PHILOSOPHICAL TRANSACTIONS

OF THE

ROYAL SOCIETY OF LONDON

Series A. Mathematical and Physical Sciences

No. 781 Vol 237 pp. 411–451 30 September 1938

PROPAGATION OF WAVE-PACKETS INCIDENT OBLIQUELY UPON A STRATIFIED DOUBLY REFRACTING IONOSPHERE

By

H. G. BOOKER, PH.D.

Christ's College, Cambridge and Department of Terrestrial Magnetism,
Carnegie Institution of Washington

LONDON

Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W. 1

Price Five Shillings

PROPAGATION OF WAVE-PACKETS INCIDENT OBLIQUELY UPON A STRATIFIED DOUBLY REFRACTING IONOSPHERE

By H. G. BOOKER, Ph.D.

Christ's College, Cambridge and Department of Terrestrial Magnetism, Carnegie Institution of Washington

(*Communicated by E. V. Appleton, F.R.S.—Received* 1 *December* 1937)

1. Introduction

Little attempt has so far been made to apply Appleton's magneto-ionic theory to wireless waves incident obliquely upon the ionosphere. Actually the magneto-ionic theory in the form given by Appleton (1925, 1932) and others (Nichols and Schelleng 1925; Breit 1927; Goldstein 1928) is only suitable for investigating vertical propagation in the ionosphere, and it is the object of this communication to develop a generalization of Appleton's magneto-ionic theory capable of dealing conveniently with waves incident obliquely upon the ionosphere. A general investigation into oblique propagation of electromagnetic waves through a slowly varying doubly refracting medium has already been made (Booker 1936), and the ideas there developed will now be applied to propagation through the ionosphere of wireless waves of wave-length sufficiently short (less than a kilometre, say) to regard the medium as slowly varying.

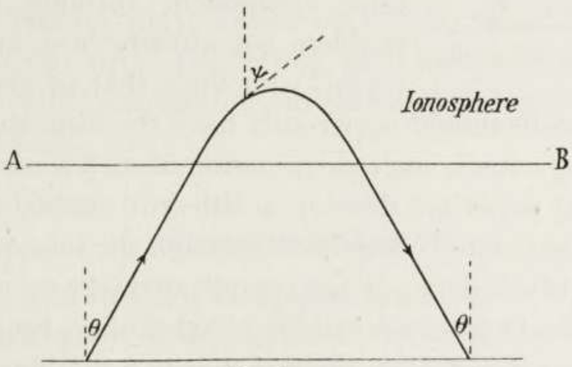

Fig. 1

The ionosphere may be thought of ideally as a horizontally stratified medium composed of free electrons, the density of which is a function only of height above the surface of the earth, increasing upwards from zero at a certain level AB (fig. 1). Let us neglect for the moment the effect of the earth's magnetic field. Then the refractive index μ of the medium is practically unity below AB and decreases as we go upwards above AB into the ionosphere. According to the simple ray theory, a wave incident upon the ionosphere at an angle θ to the vertical is bent round so that, when it reaches the level where the refractive index is μ, it is travelling at an angle ψ to the vertical, where ψ is given by Snell's Law

$$\mu \sin \psi = \sin \theta. \tag{1}$$

When the wave reaches the level where the refractive index is μ_0, given by

$$\mu_0 \sin (\tfrac{1}{2}\pi) = \sin \theta, \tag{2}$$

it returns to the earth as shown in fig. 1. If $\theta = 0$, then $\mu_0 = 0$.

The introduction of the earth's magnetic field brings with it complications. On entering the ionosphere an incident wave of arbitrary polarization is split into two characteristically polarized magneto-ionic components each of which is, under certain conditions, independently propagated through and reflected from the ionosphere. The refractive indices for each of the two magneto-ionic components are given by Appleton's magneto-ionic theory. The refractive index for a magneto-ionic component is not, however, simply a function of height above the surface of the earth: it depends also upon the angle between the direction of propagation and the direction of the earth's magnetic field. Consequently the value of μ to be put in (1) is a function—and quite a complicated function—of ψ. The solution of (1) for ψ under these circumstances is the essential difficulty which has to be overcome in studying oblique propagation of a magneto-ionic component through the ionosphere. This problem has already been attempted by Schekulin (1930), who finds that in general the ray for each magneto-ionic component should apparently have the form shown in fig. 2. He does not, however, offer any convincing interpretation of such a curious result.

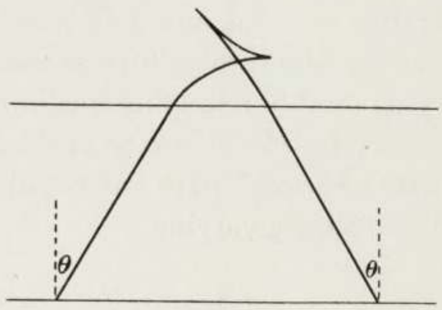

FIG. 2. Ray according to Schekulin.

In § 2 of the present paper we develop a different method for describing oblique propagation of a magneto-ionic component through the ionosphere, avoiding the use of a refractive index μ which depends in a complicated way upon an unknown angle of refraction ψ. This method makes the validity of Schekulin's result almost obvious. The explanation of the result depends on the fact that in a doubly refracting medium the direction of group-propagation in general differs from that of phase-propagation. Fig. 2 represents only the phase-propagation of a magneto-ionic component, whereas it is the group-propagation which yields a cuspless ray of more conventional type. The level of reflexion of a magneto-ionic component is the level where its direction of group-propagation becomes horizontal. By studying the propagation of wave-packets through the ionosphere we deduce in § 3 that this level is given by the upper cusp in fig. 2, and not by the lower cusp, which merely gives the level where the direction of phase-propagation becomes horizontal. The fact that the level of reflexion of a magneto-ionic component is not given by the lower cusp in fig. 2 means that it is in general wrong to deduce the level of reflexion by putting the angle of refraction equal to $\frac{1}{2}\pi$ in Snell's Law as we do when the effect of the earth's magnetic field is neglected (cf. (2)). In § 4 consideration is given to the effect of the earth's magnetic field in producing lateral deviation of wave-packets during their journey through the ionosphere, and in § 5 a study is made of the effect upon group-propagation of damping due to collisions of the free electrons with neutral air molecules. § 6 is a review of the phenomena associated with strata in which the ray theory breaks down in a medium that is structurally slowly varying. Although the terminology of the ionosphere has been employed throughout, much of the discussion in §§ 2–6 is applicable to any slowly varying plane-stratified doubly refracting medium, and it is not until we reach § 7 that it is necessary explicitly to derive the fundamental formula of the oblique incidence magneto-ionic theory. This formula is an algebraic quartic equation, and further progress depends on our ability to solve this equation. In general the only practical way of doing this is to resort to one of the standard numerical methods, and in § 8 some special cases are worked out in this way, mainly in order to see how important it is to deduce the critical values of electron density required to produce reflexion from the ionosphere by expressing the condition that the direction of group-propagation, and not phase-propagation, should be horizontal. There are, however, two special cases in which the analytical solution of the quartic equation is sufficiently simple to be of practical value, and these are discussed in §§ 9 and 10. § 9 deals with the case when the plane of phase-propagation is perpendicular to the magnetic meridian-plane (east-west transmission), and this may be taken to include the case of vertical incidence, which has been investigated by a number of writers (Appleton and Builder 1933; Taylor 1933, 1934; Ratcliffe 1933; Booker 1934, 1935; Bailey 1934; Martyn 1935; Försterling and Lassen 1933*a*, *b*; Goubau 1934, 1935*a*, *b*). § 10 deals with propagation in equatorial regions, where the earth's magnetic field is horizontal.

2. Phase-propagation

Take axis-z vertically upwards and let a plane harmonic wave of frequency $f = kc/2\pi$ be incident upon the ionosphere at an angle θ to the vertical in the yz-plane as shown in fig. 3. Consider separately the two magneto-ionic components into which the incident wave is split. In the neighbourhood of the level where the angle of refraction is ψ and the refractive index is $\mu(\psi)$, the propagation of a magneto-ionic component may be represented by the wave-function

$$\exp[ik\{ct-\mu(\psi)\,(\sin\psi\,.\,y+\cos\psi\,.\,z)\}]. \tag{3}$$

By Snell's Law (1), $\mu(\psi)\sin\psi$ is a prescribed constant, namely, $\sin\theta$. The product $\mu(\psi)\cos\psi$ we replace by a new and important variable q. (3) therefore becomes

$$\exp[ik\{ct-(\sin\theta)\,y-qz\}], \tag{4}$$

in which q is the only unknown. q actually contains the electron density N as a variable, together with the earth's magnetic field $\mathbf{H}^\circ$, the wave-frequency f and the angle of incidence θ as parameters. The propagation of the wave through the ionosphere is represented by plotting q as a function of N for given values of $\mathbf{H}^\circ$, f and θ.

The precise significance of the quantity q may be seen from fig. 4. The wave-function (3) shows that, at the level in the ionosphere where the angle of refraction is ψ and the refractive index is $\mu(\psi)$, the phase-propagation of a magneto-ionic component may be represented by a vector of length $\mu(\psi)$ making an angle ψ with the vertical as shown in fig. 4. The length of this vector is the velocity of phase-propagation divided into the

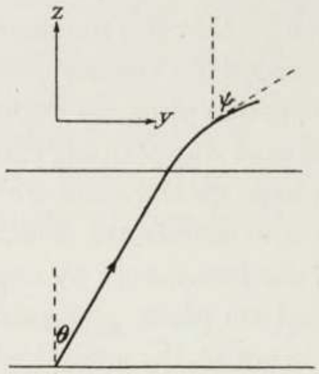

FIG. 3

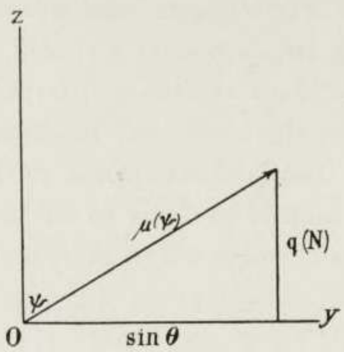

FIG. 4. Phase-propagation.

velocity of light *in vacuo*, and the direction of the vector is the direction of phase-propagation. The horizontal component of the vector is $\mu(\psi)\sin\psi$, which, by Snell's Law (1), remains constant and equal to $\sin\theta$ as the wave is propagated through the ionosphere. The vertical component of the vector is $\mu(\psi)\cos\psi$ and is therefore the quantity which we have denoted by q and which varies with the electron density in a manner to be investigated. When we use the wave-function (3) we are describing propagation by means of the quantities μ and ψ shown in fig. 4. When we use the wave-function (4) we are describing propagation by means of the alternative quantities q and $\sin\theta$ shown in fig. 4. The latter pair of quantities is the more convenient because, if the angle of incidence θ is a prescribed constant, the propagation of the wave is completely described by the variation of the single quantity q. The advantages of this method are already realized by mathematicians (see for example Hartree 1929, 1931 *a*), but little use has so far been made of it by physicists.

(4) differs from the wave-function assumed by Appleton (1932) by the inclusion of the term depending on the angle of incidence. If $\theta = 0$ (vertical incidence), we revert to Appleton's wave-function

$$\exp[ik\{ct - qz\}], \tag{5}$$

and q is then the same thing as the refractive index μ (cf. fig. 4 when $\theta = 0$). Let us recall the way in which Appleton developed the theory in the special case $\theta = 0$ so that we can adopt the same method for non-zero values of θ. By substituting the wave-function

(5) into the electromagnetic equations Appleton deduced a quartic equation for q which, since it contained no cubic or linear terms, was actually a quadratic equation for q^2 of the form

$$\alpha q^4 + \gamma q^2 + \epsilon = 0, \tag{6}$$

where α, γ and ϵ contain the electron density N as a variable together with the earth's magnetic field $\mathbf{H}^\circ$ and the wave-frequency f as parameters. When the two values of q^2 given by (6) are plotted against N for fixed values of $\mathbf{H}^\circ$ and f, they give curves of the form shown in fig. 5 (cf. Taylor 1933, fig. 3 and Ratcliffe 1933, figs. 9–13), one curve (say the broken one) referring to the extraordinary magneto-ionic component, and the other to the ordinary magneto-ionic component. Instead of regarding (6) as a quadratic for q^2, let us regard it as a quartic for q. There are then two pairs of equal and opposite roots, which, when plotted against N for fixed values of $\mathbf{H}^\circ$ and f, give four curves IA, RA, IB, RB which fit together symmetrically as shown in fig. 6. The broken and continuous curves in fig. 6 correspond to those in fig. 5. IA refers to the upgoing extraordinary wave, AR to the downcoming extraordinary wave, IB to the upgoing ordinary wave and BR to the downcoming ordinary wave. At A the pair of roots corresponding to the extra-ordinary wave passes from real to conjugate complex values via equality as N increases through N_A. N_A is the critical electron density which has to be attained in the ionosphere in order to produce reflexion of the extraordinary wave. Corresponding remarks apply to the ordinary wave. The symmetry of the curves in fig. 6 about the N-axis expresses the fact that the quartic (6) for q is actually a quadratic for q^2.

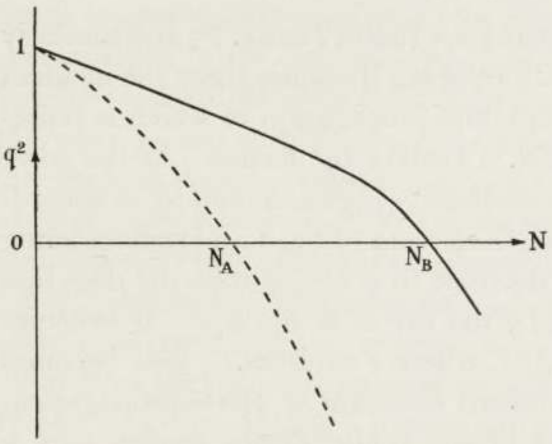

FIG. 5. q^2 as a function of N, vertical incidence.

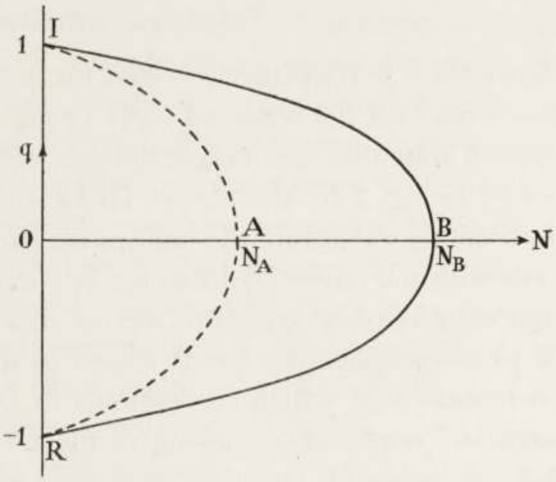

FIG. 6. q as a function of N, vertical incidence.

