15

Full wave solutions for isotropic ionosphere

15.1. Introduction

This and the remaining chapters are mainly concerned with problems in which the approximations of ray theory cannot be made, so that a more detailed study of the solutions of the governing differential equations is needed. The solution of this type of problem is called a 'full wave solution'. The subject has been studied almost entirely for plane stratified media. When the medium is not plane stratified, for example when the earth's curvature is allowed for, a possible method is to use an 'earth flattening' transformation that converts the equations into those for an equivalent plane stratified system; see §§ 10.4, 18.8. See Westcott (1968, 1969) and Sharaf (1969) for examples of solutions that do not use such a transformation.

This chapter is written entirely in terms of a horizontally stratified ionosphere and for frequencies greater than about 1 kHz, so that the effect of positive ions in the plasma may be neglected. The earth's curvature also is neglected.

The need for full wave solutions arises when the wavelength in the medium is large so that the electric permittivity changes appreciably within one wavelength and the medium cannot be treated as slowly varying (§§ 7.6, 7.10). Thus it arises particularly at very low frequencies, say 1 to 500 kHz. But even at higher frequencies the wavelength in the medium is large where the refractive index n is small, and some full wave solutions are therefore often needed for high frequencies too.

In this chapter the earth's magnetic field is neglected. This means that the differential equations to be studied are of the second order, for example (7.6), (7.68). The standard differential equations of theoretical physics are nearly all of the second order. They have been very thoroughly investigated and the number of papers and books on them is extremely large. Many of the simpler problems of radio propagation can be reduced to a study of these standard forms. The particular standard differential equation to be used depends on two functions, the electron height distribution $N(z)$ and the collision frequency $\nu(z)$. In most of the examples studied here ν is independent of the height z. For an example with a variable ν see problem 15.1. Since $N(z)$ must conform to the requirements of the standard differential equations, the forms that it can take are restricted, but still a very wide range of functions $N(z)$ can be studied. Two important uses of these standard solutions are (1) the study of some physical phenomena, for example partial penetration and reflection, §§ 15.4, 15.10, 15.17, and the various forms of tunnelling, §§ 15.9, 17.6–17.9, 19.4, 19.6; (2) for testing whether computer programs intended for more general functions $N(z)$, $\nu(z)$, are free from errors.

The coordinate system used here is the same as in ch. 6 onwards, so that z is height. The wave normals are assumed to be in the plane $y = 0$ which is called the plane of incidence. This chapter deals with the problem where the incident wave enters the ionosphere from the free space below it. It is assumed to be a plane wave with its wave normal at an angle θ to the z axis, and $S = \sin\theta$, $C = \cos\theta$. The variable q used throughout this chapter is defined by (7.2).

The differential equations for the wave fields in a horizontally stratified ionosphere were derived in § 7.2 and it was shown that they separate into two independent sets, (7.3) leading to (7.6), and (7.4) leading to (7.68). The first of these is for horizontal polarisation and most of this chapter applies to it. The other is for vertical polarisation, equation (7.68), and this is more complicated because of the term in $\mathrm{d}\mathscr{H}_y/\mathrm{d}z$. It is dealt with in §§ 15.5–15.7 where it is shown that, in some conditions, results for horizontal polarisation can be applied also for vertical polarisation.

Many of the methods and results in this chapter have been given by Jacobsson (1965) who studied the reflection of light from isotropic stratified films.

15.2. Linear electron height distribution

The case where $N(z)$ within the ionosphere is a linear function of z has already been dealt with in ch. 8. The differential equation for horizontal polarisation was (8.5) and it was transformed to the standard equation (8.7) which is the Stokes equation. It uses the independent variable ζ given by (8.6). This method will now be applied to the model ionosphere described in §§ 12.3, 12.7 in which there is free space below the level $z = h_0$, and above this X, proportional to $N(z)$, is given by

$$X = p(z - h_0), \quad q^2 = C^2 - \frac{p}{1 - \mathrm{i}Z}(z - h_0) \tag{15.1}$$

where p is a real constant. It is assumed that the collision frequency and thence Z is a constant. Then in (8.2), (8.5), (8.6) the height z_0 where q is zero is

$$z_0 = h_0 + (1 - \mathrm{i}Z)C^2/p, \tag{15.2}$$

the constant a in (8.2) is

$$a = p/(1 - \mathrm{i}Z), \tag{15.3}$$

and (8.6) gives

$$\zeta = \left(\frac{k^2 p}{1 - \mathrm{i}Z}\right)^{\frac{1}{3}}\{z - h_0 - (1 - \mathrm{i}Z)C^2/p\}. \tag{15.4}$$

