of a substance with respect to any other substance by writing down the ratio of their refractive indices with respect to air. If we exchange the substances, the refractive index assumes the reciprocal of its previous value. Accordingly, the refractive index with respect to a vacuum—the so-called "absolute" refractive index—is the product of the refractive index with respect to air and of the absolute refractive index of air—namely 1·0003; as we see, its value differs in most cases only inappreciably from the ordinary refractive index.

If we allow the angle of incidence θ to vary from 0 (normal incidence) to $\frac{\pi}{2}$ (grazing incidence), the angle of refraction θ_1 increases from 0 to $\sin^{-1}\frac{1}{n}$ (limiting angle). But there is a point of fundamental importance which must not be overlooked. It is only when $n>1$, or if, as we say, the second substance is optically denser than the first, that the limiting angle is real. Then the angle of refraction θ_1 is always smaller than the angle of incidence θ—that is, the ray is bent towards the incident normal by the refraction, and the limiting angle denotes the greatest value which the refractive index can assume at all. But if $n<1$—that is, if we exchange the two substances with each other—the angles of incidence and refraction also exchange their rôles, and the angle of refraction becomes greater than the angle of incidence; it attains the value $\frac{\pi}{2}$ only when the angle of incidence has reached the value of the limiting angle. If the angle of refraction is allowed to go beyond the limiting angle, then (20) leads to an imaginary value for the angle of refraction, and the solution which we have found for the problem of refraction becomes meaningless. As there is nothing to prevent our giving the angle of incidence any arbitrary value between 0 and $\frac{\pi}{2}$, a special question arises here, which we shall, however, deal with on a later occasion (§ 12); for the present we shall restrict ourselves to considering those cases for which the law of refraction yields a real value for the angle of refraction θ_1.

§ 9. But the electromagnetic theory of the refraction of light states more than that the refractive index is independent of the value of the angle of incidence; it also tells us the value of the refractive index. For by (20) this is equal to the square root of the ratio of the dielectric constants, or, if we take as our basis the absolute refractive index:

$$n = \sqrt{\epsilon_1}. \qquad (22)$$

If we compare this relationship with observed facts, we find, in general, crass disagreement. For example, for water $n = 1\cdot3$, while $\epsilon_1 = 80$. But even apart from this the fact that (22) cannot be generally valid follows from the fact that by definition the dielectric constant ϵ is independent of the form of the wave-functions f and g, whereas the refractive index n, in the case of *all* substances, depends more or less markedly on the form of the light-waves, that is, on the colour of the light. This phenomenon, dispersion, long constituted a serious obstacle to the acceptance of Maxwell's theory. If we wish to take adequate account of it in the theory here described, nothing remains but to conclude that the fundamental assumption which was introduced at the beginning of this chapter into the field-equations for the optics of homogeneous and isotropic bodies—namely, the relation (4), which states that the electric induction is proportional to the electric intensity of field—does not in general correspond with reality in the case of rapid optical vibrations. To obtain a satisfactory theory of dispersion we shall therefore have to replace this relationship by one that is more general. This will be done in the third part of the present volume, where it will be found that this generalization will have to be based on the circumstance that in the case of refined optical phenomena in nature the assumption that matter is absolutely continuous and homogeneous is no longer justified, but must be modified by the introduction of characteristic structural properties to a certain extent.

If this view is correct, an important significance will still have to be attached to the relation (22)—namely, that of a limiting law which is the better fulfilled the less the dispersion makes itself observed. If we carry out a test in this direction, the relationship in question is found to be definitely confirmed. For the substances which disperse least are gases, and the earliest measurements, by L. Boltzmann, have accurately confirmed the formula (22) in their case. A particularly noteworthy feature is the exact quantitative parallelism between the dependence of the refractive index and of the dielectric constant on the pressure in the case of gases, and this occurs in the sense of equation (22). Hence, with reference to this equation, we are right in speaking of a far-reaching confirmation of the electromagnetic theory within the admissible range of application.

§ 9*a*. But besides giving the *directions* of the reflected and the refracted rays, the theory also gives the *form* of the reflected and the refracted rays, by demanding that the wave-functions in question shall be proportional to the corresponding wave-functions of the incident wave. If in the formulæ (15) to (18) we replace the refractive indices n according to (20) by the angles θ and θ_1, and the angle θ' by $\pi-\theta$, the constants of proportionality assume the following values :

for the reflected wave (f', g') :

$$\mu = \frac{\tan(\theta-\theta_1)}{\tan(\theta+\theta_1)},\ \sigma = \frac{\sin(\theta-\theta_1)}{\sin(\theta+\theta_1)} \quad . \quad . \quad (23)$$

for the refracted wave (f_1, g_1) :

$$\mu_1 = \frac{\sin 2\theta}{\sin(\theta+\theta_1)\cos(\theta-\theta_1)},\ \sigma_1 = \frac{\sin 2\theta}{\sin(\theta+\theta_1)} \quad (24)$$

According to these formulæ (known as Fresnel's formulæ) there is a fundamental difference between the two wave-functions f and g, which corresponds to the physical circumstance that, according to (8), in the case of the f-wave the electric intensity of field lies in the plane of incidence, whereas in the case of the g-wave the electric intensity of field is in a direction perpendicular to the plane of incidence. The coefficients μ correspond to the former, the coefficients σ to the latter.

To test the theory we have to measure the radiant energy. Let us first consider the *reflected wave*. From equation (11), using (15) and (17), we get for the ratio of the intensity of radiation of the reflected wave to