If we now take the full wave-function (4) and substitute it into the electromagnetic equations as in the case when $\theta = 0$, we again obtain a quartic equation for q, which, however, is *no longer simply a quadratic for* q^2: as we shall see explicitly in § 7, the equation is of the form

$$\alpha q^4 + \beta q^3 + \gamma q^2 + \delta q + \epsilon = 0, \tag{7}$$

where $\alpha, \beta, \gamma, \delta$ and ϵ contain N as a variable together with $\mathbf{H}^\circ, f$ and θ as parameters, and β and δ in general differ from zero. Consequently if we plot the four roots of (7) against N for fixed values of $\mathbf{H}^\circ, f$ and θ as we did in fig. 6 for the case $\theta = 0$, we obtain curves as shown in fig. 7 which are *no longer symmetrical about the N-axis.*

The significance of this result becomes apparent when we interpret fig. 7 with the aid of fig. 4. Below the ionosphere the electron density vanishes and the wave-function for the incident wave is

$$\exp[ik\{ct-(\sin\theta)\,y-(\cos\theta)\,z\}]. \tag{8}$$

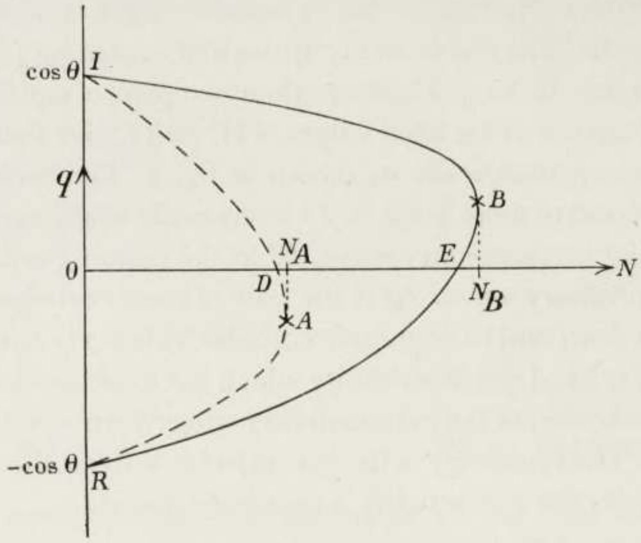

FIG. 7. q as a function of N, oblique incidence.

Hence for $N = 0$ we have $q = \cos\theta$ for the incident wave (point I in fig. 7), and similarly $q = -\cos\theta$ for the reflected wave (point R). On entering the ionosphere the incident wave is split into two magneto-ionic components the propagation of which is represented in fig. 7 by the curves *IDAR* and *IBER*. Consider the former. As the wave proceeds upwards into the ionosphere the electron density increases, and so, in accordance with the curve *IDAR* in fig. 7, q decreases. Referring to fig. 4 and remembering that $\sin\theta$ remains constant, we see that this decrease in q means that the direction of phase-propagation bends round as shown by the ray *id* in fig. 8(a). It becomes horizontal at d, which corresponds to D in fig. 7, where q vanishes. q now becomes negative, implying according to fig. 4 a downward direction of phase-propagation. But the curve *IDAR* in fig. 7 *does not turn back towards smaller electron densities until N has further increased to* N_A. In consequence the portion of the phase-ray corresponding to *DA* in fig. 7 must be as shown by *da* in fig. 8(a). The complete phase-ray corresponding to the curve *IDAR* in fig. 7 is therefore represented by *idar* in fig. 8(a), and similarly the phase-ray corresponding to the curve *IBER* in fig. 7 is represented by *iber* in fig. 8(b). The arrows indicate direction of phase-propagation.

Schekulin's result is thus confirmed. Moreover it should be emphasized that phase-rays of the form shown in fig. 8 are not exceptional phenomena which only occur as a result of some special choice of the conditions of propagation. They are general phenomena which can only be avoided by special choice of the conditions of propagation. The unusual form of the phase-rays merely expresses the effect of the earth's magnetic field in producing asymmetry between the propagation of the upgoing and downcoming waves. The explanation of this result depends on the fact that what we have been discussing so far is only the phase-propagation, whereas it is the group-propagation which yields cuspless rays of more conventional type than those shown in fig. 8.

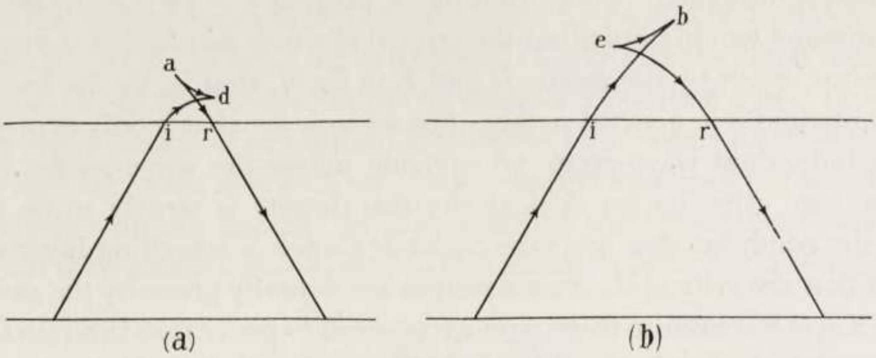

FIG. 8. Phase-rays of magneto-ionic components.

3. Group-propagation

Consider a group of plane harmonic waves whose frequencies are grouped around a mean frequency f and whose directions of incidence upon the ionosphere are grouped around a mean direction making an angle θ with the vertical. Such a group of waves forms a wave-packet consisting of a portion of a plane harmonic wave of frequency f concentrated in the neighbourhood of a certain point moving upwards towards the ionosphere with the velocity of light *in vacuo* at an angle θ to the vertical. So long as the wave-packet is still below the ionosphere, individual wave-crests within it are moving with the same velocity and in the same direction as the wave-packet, so that they are stationary relative to it. When the wave-packet enters the ionosphere it is magneto-ionically split into two characteristically polarized wave-packets which may be considered separately. Moreover, for each of these wave-packets the further effect of the earth's magnetic field in general is to cause the velocity of group-propagation to differ from that of phase-propagation not merely in magnitude but also in direction. Hence, as a wave-packet is propagated through the ionosphere, individual wave-crests within it are in general moving across the wave-packet with a velocity which differs both in magnitude *and direction* from the velocity of the wave-packet as a whole (cf. fig. 9). This clears up the difficulty with regard to fig. 8. A ray such as that shown in fig. 8 (a) or (b) does not indicate the path actually followed by a characteristically polarized wave-packet incident upon the ionosphere at an angle θ to the vertical. It merely indicates for each level in the ionosphere, both for the upward and downward journeys, the directions in which individual wave-crests move across the wave-packet when situated at that level. The actual path followed by the wave-packet still remains to be investigated.

There emerges from this interpretation of figs. 7 and 8 an interesting and important fact concerning the critical electron density which must be attained in the ionosphere in order to return a magneto-ionic component to the surface of the earth. It might be thought that this critical electron density could be deduced by putting the angle of refraction ψ equal to $\frac{1}{2}\pi$ in Snell's Law as we did in (2) for the case when the earth's magnetic field is neglected. Since, from fig. 4, putting $\psi = \frac{1}{2}\pi$ is equivalent to putting $q = 0$, this method would imply that the critical electron densities of the magneto-ionic components are given by the points D and E in fig. 7, that is, by the levels d and e in fig. 8. This method is in general wrong. For we now see that it only expresses the condition that individual wave-crests are moving across the wave-packet horizontally, whereas the true criterion for critical electron density is clearly to be obtained by expressing the condition that the *wave-packet as a whole* is travelling horizontally.* We may expect that the critical electron densities are actually given by the points A and B in fig. 7 since $\partial q/\partial N$ is infinite there, and we proceed to prove that this is in fact the case. It will follow that the levels of reflexion of the magneto-ionic components in fig. 8 are a and b, not d and e.

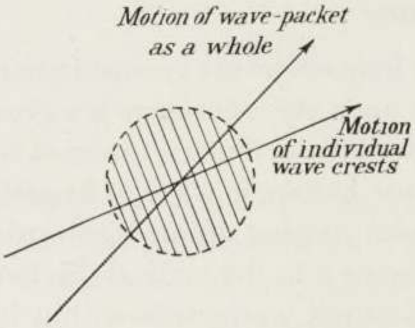

FIG. 9. Propagation of a wave-packet in a doubly refracting medium.

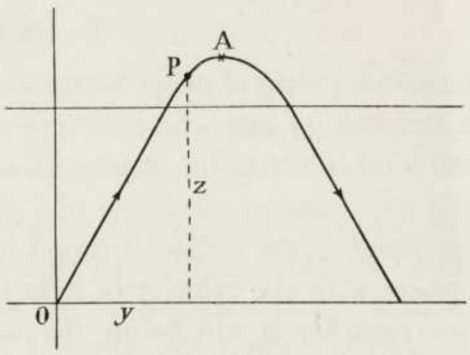

FIG. 10. The path of a wave-packet.

It will now be convenient to take the wave-function of a magneto-ionic component in the form

$$\exp\left\{ik\left(ct - Sy - \int^z q\,dz\right)\right\}. \tag{9}$$

(9) differs from (4) in that we have now written S for the sine of the angle of incidence, and have replaced qz by $\int^z q\,dz$ because q is a function of z and we now wish to have a wave-function simultaneously applicable for all values of z. The functional dependence

* Or, as may easily be proved to be the same thing, by expressing the condition that the direction of energy-propagation, as represented by the Poynting vector, is horizontal (cf. Booker 1936).

of q upon z is given by one or other of the curves IA, RA, IB, RB in fig. 7 when we know the distribution of the electron density N with height z above the surface of the earth.

A group of waves of the type (9) is represented by the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(k,S)\exp\left\{ik\left(ct-Sy-\int^{z} q\,dz\right)\right\}dk\,dS, \tag{10}$$

and if we suppose that for real values of k and S the amplitude function $A(k, S)$ is only appreciable in the neighbourhood of a certain value k_0 of k and of a certain value S_0 of S, (10) represents a wave-packet of mean frequency $k_0c/2\pi$ incident upon the ionosphere at an angle $\sin^{-1} S_0$ to the vertical in the yz-plane as described above (p. 417). The position of the wave-packet at any instant is obtained by Kelvin's method of stationary phase. We locate the point at which the various component waves of the group interfere constructively with one another by expressing the conditions that small variations of k and S from their mean values k_0 and S_0 make no first-order variation in the phase-function

$$k\left(ct-Sy-\int^{z} q\,dz\right). \tag{11}$$

These conditions, obtained by making the partial derivatives of (11) with respect to k and S vanish for $k = k_0$, $S = S_0$, are

$$ct-Sy = \int_0^z \frac{\partial(kq)}{\partial k}dz, \tag{12}$$

$$y = -\int_0^z \frac{\partial q}{\partial S}dz, \tag{13}$$

in which k and S are supposed to have their mean values k_0 and S_0. The two relations (12) and (13) between y, z and t give the position (y, z) of the wave-packet at time t. The lower limit of the integral in (11) has been chosen to make y, z and t vanish together in (12) and (13), that is, to make the wave-packet pass through the origin at time zero.

(13) gives y explicitly as a function of z, that is, it gives the equation of the path traced out by the motion of the wave-packet. Fig. 10 is a representation of this path, P being the position of the wave-packet at time t after leaving the origin O. When the value of y corresponding to a given value of z has been obtained from (13), it may be substituted into (12), which then gives the time taken by the wave-packet to reach the height z. ct is usually called the equivalent path or group-path of the wave-packet. It may be mentioned that, if in differentiating (11) partially to obtain (12) and (13) we take the independent parameters to be, not k and S, but k and (kS), (12) is replaced by a relation giving ct explicitly as a function of z without involving y.

Our immediate object is to calculate the critical electron density which must be attained in the ionosphere in order to return the wave-packet to the surface of the earth. This is the electron density corresponding to the point A in fig. 10 where the direction of motion of the wave-packet is horizontal. The level at which the electron density is critical is therefore obtained by putting

$$\frac{dy}{dz} = \infty, \tag{14}$$

where y is given as a function of z by (13). But from (13) we deduce by differentiating with respect to z that

$$\frac{dy}{dz} = -\frac{\partial q}{\partial S}. \tag{15}$$

Hence the condition (14) for critical electron density becomes

$$\frac{\partial q}{\partial S} = \infty. \tag{16}$$

Now q is given as a function of S by the quartic equation (7), the coefficients of which depend on S. Denote the left-hand side of (7), regarded as a polynomial in q, by $F(q)$, so that the quartic may be written

$$F(q) \equiv \alpha q^4 + \beta q^3 + \gamma q^2 + \delta q + \epsilon = 0, \tag{17}$$

where α, β, γ, δ and ϵ, and consequently q, depend on the electron density, the earth's magnetic field, the wave-frequency, and in particular upon S, the sine of the angle of incidence. Differentiate (17) partially with respect to S. We obtain

$$\frac{\partial F}{\partial q}\frac{\partial q}{\partial S} + \left(\frac{\partial \alpha}{\partial S}q^4 + \frac{\partial \beta}{\partial S}q^3 + \frac{\partial \gamma}{\partial S}q^2 + \frac{\partial \delta}{\partial S}q + \frac{\partial \epsilon}{\partial S}\right) = 0, \tag{18}$$

where the first term comes from differentiating the powers of q and the term in brackets from differentiating the coefficients. We may anticipate that the latter term is always finite under the conditions in which we are interested, and this is confirmed by the analytical expressions for α, β, γ, δ and ϵ to be deduced in § 7. It therefore follows from (18) that the condition (16) for critical electron density is equivalent to

$$\frac{\partial F}{\partial q} = 0. \tag{19}$$

But this is the well-known condition that the quartic equation (17) for q should have a pair of equal roots. Now the four roots of (17) are represented in fig. 7 as functions of the electron density N by the four curves IA, RA, IB, RB. At the point A the pair of roots corresponding to the extraordinary wave passes from real to conjugate complex values via equality as N increases through N_A. At the point B the pair of roots corresponding to the ordinary wave passes from real to conjugate complex values via equality as N increases through N_B. N_A and N_B are therefore the electron densities for which a pair of roots of the quartic equation (7) for q are equal, and are consequently the critical electron densities for the extraordinary and ordinary waves. IA represents the upgoing extraordinary wave, AR the downcoming extraordinary wave, IB the upgoing ordinary wave and BR the downcoming ordinary wave. The levels in fig. 8 corresponding to the critical electron densities N_A and N_B are a and b, corresponding to the points A and B in fig. 7. We have therefore shown that, in fig. 8, although d and e are the levels where individual wave-crests are moving across the wave-packets horizontally, nevertheless a and b are the levels where the wave-packets as a whole are travelling horizontally. In other words, the critical electron densities are given in fig. 7 by the points A and B where two roots of the quartic (7) for q are equal, not by the points D and E where $q = 0$. Of course, in the special case of vertical incidence represented in fig. 6, it happens that, owing to symmetry of the (q, N) curves about the N-axis, the points A and B coincide with the points D and E, so that in this particular case it is legitimate to calculate the critical electron densities from the condition $q = 0$.