The new independent variable ζ is thus a new measure of height, in which the origin is taken where $q = 0$, that is at the height $z = z_0$ (15.2), which is complex if $Z \neq 0$. The scale is changed by the complex factor

$$g = \{k^2 p/(1 - \mathrm{i}Z)\}^{\frac{1}{3}} \tag{15.5}$$

and since this is a cube root it may be chosen in three different ways. In § 8.20, for the collisionless plasma $Z = 0$, it was chosen to be real and positive. We now choose the value that goes continuously into a real positive value when $Z \to 0$. Hence

$$\arg g = \tfrac{1}{3}\arctan Z. \tag{15.6}$$

The change of scale includes a rotation in the complex z plane through an angle arg g, which must be less than $\frac{1}{6}\pi$.

A solution of (8.7) must now be chosen that satisfies the physical conditions at great heights, that is when z is real, positive and large. Then $|\zeta|$ is large and $\arg\zeta \to \arg g$ so that $\arg\zeta$ is within the range $0 \leqslant \arg\zeta < \frac{1}{6}\pi$. The asymptotic approximation of the solution Ai(ζ) is given by (8.52) which contains the factor $\exp(-\frac{2}{3}\zeta^{\frac{3}{2}})$ and $\zeta^{\frac{3}{2}}$ has a positive real part and a positive imaginary part. Hence this factor represents a wave whose amplitude decreases as $|\zeta|$ and $|z|$ increase, and whose phase is propagated upwards. Therefore Ai(ζ) is an acceptable solution. The asymptotic approximation for Bi(ζ) is given by (8.56) and includes a term with the factor $\exp(+\frac{2}{3}\zeta^{\frac{3}{2}})$. This represents a wave whose amplitude increases as $|\zeta|$ and $|z|$ increase, and whose phase is propagated downwards. It can be shown that the energy in this wave flows downwards, and it could not therefore be produced by a wave incident on the ionosphere from below. Hence the correct solution is proportional to Ai(ζ). Notice that when collisions are neglected, $\arg g = 0$, the asymptotic approximation for Ai(ζ) at great heights resembles an evanescent wave, § 2.14, since ζ is real and there is no variation of phase with height. When collisions are included there is some upward phase propagation since ζ is no longer real.

The horizontal component $\mathscr{H}_x$ of the magnetic field is found from the first equation (7.3), which comes from Maxwell's equations. Above the level $z = h_0$ the fields are therefore given by

$$E_y = K\,\mathrm{Ai}(\zeta), \quad \mathscr{H}_x = -\mathrm{i}\frac{gK}{k}\mathrm{Ai}'(\zeta) \tag{15.7}$$

where K is a constant. Below the level $z = h_0$ is free space so that $q = C$ is a constant. Then two independent solutions of (7.6) for E_y are $\exp\{\mp \mathrm{i}kC(z - h_0)\}$. The upper sign gives the obliquely upgoing wave, that is the incident wave, and the lower sign gives the downgoing, reflected wave. The solution of (7.3), (7.6) below the level $z = h_0$

is then

$$E_y = \exp\{-\mathrm{i}kC(z-h_0)\} + R\exp\{\mathrm{i}kC(z-h_0)\}$$
$$\mathscr{H}_x = -C\exp\{-\mathrm{i}kC(z-h_0)\} + RC\exp\{\mathrm{i}kC(z-h_0\} \tag{15.8}$$

where R is the reflection coefficient for $z = h_0$ as the reference level; § 11.2. Now the boundary conditions for the horizontal components E_y, $\mathscr{H}_x$ require that the values of each on the two sides of the boundary plane must be the same. Hence (15.7), (15.8) with $z = h_0$ give two equations from which K may be eliminated to give

$$R = \frac{Ck\,\mathrm{Ai}(\zeta_0) - \mathrm{i}g\,\mathrm{Ai}'(\zeta_0)}{Ck\,\mathrm{Ai}(\zeta_0) + \mathrm{i}g\,\mathrm{Ai}'(\zeta_0)} \tag{15.9}$$

where, from (15.4), (15.5):

$$\zeta_0 = -(Ck/g)^2. \tag{15.10}$$

If collisions are neglected, ζ_0 is real and the numerator and denominator of (15.9) are complex conjugates, so that $|R| = 1$. The reflection is therefore complete which was to be expected since all the energy in the incident wave should be reflected. The phase difference at $z = h_0$ between the incident and reflected waves is $\arg R$.

The functions Ai and Ai′ in (15.9) may be replaced by their asymptotic approximations (8.53), (8.55), for $\frac{2}{3}\pi < \arg\zeta_0 \leqslant \pi$ from (15.10), if $|\zeta_0|$ is large enough, that is if the gradient p in (15.1) is small and C is not too small. This is equivalent to the condition that the W.K.B. solutions of (7.6) must be good approximations just above the boundary $z = h_0$, whence (8.29) gives as the condition

$$|\zeta_0| \gg 1. \tag{15.11}$$

Then (15.9) gives

$$R \approx \mathrm{i}\exp\{-\tfrac{4}{3}\mathrm{i}(-\zeta_0)^{\frac{3}{2}}\} = \mathrm{i}\exp\{-\tfrac{4}{3}\mathrm{i}(Ck/g)^3\}. \tag{15.12}$$

An expression for R, based on the W.K.B. solutions, was given at (7.152). For a reference level at $z = h_0$, the lower limit of the integral in it must be changed from 0 to h_0. Then on using q from (15.1) in the integral, the modified (7.152) gives the same expression (15.12).

The agreement between (15.12) and the exact value (15.9) is good even when (15.11) is not satisfied. For example if collisions are neglected, ζ_0 is real and negative, and both (15.12) and (15.9) give $|R| = 1$. Then arg R is greater for (15.12) when $-\zeta_0 > 0.5$. The difference is less than 2° when $-\zeta_0 \geqslant 2$; it has a maximum value of about 11.4° when $-\zeta_0 = 1$, and it is 6° when $-\zeta_0 = 0.63$ and $-25°$ when $-\zeta_0 = 0.2$.

15.3. Reflection at a discontinuity of gradient

Electromagnetic waves are reflected at the bounding surface between two media with different refractive indices, and §§ 11.8–11.12 were devoted to this topic. There is then a discontinuity in the refractive index. It is now of interest to ask whether reflection can occur at a surface where the refractive index is continuous, but the gradient of the refractive index is discontinuous.

A problem where there was such a discontinuity of gradient was studied in § 15.2. Below the level $z = h_0$ is free space with $N(z) = 0$, $\mathrm{d}N/\mathrm{d}z = 0$, and above it N is proportional to $z - h_0$ so that $N(z)$ is continuous but there is a discontinuity of $\mathrm{d}N/\mathrm{d}z$. The expression (15.9) was deduced for the reflection coefficient of the whole ionosphere. In §§ 7.18, 7.19 it was suggested that reflection of radio waves occurs near the level $z = z_0$ where $q = 0$, and in earlier chapters this level has been called a 'level of reflection'. If there is also some reflection from the discontinuity at $z = h_0$ then (15.9) must be the result of combining these two reflections. Hence we must now ask whether (15.9) includes any contribution from the level $z = h_0$.

The field immediately above $z = h_0$ is given by (15.7) with $\zeta = \zeta_0$. If $|\zeta_0|$ is large enough the functions Ai, Ai$'$ may be replaced by their asymptotic approximations (8.53), (8.55). Each of these contains two terms of which one is a downgoing wave. Some or all of this is transmitted through the discontinuity and gives a downgoing wave below it, but this is not part of the reflection at $z = h_0$ and must be excluded. We therefore need to consider a different field configuration in which there is only an upgoing wave just above the level $z = h_0$.

First suppose that the gradients of $N(z)$, $n(z)$, $q(z)$ just above $z = h_0$ are so small that the W.K.B. solutions may be used. For an upgoing wave the W.K.B. solution satisfies (7.62) with the $-$ sign, so that the wave admittance (11.36) is

$$A_{22} = -\mathscr{H}_x/E_y = q. \tag{15.13}$$

Immediately above the boundary, $q = C$ so that $A_{22} = C$ is exactly the same as for an upgoing progressive wave in the free space below the boundary. For these two upgoing waves, above and below the boundary, the boundary conditions are satisfied and so there is no reflected wave. We conclude that if the W.K.B. solutions are good approximations on both sides of a discontinuity of gradient of $N(z)$, the reflection coefficient is zero to the same degree of approximation.