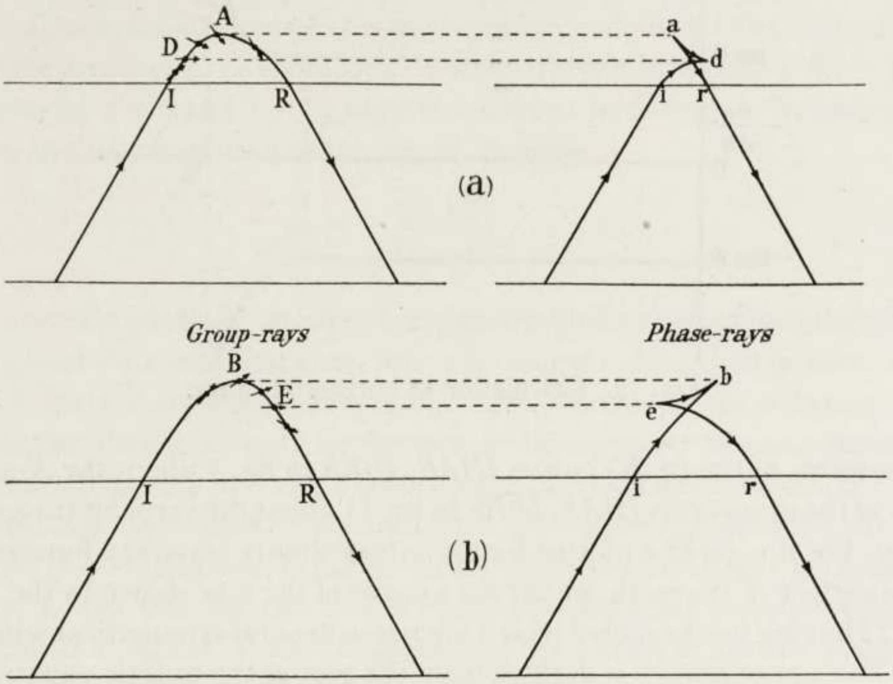

FIG. 11. Group- and phase-rays of magneto-ionic components.

We are now in a position to form a complete picture of the propagation of wave-packets through the ionosphere under the influence of the earth's magnetic field. An arbitrarily polarized wave-packet incident upon the ionosphere at the point I in fig. 11 is magneto-ionically split into the two characteristically polarized wave-packets, the paths of which are represented by the two group-rays $IDAR$, $IBER$ shown in the upper and lower sections of the figure. IA is the path followed by the extraordinary wave-packet on its upward journey through the ionosphere, and corresponds to the root of the quartic equation (7) for q represented by the curve IA in fig. 7; D in fig. 11 shows the position of the upgoing wave-packet when it reaches the level in the ionosphere where the electron density is that corresponding to the point D in fig. 7. AR in fig. 11 is the path followed by the extraordinary wave-packet on its downward journey through the ionosphere and corresponds to the root of the quartic equation (7) for q represented by the curve AR in fig. 7. In the same way, IB in fig. 11 is the path followed by the ordinary wave-packet on its upward journey, corresponding to the root of the quartic for q represented by the curve IB in fig. 7, and BR in fig. 11 is the path followed by the ordinary wave-packet on its downward journey, corresponding to the root of the quartic for q represented by the curve BR in fig. 7; E in fig. 11 shows the position of the downcoming wave-packet when it is at the level in the ionosphere where the electron density is that corresponding to the point E in fig. 7. The spots at various points along the two group-rays represent the two characteristically polarized wave-packets at various stages during their propagation through the ionosphere.

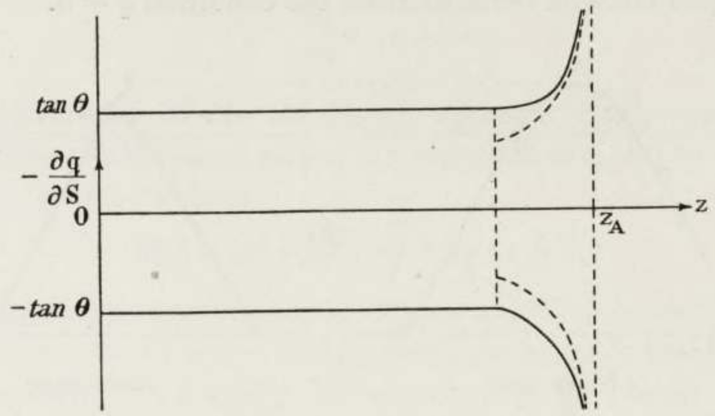

FIG. 12. Calculation of horizontal range.

The asymmetry of the (q, N) curves $IDAR$, $IBER$ in fig. 7 about the N-axis leads to asymmetry of the group-rays $IDAR$, $IBER$ in fig. 11 about the verticals through A and B respectively. For if $-\partial q/\partial S$ is plotted for the extraordinary wave as a function of height z above the surface of the earth, we obtain a curve of the type shown by the continuous line in fig. 12 having two branches which are not in general symmetrical with respect to the z-axis. The upper branch is derived from the root of the quartic equation (7) for q represented by the curve IA in fig. 7 and corresponds to the upgoing wave, while the lower branch is derived from the root of the quartic for q represented by the curve AR in fig. 7 and corresponds to the downcoming wave. It is clear from (15) that $-\partial q/\partial S$ is positive for an upgoing wave-packet and negative for a downcoming wave-packet. Below the ionosphere we have (cf. (8))

$$-\frac{\partial q}{\partial S} = -\frac{d(\pm\cos\theta)}{d(\sin\theta)} = \pm\tan\theta \tag{20}$$

for the upgoing and downcoming waves respectively. In the ionosphere the two values of $-\partial q/\partial S$ in general cease to be equal and opposite, but in accordance with (16) they tend to $\pm\infty$ as $z \to z_A$ (the height of the point A in fig. 11 (a)). As z increases through z_A, the two values of $-\partial q/\partial S$, like those of q, become conjugate complex quantities. Now

(13) shows that the horizontal displacement y undergone by a wave-packet as it ascends to the height z is represented by the area under the $-\partial q/\partial S$ curve. In particular the horizontal displacement undergone by the extraordinary wave-packet in reaching its highest point A in fig. 11(a) is given by the area in fig. 12 contained between the upper branch of the $-\partial q/\partial S$ curve, the z-axis and the ordinates $z = 0$ and $z = z_A$. The further horizontal displacement undergone by the wave-packet in returning to the surface of the earth is given by the area in fig. 12 contained between the lower branch of the $-\partial q/\partial S$ curve, the z-axis and the ordinates $z = 0$ and $z = z_A$. Since these two contributions to the horizontal range of the wave-packet (that is, its total horizontal displacement from the time it leaves the surface of the earth until the time it returns to it) are in general unequal, it is clear that the group-ray shown in fig. 11(a) must in general be asymmetrical about the vertical through A. It may be mentioned that, in using fig. 12 for a numerical calculation of the horizontal range of a wave-packet, $-\partial q/\partial S$ is most easily derived from (18), while the area in fig. 12 contained between the two branches of the $-\partial q/\partial S$ curve and the ordinates $z = 0$ and $z = z_A$ may be obtained by fitting to the infinities curves, as shown by the broken curves in fig. 12, of the form

$$\frac{\pm K}{\surd(z_A - z)}, \tag{21}$$

where K is a certain constant; the area between the broken curves may then be obtained analytically, and the remaining area, which no longer contains an infinite ordinate, by a standard numerical method. There is, of course, a curve for the ordinary wave of the same type as that shown in fig. 12 for the extraordinary wave. We may remark that, for the same mean frequency and angle of incidence, the horizontal ranges of the ordinary and extraordinary wave-packets are different. Consequently although the points I in fig. 11(a) and (b) are identical, being the point where the arbitrarily polarized incident wave-packet is split into the two characteristically polarized wave-packets, nevertheless the points R in fig. 11(a) and (b) where the two characteristically polarized wave-packets leave the ionosphere are not identical.

The integral on the right-hand side of (12), which is required for calculating the equivalent path ct of a characteristically polarized wave-packet, may be discussed in the same way as the integral on the right-hand side of (13), which gives the horizontal displacement y of the wave-packet.*

The phase-rays of fig. 8 are redrawn on the right-hand side of fig. 11. The points i, r, d, a, e, b of the phase-rays correspond to and are at the same levels as the points I, R, D, A, E, B of the group-rays. Now it will be remembered that the phase-rays merely indicate for each level in the ionosphere both for the upward and downward journeys the directions in which individual wave-crests are moving across the wave-packets when situated at that level. For each of the selected positions of the wave-packet along

* Cf. Goubau 1934, whose method of calculating the integrand on the right-hand side of (12) and of performing the numerical integration is not quite so convenient (or accurate) as that suggested here.

the group-ray *IDAR* in fig. 11(*a*) we have therefore indicated by an arrow the direction in which, according to the phase-ray *idar*, individual wave-crests are travelling across the wave-packet. When the wave-packet enters the ionosphere at *I* individual wave-crests are moving with the same velocity and in the same direction as the wave-packet as a whole, so that they are stationary relative to it. But as the wave-packet travels upwards into the ionosphere the direction of motion of individual wave-crests ceases to be identical with the direction of motion of the wave-packet as a whole (cf. fig. 9). At a point on the group-ray between *I* and *D* in fig. 11(*a*) the direction of motion of individual wave-crests across the wave-packet is more inclined to the vertical than the direction of motion of the wave-packet as a whole. When the wave-packet has reached the point *D*, individual wave-crests are travelling across it horizontally although the wave-packet as a whole is still moving upwards. Along the portion *DA* of the group-ray the direction of motion of the wave-packet as a whole continues to slope upwards in spite of the fact that the direction of motion of individual wave-crests across it is now sloping downwards. After the wave-packet has passed its highest point *A*, both the direction of motion of the wave-packet as a whole and the direction of motion of individual wave-crests across it have downward slopes, which tend to identity as the wave-packet approaches the point *R* where it leaves the ionosphere. There is a corresponding interpretation of fig. 11(*b*): when the wave-packet is between *B* and *E* its direction of motion as a whole has a downward slope whereas the direction of motion of individual wave-crests across it has an upward slope.

4. Lateral deviation

The group of waves represented by the integral (10) does not represent a wave-packet which is limited in all directions. The reason for this is that the directions of incidence upon the ionosphere of the component waves in (10), although grouped around a certain mean direction of incidence, nevertheless all lie in the yz-plane. Consequently (10) represents a wave-packet which is limited in all directions in the yz-plane but is infinite in the direction perpendicular to the yz-plane, that is, a wave-packet which, instead of being concentrated in the neighbourhood of a single point moving towards the ionosphere in the yz-plane, is concentrated in the neighbourhood of a line perpendicular to the yz-plane moving at right angles to its own length. The group-rays shown in fig. 11 are the loci of the intersections of such lines with the yz-plane.

Let us now add to the group some waves whose frequencies and directions of incidence are grouped around the same mean frequency and mean direction of incidence as those of the waves already present but whose directions of incidence nevertheless do not lie quite in the yz-plane. These new waves interfere with the waves already present in such a way as to destroy the wave-packet except in the neighbourhood of a single point moving towards the ionosphere parallel to the yz-plane. The wave-packet is now limited in all directions, and, by proper adjustment of the phases of the newly introduced waves, its line of travel before entering the ionosphere may be made to lie actually in the *yz*-plane. When this wave-packet enters the ionosphere its direction of motion as a whole in general ceases to coincide with the direction of motion of individual wave-crests across it. This means that, although the direction of motion of individual wave-crests continues to be parallel to the *yz*-plane, the wave-packet as a whole does not in general remain in the *yz*-plane. It suffers lateral deviation and follows a twisted (non-planar) path through the ionosphere, which it ultimately leaves in a plane parallel to, but not in general coincident with, the *yz*-plane. The curves on the left-hand side of fig. 11 are the projections on to the *yz*-plane of these twisted paths traced out by the motion through the ionosphere of wave-packets limited in all directions. Thus the full effect of the earth's magnetic field in general is not only to make the curves of fig. 11 asymmetrical but also to deviate the characteristically polarized wave-packets laterally out of the plane of incidence.* An expression for the horizontal displacement of a wave-packet perpendicular to the *yz*-plane similar to the integral (13) for the horizontal displacement parallel to the *yz*-plane may easily be deduced, but it has been omitted from the present paper for the sake of simplicity.

Considerations of symmetry with respect to the magnetic meridian-plane show that there can be no lateral deviation for propagation in the magnetic meridian-plane (north-south transmission). Nor can there be any *net* lateral deviation when the plane of phase-propagation is perpendicular to the magnetic meridian-plane (east-west transmission): although the characteristically polarized wave-packets do in general leave the plane of incidence on their upward journeys through the ionosphere, they nevertheless return to the plane of incidence on their downward journeys. A characteristically polarized wave-packet incident vertically upon the ionosphere is in general deviated sideways in the magnetic meridian-plane during its upward journey through the ionosphere. It retraces its steps along the *same* path, however, on its downward journey. This may be regarded as lateral deviation in a degenerate case of east-west transmission, or as asymmetry of the group-ray in a degenerate case of north-south transmission.

5. Effect of damping

Throughout the above discussion of propagation through the ionosphere we have, for the sake of simplicity, completely neglected the effect of collisional damping. Let us now examine the effect of taking such damping into account. We know from the magneto-ionic theory as developed by Appleton (1932) for the case of vertical incidence that the introduction of collisional damping makes the coefficients α, γ and ϵ in the quadratic equation (6) for q^2 complex, and in § 7 we shall see that the same is true for the

* The existence of such lateral deviation has been pointed out by Baker and Green (1933), whose statements concerning its effect are not however correct.

quartic equation (7) for q in the general case of oblique incidence. When damping is taken into account, therefore, the roots of (7) are always complex, which appears to make the right-hand sides of (12) and (13) also complex. This obviously cannot mean that the time of flight and horizontal range of a wave-packet are complex, and so it is necessary to reconsider our interpretation of equations (12) and (13).

Hitherto we have supposed that a wave-packet composed of a group of waves having a given mean frequency $kc/2\pi$ and a given mean angle of incidence $\sin^{-1}S$ leaves the origin at time zero, and we have regarded (12) and (13) as giving the position (y, z) of the wave-packet at a later time t. The relations between y, z and t depend of course upon the given values of k and S, which occur in the right-hand sides of (12) and (13) as parameters. Let us now suppose however that it is the values of y, z and t which are prescribed and let us imagine that (12) and (13) are solved for k and S. These values of k and S give the mean frequency and mean angle of incidence which a wave-packet leaving the origin at time zero must have in order to reach the point (y, z) at time t. If damping is now taken into account, the value of q occurring in the right-hand sides of (12) and (13) becomes complex. This means that for real values of y, z and t the values of k and S obtained by solving (12) and (13) also become complex. Consequently the wave-packet as received at the point (y, z) at time t is a portion of a wave of the form

$$\exp\left\{ik\left(ct - Sy - \int_0^z q\,dz\right)\right\}, \tag{22}$$

with complex values of k and S. The complex nature of k and S indicates a variation of amplitude with t and y in addition to the variation of amplitude with z resulting from the complex nature of q.