Next suppose that $\mathrm{d}N/\mathrm{d}z$ above the discontinuity is so large that the W.K.B. solutions cannot be used there. This means that there is no field configuration that can be regarded as a progressive wave. It is therefore not possible to define a purely upgoing wave in the region above $z = h_0$. Thus the problem of finding the reflection coefficient has no meaning. The reflection coefficient of the whole ionosphere is still given by (15.9) and this is influenced by the presence of the discontinuity of gradient. For example it would be changed if the discontinuity of gradient was 'smoothed out' by replacing the sharp angle in the $N(z)$ curve near $z = h_0$ by an arc extending over a small range of z. But still the reflection coefficient (15.9) cannot be separated into a part from the discontinuity and a part from the rest of the ionosphere.

The reflection process associated with the zero of q at $z = z_0$ does not occur at this one point only but must be thought of as occurring in a region of the complex z plane surrounding it and extending out to where the W.K.B. solutions can take over. If there is a discontinuity within this region the reflection coefficient is modified, but the effects associated with the discontinuity and with the zero of q cannot be separated.

An example is given in fig. 15.6 where there is a discontinuity of $\mathrm{d}N/\mathrm{d}z$ at the base of the ionosphere. It shows evidence that the reflected wave contains two components that can interact and give an interference pattern when the frequency is varied.

15.4. Piecewise linear models

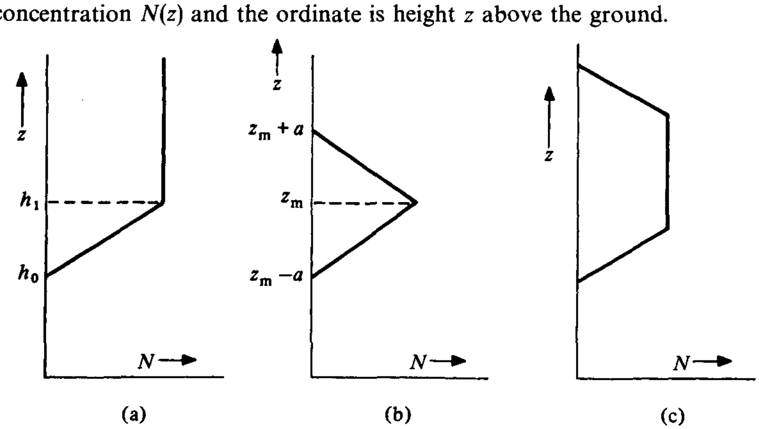
Two simple models of the ionosphere were of interest in the early days of the theory of radio reflections from the ionosphere, before computers were available. They are sketched in fig. 15.1(a, b). They were studied by Hartree (1929). The Airy integral functions were not then well known, and Hartree's solutions were given in terms of Bessel functions of order 1/3; see §8.7. The reflection coefficients for these two models are still of some historical interest and their derivation is useful as an illustration of methods. A third model, fig. 15.1(c), has been studied by B. Jeffreys (1956) for the application to wave mechanics, but it could equally well be used as a model of the ionosphere.

We consider only horizontally polarised waves, and electron collisions are neglected. For the model of fig. 15.1(a) the base of the ionosphere is at $z = h_0$. In the height range $h_0 \leqslant z \leqslant h_1$, X, proportional to $N(z)$, and q^2 are given by (15.1) with $Z = 0$. For $z \geqslant h_1$, X and $N(z)$ are constant. Thus

$$q = \begin{cases} C^2 & \text{for } z \leqslant h_0, \\ C^2 - p(z - h_0) & \text{for } h_0 \leqslant z \leqslant h_1, \\ C^2 - p(h_1 - h_0) = q_1^2, \text{say,} & \text{for } h_1 \leqslant z. \end{cases} \tag{15.14}$$

The constant q_1 is chosen to be either negative imaginary, or positive.

Fig. 15.1. Piecewise linear models of the ionosphere. The abscissa is the electron concentration $N(z)$ and the ordinate is height z above the ground.



If $h_1 - h_0$ is zero there is then a sharp boundary between two homogeneous media. The use of a non-zero $h_1 - h_0$ has the effect of removing the sharpness of the boundary, and thereby modifying the reflection coefficient as given by the Fresnel formula (11.57). This is one reason why the study of this model is of interest.

For $z > h_1$ there can only be an upgoing wave, or an evanescent wave whose amplitude decreases with increasing z, so that here, from (11.36)

$$A_{22} = q_1. \tag{15.15}$$

This is independent of z and is the value at the boundary $z = h_1$.

For $h_0 \leqslant z \leqslant h_1$ the function $q^2(z)$ is linear, and E_y satisfies the Stokes equation (8.7) with ζ given by (15.4) with $Z = 0$. Then at the boundary $z = h_1$

$$\zeta = \zeta_1 = (k^2 p)^{\frac{1}{3}}(h_1 - h_0 - C^2/p) \tag{15.16}$$

and at the boundary $z = h_0$

$$\zeta = \zeta_0 = -(k^2 p)^{\frac{1}{3}} C^2/p. \tag{15.17}$$

Between these boundaries E_y is a linear combination of any two independent solutions $\mathscr{S}_{\mathrm{A}}(\zeta)$, $\mathscr{S}_{\mathrm{B}}(\zeta)$ of (8.7) and $\mathscr{H}_x$ is found from (7.3). Hence

$$E_y = \mathscr{S}_{\mathrm{A}}(\zeta) + K\mathscr{S}_{\mathrm{B}}(\zeta), \quad \mathscr{H}_x = -\mathrm{i}w\{\mathscr{S}'_{\mathrm{A}}(\zeta) + K\mathscr{S}'_{\mathrm{B}}(\zeta)\}, \tag{15.18}$$

where K is a constant and $w = (p/k)^{\frac{1}{3}}$. The wave admittance (11.36) is then

$$A_{22} = \mathrm{i}w\{\mathscr{S}'_{\mathrm{A}}(\zeta) + K\mathscr{S}'_{\mathrm{B}}(\zeta)\}/\{\mathscr{S}_{\mathrm{A}}(\zeta) + K\mathscr{S}_{\mathrm{B}}(\zeta)\}. \tag{15.19}$$

For $z = h_1$, $\zeta = \zeta_1$, this is the same as (15.15) so that

$$K = -\{\mathrm{i}q_1\mathscr{S}_{\mathrm{A}}(\zeta_1) + w\mathscr{S}'_{\mathrm{A}}(\zeta_1)\}/\{\mathrm{i}q_1\mathscr{S}_{\mathrm{B}}(\zeta_1) + w\mathscr{S}'_{\mathrm{B}}(\zeta_1)\}. \tag{15.20}$$

For $z = h_0$, $\zeta = \zeta_0$, (15.19) is the same as A_{22} in the free space below the boundary, so that the reflection coefficient R_{22} with reference level $z = h_0$ can be found from the third equation (11.50). Let

$$\left.\begin{matrix}\Delta_+\\ \Delta_-\end{matrix}\right\} = \begin{vmatrix} \mathrm{i}q_1\mathscr{S}_{\mathrm{A}}(\zeta_1) + w\mathscr{S}'_{\mathrm{A}}(\zeta_1) & \mathrm{i}q_1\mathscr{S}_{\mathrm{B}}(\zeta_1) + w\mathscr{S}'_{\mathrm{B}}(\zeta_1) \\ \pm\mathrm{i}C\mathscr{S}_{\mathrm{A}}(\zeta_0) + w\mathscr{S}'_{\mathrm{A}}(\zeta_0) & \pm\mathrm{i}C\mathscr{S}_{\mathrm{B}}(\zeta_0) + w\mathscr{S}'_{\mathrm{B}}(\zeta_0) \end{vmatrix}. \tag{15.21}$$

Then it can be shown that

$$R_{22} = \Delta_+/\Delta_-. \tag{15.22}$$

The choice of the functions $\mathscr{S}_{\mathrm{A}}$ and $\mathscr{S}_{\mathrm{B}}$ is a matter of convenience in calculation. A fairly obvious choice would be $\mathscr{S}_{\mathrm{A}}(\zeta) = \mathrm{Ai}(\zeta)$, $\mathscr{S}_{\mathrm{B}}(\zeta) = \mathrm{Bi}(\zeta)$. Most large computers have library subroutines for calculating these functions and their derivatives.

Some special cases of (15.22) may be noted. Suppose that in the range $h_0 \leqslant z \leqslant h_1$ the W.K.B. approximations can be used in some part of the range.