Now it seems unlikely that, owing to the effect of damping in the ionosphere, it is necessary in practice to take into account the complex nature of the values of k and S associated with a downcoming wave-packet received at the surface of the earth, and this may be verified by expressing the conditions that the contributions of the imaginary part of q to the right-hand sides of (12) and (13) are negligible compared with those of the real part. At vertical incidence we derive in this way the condition that the percentage variation of the amplitude reflexion coefficient with frequency per kilocycle per second due to variation of absorption with frequency should be negligible compared with twice the equivalent path measured in kilometres. This condition is abundantly satisfied under all conditions when radio exploration of the ionosphere is possible with present equipment. We feel confident that collisional damping has no detectable effect upon the path or time of flight of wave-packets in the ionosphere and may be omitted from equations (12) and (13).

Since we are now convinced that the values of k and S in (22) may be regarded as real, the only variation of amplitude to be considered arises from the imaginary part of q. If in (22) we put

$$q = \xi + i\eta, \tag{23}$$

where ξ and η are real, we see that the phase-propagation of a magneto-ionic component is given by the factor

$$\exp\left\{ik\left(ct - Sy - \int_0^z \xi\,dz\right)\right\}, \tag{24}$$

while there is in addition a spatial variation in the amplitude of the wave represented by the factor

$$\exp\left(k\int_0^z \eta\,dz\right). \tag{25}$$

η is negative for an upgoing wave and positive for a downcoming wave. It should be noticed that the surfaces of constant amplitude given by the factor (25) are horizontal planes, and do not in general coincide with the surfaces of constant phase given by (24): the attenuation of a wave has to be calculated by integrating the imaginary part of q vertically even in the case of oblique incidence.

In order to evaluate the attenuation suffered by a wave in the ionosphere, it is often convenient to divide the ionosphere into a deviating region and a non-deviating region, the latter being situated below the former. The non-deviating region is one in which propagation can be considered rectilinear but in which absorption may not be negligible. In this region there is no appreciable variation with height in the angle between the direction of propagation of the wave and the direction of the earth's magnetic field, and consequently the main difficulty which the technique of the present paper was developed to overcome is non-existent. The attenuation suffered by a magneto-ionic component in the non-deviating region of the ionosphere may be calculated by integrating the imaginary part of Appleton's complex refractive index along the oblique rectilinear path of the component in the non-deviating region, and this is true in spite of the fact that the planes of constant amplitude do not coincide with the planes of constant phase.

In the deviating region of the ionosphere the following method is available for calculating the integrand of (25). Reflexion from the ionosphere is only appreciable provided the effect of damping in the deviating region is sufficiently small (see p. 429). In these circumstances Newton's method of correcting approximate roots of equations may be used to deduce the roots of the quartic equation (7) from the roots which are obtained when damping is entirely neglected and which are represented by the curves *IA*, *RA*, *IB*, *RB* in fig. 7. Let q now denote, not as in (23) a root of (7) for the case when damping is taken into account, but a root of (7) for the case when damping is entirely neglected. Let ζ be the small complex correction to this root necessitated by the introduction of a small amount of damping; the root then becomes

$$\xi + i\eta = q + \zeta. \tag{26}$$

Let α, β, γ, δ and ϵ be the (real) coefficients in the quartic equation (7) when damping is entirely neglected, and let $i\alpha'$, $i\beta'$, $i\gamma'$, $i\delta'$ and $i\epsilon'$ be the small corrections to these coefficients when a small amount of damping is taken into account. α', β', γ', δ' and ϵ' are real and depend upon the electron density, the earth's magnetic field, the frequency of electronic collisions, the wave-frequency and the angle of incidence in a manner which will be investigated in § 7. The quartic equation (7) for the case when damping is entirely neglected is therefore

$$\alpha q^4 + \beta q^3 + \gamma q^2 + \delta q + \epsilon = 0, \tag{27}$$

and for the case when a small amount of damping is taken into account is

$$(\alpha + i\alpha')(q+\zeta)^4 + (\beta + i\beta')(q+\zeta)^3 + (\gamma + i\gamma')(q+\zeta)^2 + (\delta + i\delta')(q+\zeta) + (\epsilon + i\epsilon') = 0. \tag{28}$$

If we subtract (27) from (28) and retain only the terms of the first order of magnitude, we obtain without difficulty

$$\zeta = -i\frac{\alpha' q^4 + \beta' q^3 + \gamma' q^2 + \delta' q + \epsilon'}{4\alpha q^3 + 3\beta q^2 + 2\gamma q + \delta}. \tag{29}$$

This is a pure imaginary, and so we deduce from (26) that

$$\begin{cases} \xi = q, & (30) \\ \eta = -\dfrac{\alpha' q^4 + \beta' q^3 + \gamma' q^2 + \delta' q + \epsilon'}{4\alpha q^3 + 3\beta q^2 + 2\gamma q + \delta}. & (31) \end{cases}$$

(30) implies that the propagation of phase in the ionosphere is unaffected by the introduction of a small amount of damping. (31) gives, in terms of the roots of the quartic equation (7) for the case when damping is entirely neglected, the quantities which, in accordance with (25), have to be integrated with respect to height in order to calculate the attenuation suffered by the magneto-ionic components in the deviating region of the ionosphere. Corresponding to the four roots of (7) which are obtained when damping is entirely neglected and which are represented by the curves *IA*, *RA*, *IB*, *RB* in fig. 7, there are four values of η which refer respectively to the upgoing extraordinary wave, the downcoming extraordinary wave, the upgoing ordinary wave and the downcoming ordinary wave. To calculate the total attenuation of the extraordinary wave in the deviating region we have to integrate η with respect to height up to the level where the electron density has the value N_A in fig. 7, using the value of η corresponding to the upgoing extraordinary wave and then down again using the value of η corresponding to the downcoming extraordinary wave. A similar procedure has to be followed for the ordinary wave.

6. Validity of the ray theory

In the present paper we are assuming the validity of the ray theory of propagation of wireless waves through the ionosphere. This is an assumption which requires a good deal more discussion when the earth's magnetic field is taken into account than in the simple case of an isotropic medium. The problem has been investigated elsewhere (Booker 1936), and it turns out that, provided the variation in the constitution of the ionosphere with height, per vacuum wave-length upon 2π, is small, the wave-function of a magneto-ionic component may legitimately be taken in the form (9) so long as the root of the quartic equation (7) for q occurring in (9) is not approximately equal to another root of this quartic. Two roots of the quartic (7) are said to become approximately equal when their difference ceases to be large compared with their rates of change with height per vacuum wave-length upon 2π.

Let us assume that the variation in the constitution of the ionosphere with height is small enough to insure the validity of the ray theory in general and let us consider the significance of the exceptional strata where the quartic equation (7) for q has a double root. At the point A in fig. 7 the root of (7) corresponding to the upgoing extraordinary wave is equal to the root corresponding to the downcoming extraordinary wave. Hence, in the neighbourhood of the height where the electron density has the critical value N_A corresponding to the point A in fig. 7, there exists a stratum within which the difference between the values of q for the upgoing and downcoming extraordinary waves is not large compared with their rates of change per vacuum wave-length upon 2π. Consequently the ray theory fails to describe the propagation of the extraordinary wave in this stratum and the wave-function fails to be of the form (9). This is simply because reflexion of the extraordinary wave is actually taking place in this stratum where the value of q for the upgoing extraordinary wave is approximately equal to the value of q for the downcoming extraordinary wave. Similarly, in the neighbourhood of the height where the electron density has the critical value N_B corresponding to the point B in fig. 7, there is a stratum within which the ray theory ceases to describe the propagation of the ordinary wave owing to the fact that the wave is in process of being reflected. To describe what is actually happening within a reflecting stratum it is necessary to obtain a more exact solution of the electromagnetic equations than that upon which the ray theory is based. This has been done (Gans 1915; Hartree 1931 *b*) under certain simplifying restrictions, and, so long as damping may be neglected, it appears that, although the ray theory fails to describe even approximately what is actually happening within a reflecting stratum, nevertheless it does give practically the right expression for the reflected wave.

If the effect of damping is now taken into account, a complication arises from the fact that the values of q corresponding to the upgoing and downcoming waves, being complex, no longer attain actual equality, with the result that reflexion becomes partial instead of total. So long as the difference between the two values of q decreases to a value small compared with their rates of change per vacuum wave-length upon 2π, reflexion may still be considered total. But if the difference between the values of q decreases only to a value comparable with their rates of change per vacuum wave-length upon 2π, reflexion is only partial. If the effect of damping is so marked that the value of q for the upgoing wave always differs from the value of q for a downcoming wave by an amount large compared with their rates of change per vacuum wave-length upon 2π, no question of reflexion of the upgoing wave arises: it simply travels upwards until it is absorbed. Calculation of the precise value of the reflexion coefficient in a case when the effect of damping is such that the partial reflexion phenomenon is important requires a more exact solution of the electromagnetic equations than that upon which the ray theory is based. The fact that reflexion is only partial when damping is taken into account means that the attenuation to be associated with the reflecting stratum should be considered somewhat greater than that given by simple integration of the imaginary part of q up to, and down again from, the level where the difference between the values of q for the upgoing and downcoming waves is a minimum. If however we use in the reflecting stratum the approximate formula (31) for the imaginary part of q, a slight error is made which tends to compensate that arising from neglect of the partial reflexion phenomenon.

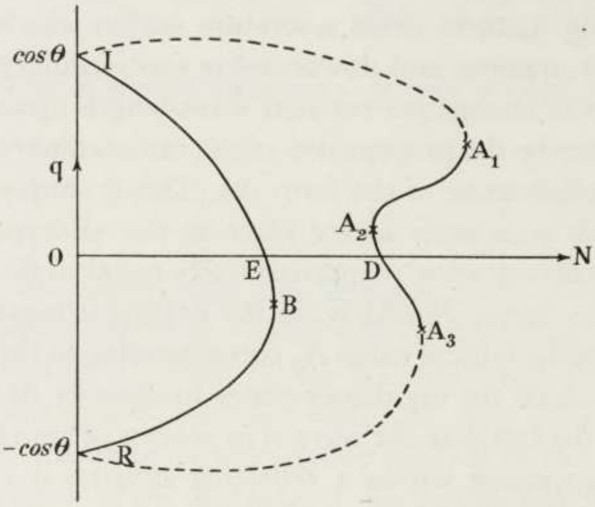

FIG. 13. q as a function of N.

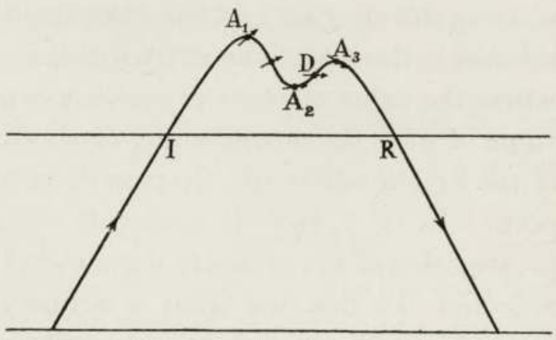

FIG. 14. A group-ray.

Although throughout the above discussions we have found it convenient, in the case when damping is neglected, to take the roots of the quartic equation (7) for q to vary with electron density in the manner shown in fig. 7, it should be understood that the (q, N) curves are by no means always of this particular form. Another form which the (q, N) curves can take is shown in fig. 13. The root of the quartic equation (7) for q corresponding to the upgoing ordinary wave is here represented by the curves IB and A_2A_3, while the root corresponding to the downcoming ordinary wave is represented by the curves RB and A_2A_1. Between B and A_2 the values of q corresponding to the ordinary wave are conjugate complex quantities. The roots of the quartic for q corresponding to the upgoing and downcoming extraordinary waves are represented by IA_1 and RA_3 respectively. At A_1 a root of (7) corresponding to the upgoing extraordinary wave becomes equal to a root corresponding to the downcoming ordinary wave, and in consequence an upgoing extraordinary wave-packet, on reaching the level in the ionosphere where the electron density is that corresponding to the point A_1, is reflected, and is subsequently propagated in the manner of a downcoming ordinary wave-packet.

At A_2 a root of (7) corresponding to the downcoming ordinary wave becomes equal to a root corresponding to the upgoing ordinary wave, so that, when the downcoming wave-packet has descended to the level in the ionosphere where the electron density is that corresponding to the point A_2, it is again reflected and travels up to the level where the electron density is that corresponding to the point A_3. Here a root of (7) corresponding to the upgoing ordinary wave becomes equal to a root corresponding to the downcoming extraordinary wave, and in consequence the upgoing wave-packet is reflected for the third time and thereafter resumes the type of propagation appropriate to the extraordinary wave. The path followed by the wave-packet is therefore of the form shown in fig. 14. The arrows indicate as in fig. 11 the directions in which individual wave-crests are travelling across the wave-packet in various positions. Between A_1 and A_2, because q is positive, the direction in which individual wave-crests are travelling across the wave-packet is sloping upwards although the direction of motion of the wave-packet as a whole is sloping downwards. Between D and A_3, because q is negative, the direction in which individual wave-crests are travelling across the wave-packet is sloping downwards although the direction of motion of the wave-packet as a whole is sloping upwards. The justification for this interpretation of the (q, N) curve $IA_1A_2DA_3R$ in fig. 13, and for a number of other statements in this section, is to be found in the paper already referred to (Booker 1936).

We have seen that strata within which the ray theory breaks down because a root of the quartic equation (7) for q corresponding to an upgoing wave is approximately equal to a root corresponding to a downcoming wave are associated with the phenomenon of reflexion. We now turn our attention to a totally different phenomenon that is associated with strata within which the ray theory breaks down either because a root of the quartic for q corresponding to the upgoing ordinary wave is approximately equal to a root corresponding to the upgoing extraordinary wave, or because a root corresponding to the downcoming ordinary wave is approximately equal to a root corresponding to the downcoming extraordinary wave. In strata such as these the phenomenon is one of breakdown in the independence of propagation of the magneto-ionic components.

For waves incident vertically upon the ionosphere at the magnetic equator propagation of one of the magneto-ionic components is entirely unaffected by the presence of the earth's magnetic field. It is this magneto-ionic component which is called the ordinary wave, the other being called the extraordinary wave, and it is by continuity with this special case that the four roots of the quartic equation (7) for q are in general classified for each level in the ionosphere into an upgoing ordinary wave, an upgoing extraordinary wave, a downcoming ordinary wave and a downcoming extraordinary wave. If in the special case of vertical incidence at the magnetic equator these four roots are plotted as functions of height, it turns out, as one would expect, that the root corresponding to the upgoing ordinary wave at any one level in the ionosphere is continuously connected to the root corresponding to the upgoing ordinary wave at any other level in the ionosphere, and likewise for the other three roots. Now it might be thought that this continues to be true for any angle of incidence and any direction of the earth's magnetic field, but this is not so. It can happen for instance that, as we pass away from the special case of vertical incidence at the magnetic equator, the two roots corresponding to the upgoing ordinary wave and the upgoing extraordinary wave for a certain level $z = z_T$ in the ionosphere become equal and then separate again. After passing through this transition condition the situation is (cf. Booker 1934, fig. 8 and 1936, fig. 4) that the root corresponding to the upgoing ordinary wave for $z<z_T$ is continuously connected to the root corresponding to the upgoing extraordinary wave for $z>z_T$, while the root corresponding to the upgoing extraordinary wave for $z<z_T$ is continuously connected to the root corresponding to the upgoing ordinary wave for $z>z_T$. If in these circumstances an upgoing ordinary wave-packet passes the level $z=z_T$, it thereupon acquires the propagation characteristics of the upgoing extraordinary wave; if the wave-packet subsequently returns to the surface of the earth, it does so as an extraordinary wave-packet unless as a result of a second transition it regains the propagation characteristics of the ordinary wave. A wave-packet which has the propagation characteristics of the ordinary wave both on entering and on leaving the ionosphere is often called an ordinary wave-packet throughout its entire journey through the ionosphere even though along part of its path it is propagated like an extraordinary wave-packet. An example of this is provided by an ordinary wave-packet incident vertically upon the ionosphere at a magnetic pole. In this case the upgoing ordinary wave-packet acquires at a certain level in the ionosphere the propagation characteristics of the extraordinary wave and as a result is reflected at a level which in the case of vertical incidence at the magnetic equator would be the level of reflection of the extraordinary wave; on its downward journey through the ionosphere the wave-packet regains at a certain (the same) level the propagation characteristics of the ordinary wave (cf. Ratcliffe 1933, fig. 6, and Booker 1934, figs. 10–15). In the same way a wave-packet which has the propagation characteristics of the extraordinary wave both on entering and on leaving the ionosphere is conveniently called an extraordinary wave-packet throughout its entire journey through the ionosphere even though along part of its path it is propagated like an ordinary wave-packet. An example of this is provided by figs. 13 and 14.

What actually happens under transition conditions is as follows. Suppose that at a certain angle of incidence $\theta = \theta_T$, the values of q for the upgoing ordinary wave and the upgoing extraordinary wave are equal at the level $z = z_T$. Then when $\theta = \theta_T$ there is in the neighbourhood of the level $z = z_T$ a stratum within which the difference between the values of q for the upgoing ordinary wave and the upgoing extraordinary wave is not large compared with their rates of change per vacuum wave-length upon 2π. Consequently the independence of propagation of the upgoing magneto-ionic components breaks down within this stratum: an incident upgoing ordinary wave-packet below this stratum produces above the stratum a pair of transmitted upgoing wave-packets, one of which has the propagation characteristics of the ordinary wave and the other of the extraordinary wave. As θ varies from θ_T, the difference between the values of q at $z = z_T$ for the upgoing ordinary wave and the upgoing extraordinary wave increases from zero. Simultaneously the amplitude of one of the transmitted wave-packets decreases, and it becomes negligible when the difference between the two values of q becomes large compared with their rates of change per vacuum wave-length upon 2π. The important point is that, if the transmitted wave-packet whose amplitude becomes negligible as θ decreases from θ_T is, say, the one whose propagation characteristics are those of the ordinary wave, then the transmitted wave-packet whose amplitude becomes negligible as θ increases from θ_T is the one whose propagation characteristics are those of the extraordinary wave. Consequently as θ decreases through θ_T, the wave transmitted through the stratum in the neighbourhood of $z = z_T$ due to an incident ordinary wave passes from an ordinary wave to an extraordinary wave via a linear combination of both. When θ differs sufficiently from θ_T, there is no stratum near $z = z_T$ where the independence of propagation of the magneto-ionic components breaks down, but for $\theta < \theta_T$ an upgoing ordinary wave-packet acquires at the level $z = z_T$ the propagation characteristics of the extraordinary wave whereas for $\theta > \theta_T$ it retains those of the ordinary wave. This does not mean that for $\theta < \theta_T$ the upgoing wave-packet suffers any discontinuity at $z = z_T$: the discontinuity is only one of nomenclature, but even so is by no means without significance. It is, of course, possible to pass through a transition condition by variation of parameters other than the angle of incidence, such as for example the frequency (cf. Booker 1935, p. 275). It should be emphasized that there can be no question of reflexion being associated with a stratum in which the independence of propagation of the magneto-ionic components breaks down, unless of course *three* roots of the quartic equation (7) for q are approximately equal in the same stratum.

We notice that in fig. 7 the two roots of the quartic equation (7) for q corresponding to the upgoing ordinary and upgoing extraordinary waves become equal at the point I, while the two roots corresponding to the downcoming ordinary and downcoming extraordinary waves become equal at the point R. This means that near the bottom of the ionosphere, where the electron density is small, the independence of propagation of the magneto-ionic components should break down. However, as has been shown elsewhere (Booker 1936), the essential problem associated with this region of low ionization density is to locate the stratum which divides the doubly refracting region of the ionosphere from the region where the electron density is so low that the medium must be regarded as isotropic. This stratum is actually located where the difference between the two values of q corresponding to the upgoing or downcoming magneto-ionic components (at vertical incidence the two complex refractive indices) is of the same order of magnitude as their rates of change per vacuum wave-length upon 2π. It is in this stratum that the limiting polarization of the downcoming magneto-ionic components is determined. If the region of low electron density near the bottom of the ionosphere is one which produces little refraction and is mainly effective in producing absorption, the stratum dividing the isotropic from the doubly refracting region of the ionosphere is located where the difference between the absorption indices of the two magneto-ionic components is of the same order of magnitude as their rates of change per vacuum wave-length upon 2π.

7. The generalized magneto-ionic theory

The foregoing discussion of propagation of wave-packets incident obliquely upon the ionosphere has been based upon the quartic equation (7) for q without knowing precisely the way in which the coefficients in this equation depend upon the electron density, the earth's magnetic field, the frequency of electronic collisions, the wave-frequency and the angle of incidence. This is because, although we have employed the terminology of the ionosphere throughout, much of the above discussion is applicable to any slowly varying plane-stratified doubly refracting medium. We have now reached the stage however when we must actually evaluate by the method described in § 2 the coefficients in the quartic equation (7) for q appropriate to a medium composed of ionized air under the influence of the earth's magnetic field. We shall suppose that it is only the free electrons in the ionized air which directly affect the propagation of wireless waves through the ionosphere.

In order to investigate the magneto-ionic theory we shall require the following notation:

e = charge on an electron (e.s.u.). e is negative for negative electrons.
m = mass of an electron.
N = electron density (number of free electrons per unit volume).
$\mathbf{H}^\circ$ = earth's magnetic field (e.m.u.).
ν = mean frequency of collisions between a free electron and neutral air molecules.
$p = kc = 2\pi$ times wave-frequency.
$p_H = |e\mathbf{H}^\circ/mc| = 2\pi$ times the magneto-ionic frequency.
$x = (4\pi e^2/mp^2)\,N$.
$\mathbf{y} = (e/mcp)\,\mathbf{H}^\circ$.
$y = |\mathbf{y}| = p_H/p$.
$z = \nu/p$.
$u = 1 - iz$.
$\rho_0 = -u/x$.
$\boldsymbol{\rho} = -\mathbf{y}/x$.
θ = angle of incidence.
$S = \sin\theta$.
$C = \cos\theta$.
$\rho = 1/(q^2 - C^2) \neq |\boldsymbol{\rho}|$.

It should be noted that x, $\mathbf{y}$ and z are respectively proportional to the electron density, the earth's magnetic field, and the collisional frequency. ρ_0 is the reciprocal of the scattering index of the free electrons in an ionized gas when no magnetic field is imposed. Following Darwin (1934) we shall not apply Lorentz's correction to the theory of dispersion in an ionized gas*; the correction would require the addition of a term $\frac{1}{3}x$ to the definition of u. It may be mentioned that the derivation in this section may be regarded as covering the case of a medium in which the electrons are not free, but possess a natural angular frequency of vibration ω. The extension would be effected merely by adding to the definition of u not only the term $\frac{1}{3}x$ but also a term $-(\omega/p)^2$.

In order to avoid confusion with the quantities x, y and z just defined, we shall now take the Cartesian coordinates to be (x_1, x_2, x_3). Take axis-3 vertically upwards and the 23-plane as the plane of phase-propagation. Let the polarization vector and the electric intensity of a magneto-ionic component be (cf. (4)) the real parts of

$$\left.\begin{aligned}\mathbf{P} &= \mathbf{A}\exp\{ik(ct-Sx_2-qx_3)\},\\ \mathbf{E} &= \mathbf{B}\exp\{ik(ct-Sx_2-qx_3)\},\end{aligned}\right\} \tag{32}$$

where $\mathbf{A}$ and $\mathbf{B}$ are constant vectors. Even though we are concerned with a medium containing unbound electrons, we may represent flow of charge in the medium by the time derivative of a polarization vector. Take Maxwell's equations in the form

$$\left.\begin{aligned}\operatorname{curl}\mathbf{H} &= \frac{1}{c}(\dot{\mathbf{E}}+4\pi\dot{\mathbf{P}}),\\ \operatorname{curl}\mathbf{E} &= -\frac{1}{c}\dot{\mathbf{H}}.\end{aligned}\right\} \tag{33}$$

Elimination of $\mathbf{H}$ between these gives

$$\operatorname{curl}\operatorname{curl}\mathbf{E} = -\frac{1}{c^2}(\ddot{\mathbf{E}}+4\pi\ddot{\mathbf{P}}),$$

which is the same thing as

$$\nabla^2\mathbf{E}-\operatorname{grad}\operatorname{div}\mathbf{E}-\frac{1}{c^2}(\ddot{\mathbf{E}}+4\pi\ddot{\mathbf{P}}) = 0. \tag{34}$$

When $\mathbf{P}$ and $\mathbf{E}$ depend on time and space in the manner given by (32), (34) becomes

$$\left.\begin{aligned}(C^2-q^2)\,E_1+4\pi P_1 &= 0,\\ (1-q^2)\,E_2+SqE_3+4\pi P_2 &= 0,\\ SqE_2+C^2E_3+4\pi P_3 &= 0,\end{aligned}\right\} \tag{35}$$

or, solving for $\mathbf{E}$,

$$\left.\begin{aligned}E_1 &= 4\pi\rho P_1,\\ E_2 &= 4\pi\{C^2\rho P_2-Sq\rho P_3\},\\ E_3 &= 4\pi\{-Sq\rho P_2+(S^2\rho-1)\,P_3\}.\end{aligned}\right\} \tag{36}$$

* Recent observations (Booker and Berkner 1938) indicate however that it may be necessary to include this correction.

If we neglect the force exerted on an electron by the magnetic field of the wave, the constitutive relation between **P** and **E** is

$$\ddot{\mathbf{P}}+\nu\dot{\mathbf{P}} = (Ne^2/m)\,\mathbf{E}+(e/mc)\,\dot{\mathbf{P}}\wedge\mathbf{H}^\circ. \tag{37}$$

When **P** and **E** depend on time in the manner given by (32), (37) becomes

$$-u\mathbf{P} = (x/4\pi)\,\mathbf{E}+i\mathbf{P}\wedge\mathbf{y}, \tag{38}$$

or, solving for **E**,

$$\mathbf{E} = 4\pi\{\rho_0\mathbf{P}+i\mathbf{P}\wedge\boldsymbol{\rho}\}. \tag{39}$$

On equating the expressions (36) and (39) for **E**, we obtain

$$\left.\begin{aligned}(\rho_0-\rho)\,P_1+ \qquad i\rho_3P_2- \qquad\qquad i\rho_2P_3 &= 0,\\ -i\rho_3P_1+ \quad (\rho_0-C^2\rho)P_2+ \quad (i\rho_1+Sq\rho)P_3 &= 0,\\ i\rho_2P_1+(-i\rho_1+Sq\rho)\,P_2+(\rho_0-S^2\rho+1)\,P_3 &= 0.\end{aligned}\right\} \tag{40}$$

The first and second of equations (40) give

$$\frac{iP_1}{\rho_3(i\rho_1+Sq\rho)+\rho_2(\rho_0-C^2\rho)} = \frac{P_2}{(\rho_0-\rho)\,(i\rho_1+Sq\rho)+\rho_2\rho_3} = \frac{-P_3}{(\rho_0-\rho)\,(\rho_0-C^2\rho)-\rho_3^2}, \tag{41}$$

and substitution into the third gives, on reduction,

$$(\rho_0+1)\,(\rho_0-\rho)^2-(\rho_1^2+C^2\rho_2^2+S^2\rho_3^2)\,(\rho_0-\rho)-\{S^2\rho_0\rho_2^2+(C^2\rho_0+1)\,\rho_3^2\} = 2\rho_2\rho_3Sq\rho. \tag{42}$$

Substituting for the ρ's in terms of q, u, x and **y**, (41) becomes

$$\begin{aligned}&\frac{iP_1}{[y_3\{iy_1(q^2-C^2)-Sxq\}+y_2\{u(q^2-C^2)+C^2x\}]\,(q^2-C^2)}\\ &\qquad = \frac{P_2}{\{u(q^2-C^2)+x\}\,\{iy_1(q^2-C^2)-Sxq\}+y_2y_3(q^2-C^2)}\\ &\qquad = \frac{-P_3}{\{u(q^2-C^2)+x\}\,\{u(q^2-C^2)+C^2x\}-y_3^2(q^2-C^2)^2},\end{aligned} \tag{43}$$

while (42) takes the form of the quartic equation (7) for q, viz:

$$\alpha q^4+\beta q^3+\gamma q^2+\delta q+\epsilon = 0, \tag{44}$$

where

$$\alpha = u(u^2-y^2)-x(u^2-y_3^2), \tag{45}$$

$$\beta = 2Sxy_2y_3, \tag{46}$$

$$\gamma = 2u\{C^2y^2-(u-x)\,(C^2u-x)\}-x\{y_1^2+C^2y_2^2+(1+C^2)\,y_3^2\}, \tag{47}$$

$$\delta = -2C^2Sxy_2y_3, \tag{48}$$

$$\epsilon = (u-x)\,\{(C^2u-x)^2-C^4y_2^2\}-C^2(C^2u-x)\,(y_1^2+y_3^2). \tag{49}$$

With these values of α, β, γ, δ and ϵ (44) is the fundamental quartic equation of the oblique incidence magneto-ionic theory. Its four roots correspond to the upgoing ordinary wave, the upgoing extraordinary wave, the downcoming ordinary wave and the downcoming extraordinary wave. The elliptical polarization of each of these four waves is obtained by substituting the appropriate value of q into (43).

When the effect of damping is neglected, the quantity u occurring in (45)–(49) is simply unity and we have

$$\begin{cases} \alpha = (1-y^2)-x(1-y_3^2), & (50) \\ \beta = 2Sxy_2y_3, & (51) \\ \gamma = 2\{C^2y^2-(1-x)(C^2-x)\}-x\{y_1^2+C^2y_2^2+(1+C^2)\,y_3^2\}, & (52) \\ \delta = -2C^2Sxy_2y_3, & (53) \\ \epsilon = (1-x)\{(C^2-x)^2-C^4y_2^2\}-C^2(C^2-x)(y_1^2+y_3^2). & (54) \end{cases}$$

When the effect of damping is taken into account, the quantity u occurring in (45)–(49) becomes $1-iz$. The first order corrections to (50)–(54) required when a small amount of damping is taken into account are $i\alpha'$, $i\beta'$, $i\gamma'$, $i\delta'$ and $i\epsilon'$, where

$$\begin{cases} \alpha' = -z\{(3-y^2)-2x\}, & (55) \\ \beta' = 0, & (56) \\ \gamma' = 2z\{C^2\{3-y^2)-2(1+C^2)\,x+x^2\}, & (57) \\ \delta' = 0, & (58) \\ \epsilon' = -z\{C^4(3-y^2)-2C^2(2+C^2)\,x+(1+2C^2)\,x^2\}. & (59) \end{cases}$$

We notice that β' and δ' vanish, while α', γ' and ϵ' are real and are independent of the direction of the earth's magnetic field.

8. Numerical solution of the general case

From §§ 2–7 we see that study of the propagation of wireless waves through a slowly varying plane-stratified ionosphere under the influence of the earth's magnetic field reduces to an examination of the way in which the four roots of the quartic equation (44) for q vary with height above the surface of the earth. Suppose that we know the way in which the electron density and the collisional frequency vary with height. Suppose also that we are dealing with a region of the atmosphere which is sufficiently limited to be able to neglect the variation from point to point in the magnitude and direction of the earth's magnetic field. Then in (45)–(59) $\mathbf{y}$, S and C are to be regarded as prescribed constants, while x and z are known functions of height; our problem is to examine the consequent dependence upon height of the four values of q given by (44).

Although there exist analytical methods of solving an algebraic quartic equation, the only practical way of setting about this problem in general is to resort to one of the standard numerical methods (see for example Whittaker and Robinson 1932). A convenient way of doing this is first of all to neglect completely the effect of damping and to examine the dependence of the four resulting values of q upon electron density, that is, to put $z = 0$ and plot q as a function of x, which is proportional to N, thus obtaining curves of the type shown in fig. 7. If in this way we can obtain the roots of (44) for the case when damping is neglected, it would then be possible to take into account the effect of damping in the way described in § 5; the values of α, β, γ, δ, ϵ, α', β', γ', δ' and ϵ' required in (31) are given by (50)—(59).

We know that, owing to the effect of the earth's magnetic field in producing asymmetry between the propagation of the upgoing and downcoming waves, the (q, N) curves are in general asymmetrical with respect to the N-axis. In the same way the (q, x) curves are in general asymmetrical with respect to the x-axis. To illustrate this phenomenon clearly let us consider a case in which the direction of incidence is parallel to the earth's magnetic field and the direction of reflexion perpendicular to it. This occurs at an angle of incidence of 45° for propagation in the magnetic meridian-plane

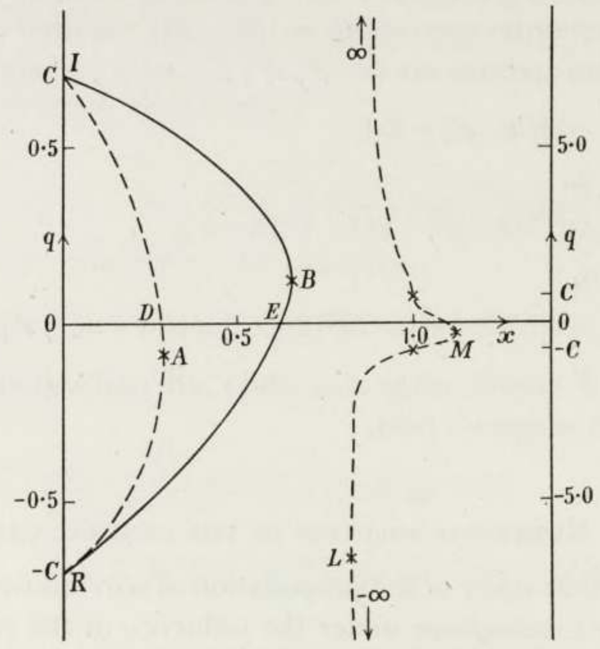

FIG. 15. q as a function of x, $y < 1$.

in latitudes where the magnetic dip is $\pm 45°$. Let us therefore draw the (q, x) curves for the case when

$$y_1 = 0, \quad y_2 = y_3 = \pm y/\sqrt{2}, \quad S = C = 1/\sqrt{2}. \tag{60}$$

Fig. 15 shows the (q, x) curves obtained for a wave-frequency twice the magneto-ionic frequency $(y = \frac{1}{2})$, and fig. 16 those for a wave-frequency half the magneto-ionic frequency $(y = 2)$. IR, AR represent the upgoing and downcoming extraordinary waves, while IB, BR represent the upgoing and downcoming ordinary waves. The values x_A and x_B of x corresponding to the points A and B give the critical electron densities required for reflexion of the extraordinary and ordinary waves respectively. The values x_D and x_E corresponding to the points D and E give the electron densities for which the directions of phase-propagation in the upgoing extraordinary wave and the downcoming ordinary wave respectively are horizontal. We can now see how important it is to calculate the critical electron densities required for reflexion of the extraordinary and ordinary waves from the values x_A and x_B of x, and not x_D and x_E. In the case of the ordinary wave in fig. 16 the electron density corresponding to the point B is about 18 % greater than that corresponding to the point E; it is about 60 % greater than it would be in the absence of the earth's magnetic field, when the critical electron density would be given by $x = C^2 = \frac{1}{2}$. The corresponding percentages for the ordinary wave in fig. 15 are 5 % and 30 %, the influence of the earth's magnetic field being less for the higher frequency. We notice that the discrepancy between the electron density for which the direction of group-propagation becomes horizontal and the electron density for which the direction of phase-propagation becomes horizontal is less marked for the extraordinary wave than for the ordinary wave. This can be connected up with the fact

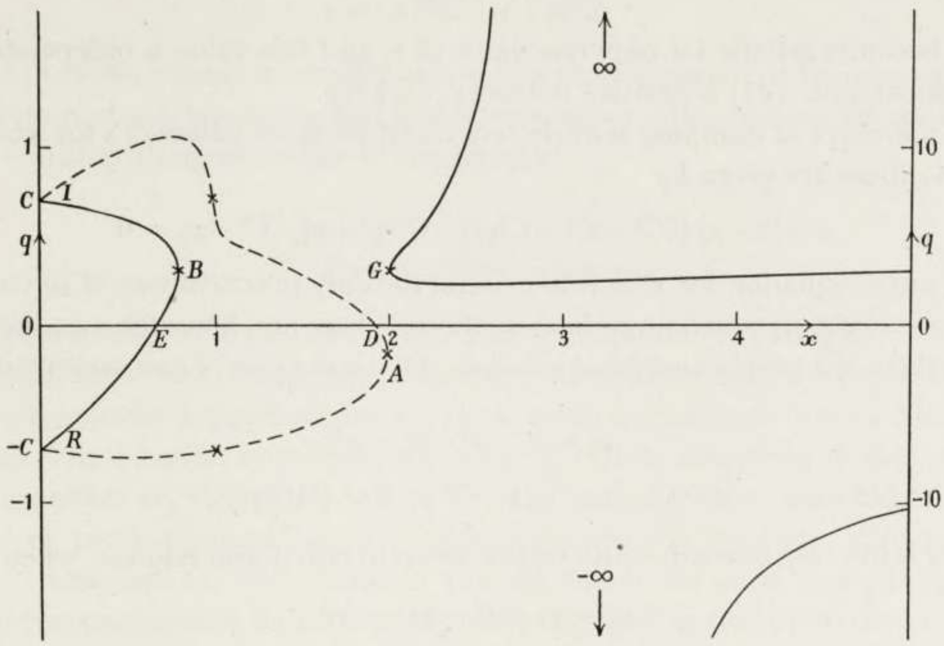

FIG. 16. q as a function of x, $y > 1$.

that the critical electron density required to reflect the ordinary wave in the case of vertical incidence at a magnetic pole differs from what it is in the case of vertical incidence at the magnetic equator, whereas the critical electron density required to reflect the extraordinary wave is the same in both cases.

The branches of the (q, x) curves in figs. 15 and 16 occurring for values of x greater than those which produce reflexion are of no interest so long as we are concerned with waves incident from below upon an ionosphere the variation of whose constitution with height, per vacuum wave-length upon 2π, is small. The scale of q for these branches of the (q, x) curves has been reduced by a factor of ten.

It may be mentioned that, if the direction of incidence and the direction reverse to that of reflexion are interchanged, the (q, x) curves are simply reflected in the x-axis. For (44)–(49) show that reversal of the sign of S merely reverses the sign of q. This implies that at every level in the ionosphere on both the upward and downward journeys the phase-propagation and the group-propagation of the separate magneto-ionic components are simply reversed, while from (43) their absolute elliptical polarizations in space remain unaltered.

Although the zeros and infinity of q are not in general associated with the phenomenon of reflexion, it is nevertheless helpful to be able to locate them easily when plotting diagrams such as figs. 15 and 16. We see from (50) that, when the effect of damping is neglected, a root of (44) is infinite when

$$\alpha \equiv (1-y^2)-x(1-y_3^2) = 0,$$

that is, when
$$x = \frac{1-y^2}{1-y_3^2}. \tag{61}$$

Hence q becomes infinite for only one value of x, and this value is independent of the angle of incidence. (61) is positive unless $|y_3| \leqslant 1 \leqslant y$.

When the effect of damping is neglected, there are three values of x for which $q = 0$. From (54) these are given by

$$\epsilon \equiv (1-x)\{(C^2-x)^2-C^4y_2^2\}-C^2(y_1^2+y_3^2)(C^2-x) = 0. \tag{62}$$

This is a cubic equation for x, and in general the only practical way of solving it is to resort to one of the standard numerical methods. There are, however, two special cases in which there is a simple analytical solution. One is the case of east-west transmission: when

$$y_2 = 0, \quad y_1^2+y_3^2 = y^2,$$

the roots of (62) are
$$x = C^2 \quad \text{or} \quad \tfrac{1}{2}\{1+C^2 \pm \surd(S^4+4C^2y^2)\}. \tag{63}$$

The other is the case of north-south transmission in equatorial regions: when

$$y_1 = y_3 = 0, \quad y_2 = \pm y,$$

the roots of (62) are
$$x = 1 \quad \text{or} \quad C^2(1 \pm y). \tag{64}$$

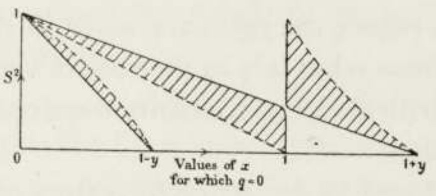

FIG. 17. The zeros of q, $y < 1$.

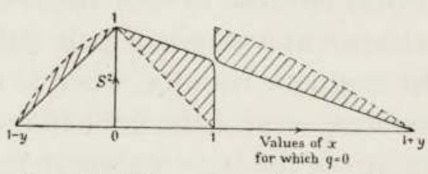

FIG. 18. The zeros of q, $y > 1$.

The nature of the dependence of the roots of (62) upon the angle of incidence is shown in figs. 17 and 18 for wave-frequencies above and below the magneto-ionic frequency respectively. The values of x for which $q = 0$ are plotted horizontally against S^2 plotted vertically. Fig. 17 refers to a wave-frequency twice the magneto-ionic frequency ($y = \frac{1}{2}$) and fig. 18 to a wave-frequency half the magneto-ionic frequency ($y = 2$). The broken curves refer to the special case of east-west transmission and are plotted from the values (63) of x, while the continuous curves refer to the special case of north-south transmission in equatorial regions and are plotted from the values (64) of x. Although in the general case the analytical solution of the cubic equation (62) appears to be of little practical value, nevertheless we can prove that its three roots are real and are represented in figs. 17 and 18 by curves lying in the three shaded regions, so that for a given angle of incidence intervals of x within which the three zeros of q must lie are given by the intercepts of these shaded regions upon a horizontal line drawn for the appropriate value of S^2. We have for any of the values (63) of x

$$\epsilon = -S^2C^2y_2^2x, \tag{65}$$

and for any of the values (64) of x

$$\epsilon = +S^2C^2(y_1^2+y_3^2)\,x. \tag{66}$$

Hence, if ϵ does not vanish on the boundary of a shaded region, it has one sign on the broken portion of the boundary and the reverse sign on the continuous portion, and therefore vanishes along some line in the interior.

9. East-west transmission

Although the analytical solution of the quartic equation (44) for q appears to be of little practical value in general, there are nevertheless some interesting special cases in which (44) is actually a quadratic for q^2; the analytical solution is then of considerable value. This simplification occurs when $\beta = \delta = 0$, which, according to (46) and (48), occurs when either $H_2^\circ = 0$, or $H_3^\circ = 0$, or $S = 0$. The first of these cases is that in which the plane of phase-propagation is perpendicular to the magnetic meridian-plane (east-west transmission); the second is that in which the earth's magnetic field is horizontal (propagation in equatorial regions); the third is that of vertical incidence. The last case may be included in the first for when incidence is vertical the plane of phase-propagation is indeterminate and so axes 1 and 2 may be chosen to make $H_2^\circ = 0$. We shall examine the combined first and third cases in the present section, and the second case in the following section.

When (44) is a quadratic for q^2, (q, x) curves of the type indicated in figs. 15 and 16 become symmetrical about the x-axis (cf. fig. 6), and it is then more convenient to draw curves showing the variation of q^2 with x (cf. fig. 5). On account of the symmetry which we now have in the (q, x) curves, the zeros and infinity of q (or q^2), which are given by (61) and (62), are double roots of (44) of the type to be associated with the phenomenon of reflexion. Moreover, transitions of the type discussed on pp. 431 *et seq.* must occur in pairs, one for the upgoing waves and the other for the downcoming waves, and so a wave-packet which has the propagation characteristics of the ordinary wave when it enters the ionosphere must necessarily have the propagation characteristics of the ordinary wave when it leaves the ionosphere even though over part of its path it may be propagated like an extraordinary wave-packet. In order not to overlook the effect of transitions it is desirable to plot, not the variation of q^2 with x when $z = 0$, but the variation of $\lim_{z \to 0} q^2$ with x.

When the plane of phase-propagation is perpendicular to the magnetic meridian plane ($H_2^\circ = 0$), it is more convenient to go back to equations (41) and (42) and put $\rho_2 = 0$, rather than to put $y_2 = 0$ in (43) and (44). If we do this, we obtain

$$\left\{\begin{array}{l} \dfrac{P_2}{P_1} = \dfrac{\rho_0 - \rho}{-i\rho_3}, \qquad (67) \\ (\rho_0+1)(\rho_0-\rho)^2 - (\rho_1^2 + S^2\rho_3^2)(\rho_0-\rho) - (C^2\rho_0+1)\rho_3^2 = 0. \qquad (68) \end{array}\right.$$

Since ρ contains q^2, (67) may be rearranged so as to express q^2 in terms of (P_2/P_1). Also (68) is a quadratic equation for $(\rho_0 - \rho)$ and in view of (67) may be rewritten as a quadratic for (P_2/P_1). If we carry out these transformations and substitute for the ρ_0, ρ_1 and ρ_3 in terms of u, x, y_1 and y_3, we obtain

$$\left\{\begin{array}{l} q^2 = C^2 - \dfrac{x}{u + iy_3(P_2/P_1)}, \qquad (69) \\ (u-x)P_2^2 - 2im_3 P_2 P_1 + (C^2u - x)P_1^2 = 0, \qquad (70) \end{array}\right.$$

where

$$m_3 = (y_1^2 + S^2y_3^2)/2y_3. \qquad (71)$$

(P_2/P_1) represents the projection on to the 12-plane of the elliptical polarization of a magneto-ionic component, and the two values of (P_2/P_1) given by the quadratic (70) refer to the ordinary and extraordinary waves. The corresponding values of q^2 are obtained by substituting the appropriate values of (P_2/P_1) into (69). The explicit expression for q^2 in terms of x, y_1, y_3, z, S and C is

$$q^2 = C^2 - \frac{x(u-x)}{u(u-x) - \frac{1}{2}(y_1^2 + S^2y_3^2) \pm \sqrt{\{\frac{1}{4}(y_1^2+S^2y_3^2)^2 + y_3^2(C^2u-x)(u-x)\}}}, \qquad (72)$$

where u is unity when damping is neglected and $1 - iz$ when damping is taken into account.

If we put $S = 0$ and $C = 1$ in (69)–(72), we recover the well-known relations which apply to the special case of vertical incidence and which have been derived by Appleton (1932). Methods for dealing with equations (69) and (70) in the case of vertical incidence have been adequately discussed by a number of writers (already cited on p. 413), and since these methods require little modification in the more general case of east-west transmission there is no need to discuss them in detail here.

Typical curves showing the variation with x of $\lim_{z \to 0} q^2$ are shown in figs. 19 and 20 for wave-frequencies above and below the magneto-ionic frequency respectively. Fig. 19 is drawn for a frequency twice the magneto-ionic frequency ($y = \frac{1}{2}$) and for an angle of incidence of 45° ($S^2 = C^2 = \frac{1}{2}$). The series of curves shows the evolution of east-west transmission in going from the magnetic equator to a magnetic pole, that is, as $|y_1/y|$ decreases from unity to zero and $|y_3/y|$ increases from zero to unity, or in other words as the quantity $|m_3|$ defined in (71) decreases from infinity to $\frac{1}{2}S^2y$. The continuous curves refer to the ordinary wave and the broken curves to the extraordinary wave. The infinity of q^2 is given by (61) and the zeros by (63). We see that for east-west transmission using wave-frequencies greater than the magneto-ionic frequency the critical electron density required for reflexion of the extraordinary wave is given by

$$x = \tfrac{1}{2}\{1+C^2-\sqrt{(S^4+4C^2y^2)}\}, \tag{73}$$

while that for the ordinary wave is given by

$$x = C^2. \tag{74}$$

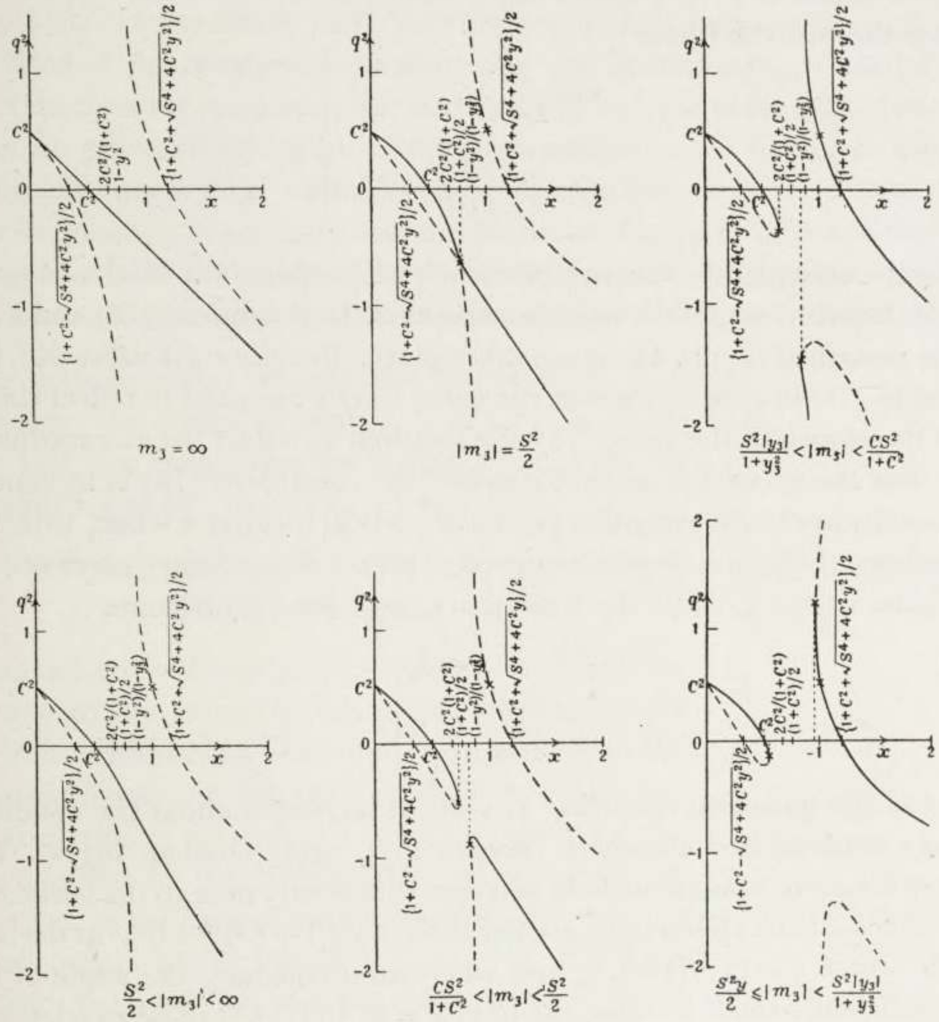

FIG. 19. East-west transmission, $y < 1$.

The value (73) of x does not vary as we go from the magnetic equator to a magnetic pole so long as there is no variation in the angle of incidence or in the ratio of the wave-frequency to the magneto-ionic frequency. The value (74) of x is entirely independent of the earth's magnetic field, so that for east-west transmission with a given angle of incidence the critical electron density required to reflect the ordinary wave is the same as if the effect of the earth's magnetic field were neglected. The linear nature of the dependence of q^2 upon x for the ordinary wave in the case $m_3 = \infty$ in fig. 19 shows that for east-west transmission at the magnetic equator the entire propagation of the ordinary wave is completely unaffected by the presence of the earth's magnetic field; it may be mentioned that in this special case variation of the angle of incidence merely involves moving the (q^2, x) curves bodily parallel to the q^2-axis.

We notice in fig. 19 that, as $|m_3|$ decreases through the value $CS^2/(1+C^2)$, the ordinary and extraordinary waves exchange their types of propagation for values of x greater than $2C^2/(1+C^2)$. This is because, when the effect of damping is taken into account, the two complex values of q^2 corresponding to the ordinary and extraordinary waves are equal when the two conditions

$$\left.\begin{aligned} x &= \frac{2C^2}{1+C^2}, \\ m_3^2 &= \left\{\frac{CS^2}{1+C^2}\right\}^2 + C^2z^2 \end{aligned}\right\} \tag{75}$$

are satisfied; consequently the independence of propagation of the magneto-ionic components breaks down in the neighbourhood of the conditions (75) and a transition of the type described on pp. 431 *et seq.* takes place. But since the value $2C^2/(1+C^2)$ of x occurring in (75) in general exceeds the value C^2 of x required to reflect the ordinary wave and therefore also the value (73) of x required to reflect the extraordinary wave, it follows that the transition occurring under the conditions (75) is in general of no practical significance. An exception occurs at vertical incidence when, with $S = 0$ and $C = 1$, the critical electron density required to reflect the ordinary wave becomes that corresponding to $x = 1$, while the transition conditions (75) become

$$\left.\begin{aligned} x &= 1, \\ \left|\frac{y_1^2}{2zy_3}\right| &= 1. \end{aligned}\right\} \tag{76}$$

The effect of the transition occurring at vertical incidence under the conditions (76) has already been studied elsewhere (Booker 1934, 1935; Goubau 1935*a*). When the direction of the earth's magnetic field becomes sufficiently near to the vertical to make (cf. (75)) m_3^2 cease to be appreciably greater than $\{CS^2/(1+C^2)\}^2 + C^2z^2$ at the level in the ionosphere where $x = 2C^2/(1+C^2)$, and when simultaneously the angle of incidence becomes sufficiently small to cause the level $x = 2C^2/(1+C^2)$ to be overlapped by the stratum in the neighbourhood of the level $x = C^2$ within which reflexion of the ordinary wave is taking place, then the ordinary wave begins to penetrate this level and to be reflected from the level where the electron density is given by

$$x = \tfrac{1}{2}\{1 + C^2 + \sqrt{(S^4 + 4C^2y^2)}\}. \tag{77}$$

When the condition of vertical incidence at a magnetic pole is reached, the ordinary wave is wholly reflected from the level (77), which then takes the simpler form $x = 1+y$. These remarks are subject of course to the restriction that the effect of damping is not so marked as to obliterate reflexion for all practical purposes.

As the wave-frequency decreases to the magneto-ionic frequency, the values (61) and (73) of x decrease to zero. When the thickness of the stratum in the ionosphere between the zero (73) and the infinity (61) of q^2 becomes comparable with a vacuum wave-length upon 2π, the extraordinary wave begins to penetrate this stratum,* penetration becoming complete when the wave-frequency reaches the magneto-ionic frequency. Penetration only begins however when y differs from unity by a quantity of the order of magnitude of the rate of change of $x(\propto N)$ with height per vacuum wave-length upon 2π near the bottom of the ionosphere, and it must be remembered that when damping is taken into account the extraordinary wave is strongly absorbed for wave-frequencies in the neighbourhood of the magneto-ionic frequency (cf. Booker 1935, p. 282). For wave-frequencies less than the magneto-ionic frequency ($y>1$), the value (73) of x is negative and so has no physical significance. As the wave-frequency decreases through the magneto-ionic frequency the value (61) of x also becomes negative, but with further decrease of frequency it in general passes to a positive value via infinity and then refers to the ordinary instead of the extraordinary wave. A typical example of the variation with x of $\lim_{z\to 0} q^2$ for a wave-frequency less than the magneto-ionic frequency is shown in fig. 20.

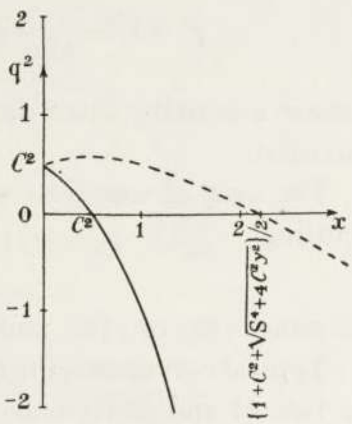

FIG. 20. East-west transmission, $y>1$.

The figure actually refers to east-west transmission in a latitude where the magnetic dip is $\pm 45^\circ$ for a wave-frequency half the magneto-ionic frequency and for an angle of incidence of 45°, that is, to the case when

$$\left.\begin{aligned} y_1^2 = y_3^2 = \tfrac{1}{2}y^2, \quad y_2 = 0, \quad y = 2, \\ S^2 = C^2 = \tfrac{1}{2}. \end{aligned}\right\} \tag{78}$$

The continuous curve refers to the ordinary wave and the broken curve to the extraordinary wave. We see that for east-west transmission using wave-frequencies less than the magneto-ionic frequency the critical electron density required for reflexion of the ordinary wave is in general given by (74), while that for the extraordinary wave is given by (77). However, as the condition of vertical incidence at a magnetic pole is approached a transition occurs under the conditions (75) and thereafter the ordinary wave is reflected from the level (77) and the extraordinary wave is not reflected at all.

* No possibility seems to exist of separate reflections from the upper and lower boundaries of the stratum in the manner suggested by Saha and Rai 1937; this seems clear physically and may be rigorously proved mathematically.

10. Propagation in equatorial regions

When the earth's magnetic field is horizontal ($H_3^\circ = 0$), we put $\rho_3 = 0$ in (41) and (42) and obtain (cf. (67) and (68))

$$\begin{cases} \dfrac{P_3}{P_1} = \dfrac{\rho_0 - \rho}{i\rho_2}, & (79) \\ (\rho_0+1)(\rho_0-\rho)^2 - (\rho_1^2 + C^2\rho_2^2)(\rho_0-\rho) - S^2\rho_0\rho_2^2 = 0, & (80) \end{cases}$$

which may be rewritten as (cf. (69) and (70))

$$\begin{cases} q^2 = C^2 - \dfrac{x}{u - iy_2(P_3/P_1)}, & (81) \\ (u-x)\,P_3^2 + 2im_2SP_3P_1 + uS^2P_1^2 = 0, & (82) \end{cases}$$

where
$$m_2 = (y_1^2 + C^2y_2^2)/2Sy_2. \tag{83}$$

(P_3/P_1) represents the projection on to the 13-plane of the elliptical polarization of a magneto-ionic component, and the two values of (P_3/P_1) given by the quadratic (82) refer to the ordinary and extraordinary waves. The corresponding values of q^2 are obtained by substituting the appropriate values of (P_3/P_1) into (81). The explicit expression for q^2 in terms of x, y_1, y_2, z, S and C is

$$q^2 = C^2 - \frac{x(u-x)}{u(u-x) - \frac{1}{2}(y_1^2 + C^2y_2^2) \pm \sqrt{\{\frac{1}{4}(y_1^2 + C^2y_2^2)^2 + S^2y_2^2u(u-x)\}}}, \tag{84}$$

where u is unity when damping is neglected and $1 - iz$ when damping is taken into account.

The case of east-west transmission at the magnetic equator may be obtained by putting

$$y_1 = \pm y, \quad y_2 = y_3 = 0$$

in either (84) or (72), and it is clear that the two formulae give the same result.

Typical curves showing the variation of q^2 with x when damping is neglected are shown in figs. 21 and 22 for wave-frequencies above and below the magneto-ionic frequency respectively. Fig. 21 actually refers to propagation in equatorial regions in a plane making an angle of 45° with the magnetic meridian-plane for a wave-frequency twice the magneto-ionic frequency and for an angle of incidence of 45°, that is, to the case when

$$\left.\begin{aligned} y_1^2 = y_2^2 = \tfrac{1}{2}y^2, \quad y_3 = 0, \quad y = \tfrac{1}{2}, \\ S^2 = C^2 = \tfrac{1}{2}. \end{aligned}\right\} \tag{85}$$

The continuous curve refers to the ordinary wave and the broken curve to the extraordinary wave. From (61) we see that the infinity of q^2 is given by

$$x = 1 - y^2, \tag{86}$$

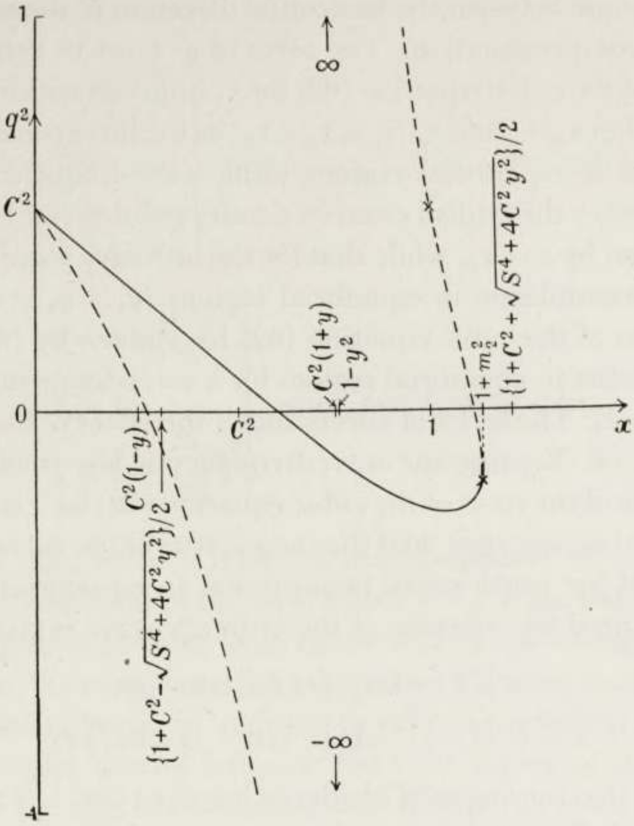

FIG. 21. Propagation in equatorial regions, $y<1$.

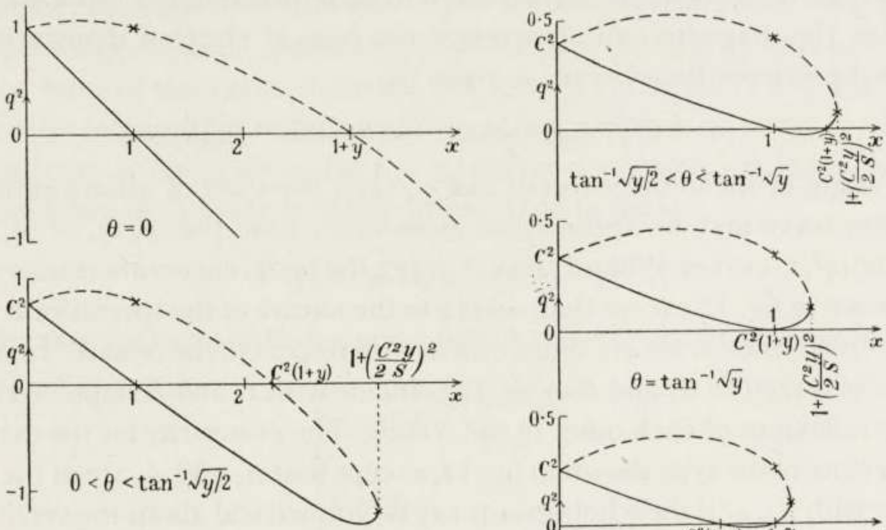

FIG. 22. North-south transmission in equatorial regions, $y>1$.

independent of the angle between the horizontal direction of the earth's magnetic field and the plane of phase-propagation. The zeros of q^2 must in general be obtained by numerical solution of the cubic equation (62) for x; intervals within which they must lie are given by fig. 17. Let x_D, x_E and x_F ($x_D \leqslant x_E \leqslant x_F$) be the three roots of (62). Then we see that for propagation in equatorial regions using wave-frequencies greater than the magneto-ionic frequency the critical electron density required for reflexion of the extra-ordinary wave is given by $x = x_D$, while that for the ordinary wave is given by $x = x_E$.

For north-south transmission in equatorial regions ($y_1 = y_3 = 0$, $y_2 = \pm y$) there is an analytical solution of the cubic equation (62) for x given by (64). Fig. 22 refers to north-south transmission in equatorial regions for a wave-frequency half the magneto-ionic frequency ($y = 2$). The series of curves shows the effect of increasing the angle of incidence from 0 to $\frac{1}{2}\pi$. Because the wave-frequency is less than the magneto-ionic frequency ($y > 1$), one of the roots of the cubic equation (62) for x is negative (cf. fig. 18) and so has no physical significance, and the same is true of the value (86) of x where q^2 is infinite. We see that for north-south transmission in equatorial regions the critical electron density required for reflexion of the ordinary wave is given by

$$x = 1 \quad \text{when} \quad 0 \leqslant \theta \leqslant \tan^{-1}\sqrt{y}, \tag{87}$$

and

$$x = C^2(1+y) \quad \text{when} \quad \tan^{-1}\sqrt{y} \leqslant \theta \leqslant \tfrac{1}{2}\pi. \tag{88}$$

We notice from (87) that for angles of incidence less than $\tan^{-1}\sqrt{y}$ the level of reflexion of the ordinary wave is independent of the value of the angle of incidence. We also see from fig. 22 that for north-south transmission in equatorial regions using wave-frequencies less than the magneto-ionic frequency the critical electron density required for reflexion of the extraordinary wave is given by

$$x = C^2(1+y) \quad \text{when} \quad 0 \leqslant \theta \leqslant \tan^{-1}\sqrt{(\tfrac{1}{2}y)}. \tag{89}$$

When the angle of incidence exceeds $\tan^{-1}\sqrt{(\frac{1}{2}y)}$, the state of affairs existing for the extraordinary wave may be appreciated more easily from the (q, x), or (q, N), curves than from the (q^2, x) curves. When $\theta > \tan^{-1}\sqrt{(\frac{1}{2}y)}$, the (q, N) curves are readily seen to take the form shown in fig. 13 except that, owing to the nature of the special case with which we are concerned, the curves are symmetrical with respect to the N-axis. In these special circumstances therefore A_2 and B in fig. 13 coincide with D and E respectively, while A_1 and A_3 are reflexions of each other in the N-axis. The group-ray for the extraordinary wave is therefore of the type shown in fig. 14, except that A_1 and A_3 are at the same level, D coincides with A_2, and the whole group-ray is symmetrical about the vertical through A_2. It is clearly the common level of A_1 and A_3 which gives the critical electron density required to reflect the wave, and it therefore follows that for north-south transmission in equatorial regions using wave-frequencies less than the magneto-ionic frequency the critical electron density required for reflexion of the extraordinary wave is given by

$$x = 1 + \left(\frac{C^2 y}{2S}\right)^2 \quad \text{when} \quad \tan^{-1}\sqrt{(\tfrac{1}{2}y)} \leqslant \theta < \tfrac{1}{2}\pi. \tag{90}$$

The extraordinary wave is reflected in succession at the levels

$$\left.\begin{aligned} x &= 1+\left(\frac{C^2y}{2S}\right)^2 \\ x &= C^2(1+y) \\ x &= 1+\left(\frac{C^2y}{2S}\right)^2 \end{aligned}\right\} \text{ when } \tan^{-1}\surd(\tfrac{1}{2}y)<\theta\leqslant\tan^{-1}\surd y, \tag{91}$$

and

$$\left.\begin{aligned} x &= 1+\left(\frac{C^2y}{2S}\right)^2 \\ x &= 1 \\ x &= 1+\left(\frac{C^2y}{2S}\right)^2 \end{aligned}\right\} \text{ when } \tan^{-1}\surd y\leqslant\theta<\tfrac{1}{2}\pi. \tag{92}$$

It should be noted that, when θ is in the neighbourhood of $\tan^{-1}\surd y$, the reflecting stratum in the neighbourhood of the level where $x = 1$ is partially penetrable both for the ordinary wave-packet arriving from below and for the extraordinary wave-packet arriving from above. We may also point out that for glancing incidence (θ nearly equal to $\frac{1}{2}\pi$) the extraordinary wave, in addition to suffering reflexion from the level where $x = 1$, would also suffer partial reflexion from the region of small electron density (provided of course the effect of damping were sufficiently small).

When the plane of propagation does not necessarily coincide with the magnetic meridian-plane, the state of affairs existing in equatorial regions for wave-frequencies less than the magneto-ionic frequency is as follows. Let x_D, x_E and x_F ($x_D\leqslant x_E\leqslant x_F$) be the three roots of the cubic equation (62) giving the zeros of q. Then the critical electron density required for reflexion of the ordinary wave is given by $x = x_E$. That for the extraordinary wave is given by $x = x_F$ unless conditions are sufficiently near to glancing incidence in the magnetic meridian-plane to make

$$4S^4y_2^2(C^2y_2^2-y_1^2)>(C^2y_2^2+y_1^2)^3. \tag{93}$$

In the latter case the extraordinary wave suffers reflexion in succession at the levels

$$x = 1+\left\{\frac{y_1^2+C^2y_2^2}{2Sy_2}\right\}^2, \tag{94}$$

$$x = x_F, \tag{95}$$

$$x = 1+\left\{\frac{y_1^2+C^2y_2^2}{2Sy_2}\right\}^2, \tag{96}$$

the critical electron density being given by (94) and (96). (94) and (96) are obtained by expressing the condition that the two values (84) of q^2 are equal, and (93) by expressing the condition that this common value of q^2 is positive.

11. Summary

A study is made of the propagation of wave-packets incident obliquely upon a slowly varying plane-stratified doubly refracting medium with a view to applying the results to the ionosphere. The magneto-ionic theory in the form given by Appleton is only suitable for investigating vertical propagation in the ionosphere. The theory is generalized so as to be capable of describing oblique propagation of a magneto-ionic component through the ionosphere without using a refractive index which depends in a complicated way upon an unknown angle of refraction. The fundamental formula of the oblique-incidence magneto-ionic theory is an algebraic quartic equation for a quantity q which depends upon the prescribed angle of incidence and which at vertical incidence becomes identical with the well-known refractive index. The four roots of the quartic equation for q correspond to the upgoing ordinary wave, the upgoing extraordinary wave, the downcoming ordinary wave and the downcoming extraordinary wave. The level in the ionosphere where individual wave-crests are moving horizontally across a characteristically polarized wave-packet is given by the condition $q = 0$, which is equivalent to putting the angle of refraction equal to $\frac{1}{2}\pi$ in Snell's Law. But the level of reflexion of the magneto-ionic component is the level where the wave-packet *as a whole* is travelling horizontally and is given by the condition that the root of the quartic equation for q corresponding to the upgoing magneto-ionic component should be equal to a root corresponding to a downcoming magneto-ionic component.

The detailed application of these results to the ionosphere depends on our ability to solve the fundamental quartic equation for q, and the only practical way of achieving this in general is to resort to one of the standard numerical methods. Special cases worked out in this way show that the critical electron density required for reflexion of a magneto-ionic component can easily be as much as 25 % in excess of the erroneous value which would be calculated by putting the angle of refraction equal to $\frac{1}{2}\pi$ in Snell's Law.

The analytical solution of the quartic equation for q is sufficiently simple to be of practical value in the case of east-west transmission (including vertical incidence) and in the case of propagation in equatorial regions. Owing to the symmetrical influence of the earth's magnetic field upon the propagation of the upgoing and downcoming waves in these special cases, it is legitimate to calculate the critical electron densities of the magneto-ionic components from the condition ($q = 0$) that the direction of phase-propagation is horizontal. Representative curves are drawn.

References

Appleton 1925 *Proc. Phys. Soc.* **37**, 16 D.
—— 1932 *J. Instn Elect. Engrs*, **71**, 642.
Appleton and Builder 1933 *Proc. Phys. Soc.* **45**, 208.
Bailey 1934 *Phil. Mag.* **18**, 516.
Baker and Green 1933 *Proc. Inst. Radio Engrs, N.Y.*, **21**, 1103.

Booker 1934 *Proc. Roy. Soc.* A, **147**, 352.
—— 1935 *Proc. Roy. Soc.* A, **150**, 267.
—— 1936 *Proc. Roy. Soc.* A, **155**, 235.
Booker and Berkner 1938 *Nature* **141**, 562.
Breit 1927 *Proc. Inst. Radio Engrs, N.Y.* **15**, 709.
Darwin 1934 *Proc. Roy. Soc.* A, **146**, 17.
Försterling and Lassen 1933*a* *Hochfrequenztech. u. Electroakust.* **42**, 158.
—— —— 1933*b* *Ann. Phys., Lpz.*, **18**, 26.
Gans 1915 *Ann. Phys., Lpz.*, **47**, 709.
Goldstein 1928 *Proc. Roy. Soc.* A, **121**, 260.
Goubau 1934 *Hochfrequenztech. u. Elektroakust.* **44**, 17, 138.
—— 1935*a* *Hochfrequenztech. u. Elektroakust.* **45**, 179.
—— 1935*b* *Hochfrequenztech. u. Elektroakust.* **46**, 37.
Hartree 1929 *Proc. Camb. Phil. Soc.* **25**, 97.
—— 1931*a* *Proc. Camb. Phil. Soc.* **27**, 143.
—— 1931*b* *Proc. Roy. Soc.* A, **131**, 428.
Martyn 1935 *Phil. Mag.* **19**, 376.
Nichols and Schelleng 1925 *Bell System Technical Journal* **4**, 215.
Ratcliffe 1933 *Wireless Engr*, **10**, 354.
Saha and Rai 1937 *Proc. Nat. Inst. Sci. India*, **3**, 359.
Schekulin 1930 *Hochfrequenztech. u. Elektroakust.* **36**, 172.
Taylor 1933 *Proc. Phys. Soc.* **45**, 245.
—— 1934 *Proc. Phys. Soc.* **46**, 408.
Whittaker and Robinson 1932 "Calculus of Observations", chap. 6.