(a) W.K.B. approximations good throughout the range

This can only happen if q is everywhere positive. Then the W.K.B. solutions are good on both sides of the discontinuities of gradient of q at $z = h_0$ and $z = h_1$. It was shown in § 15.3 that there is no appreciable reflection in these conditions. Hence the reflection coefficient of the ionosphere is negligibly small. It can be shown that if $\mathscr{S}_A$, $\mathscr{S}_B$ are the two W.K.B. solutions (8.27), (8.28), one column of Δ_+, (15.21) is a multiple of the other.

(b) W.K.B. approximations good at $z = h_1$ but q has a zero below this

In this case q_1 is negative imaginary so that the wave where $z > h_1$ is evanescent. Just below $z = h_1$ the W.K.B. solution is the corresponding evanescent approximation of the solution of the Stokes equation, that is the subdominant term. Hence the solution here is simply $\mathrm{Ai}(\zeta)$ which is the same as if the discontinuity at $z = h_1$ did not exist. The problem reduces to that solved in § 15.2, for $Z = 0$, and the reflection coefficient is (15.9). If the W.K.B. approximations are good also at $z = h_0$, the reflection coefficient can be further simplified and is given by (15.12).

(c) Sharp boundary limit

If $h_1 - h_0$ tends to zero it can be shown that (15.22) for R_{22} reduces to the Fresnel formula (11.57) for a sharp boundary.

Some further properties of the model (15.14), fig. 15.1(a), were given by Hartree (1929) and by Budden (1961a).

The model of fig. 15.1(b) is of interest because it simulates a rather crude form of symmetric ionospheric layer with free space below and above it. The plasma frequency at the maximum of this layer is the penetration frequency f_p. The theory is more complicated because there are now three discontinuities of dN/dz where the boundary conditions must be imposed. It has been given in full by Hartree (1929) in terms of Bessel functions of order 1/3, and by B. Jeffreys (1942, 1956) in terms of Airy integral functions. Jeffreys also considered a symmetrical trapezoidal form, fig. 15.1(c), for the function corresponding to $N(z)$. This is still more complicated because it has four discontinuities of dN/dz. Her papers were concerned with wave mechanics and potential barriers but the physics is the same as for the radio problem.

The formula for R_{22} for the $N(z)$ of fig. 15.1(b) is too complicated to be worth repeating here but it is given, in the same notation as used here, by Budden (1961a). For normal incidence and for frequencies sufficiently less than f_p the reflection is almost total so that $|R_{22}| \approx 1$. For frequencies sufficiently greater than f_p the wave penetrates the layer and $|R_{22}| \approx 0$. For frequencies in a range near f_p there is partial penetration and partial reflection. Curves showing how $|R_{22}|$ depends on frequency were given by Budden (1961a). They are similar in general form to those of fig. 15.5 which are for a more realistic model of the ionospheric layer. When the thickness of the layer, $2a$ in fig. 15.1(b), is small, the transition of $|R_{22}|$ from 1 to 0 occurs over a large range of f/f_p. For a thicker layer with large $2a$ the transition is more abrupt. This same behaviour can be seen in fig. 15.5.

15.5. Vertical polarisation at oblique incidence 1. Introductory theory

The term 'vertical polarisation' was explained in §7.2. It refers to obliquely incident waves in which the magnetic field has only a horizontal component. The wave normal is in the x–z plane, so that $\mathscr{H}_y$ is the only non-zero component of $\mathscr{H}$. The electric field $\boldsymbol{E}$ has both a vertical component E_z and a horizontal component E_x.

The differential equations for vertical polarisation were formulated in §7.12. One form is (7.68) in which

$$q^2 = n^2 - S^2 \tag{15.23}$$

where n is the refractive index. It shows how $\mathscr{H}_y$ depends on the height z, and it has a term in $\mathrm{d}\mathscr{H}_y/\mathrm{d}z$. It may be reduced to its 'normal form', that is a form without a first derivative term, by a change of dependent variable. Let

$$V = \mathscr{H}_y/n \tag{15.24}$$

so that (7.68) gives

$$\frac{\mathrm{d}^2 V}{\mathrm{d}z^2} + k^2(M^2 - S^2)V = 0 \tag{15.25}$$

where

$$M^2 = n^2 + \frac{1}{2k^2n^2}\frac{\mathrm{d}^2(n^2)}{\mathrm{d}z^2} - \frac{3}{4k^2n^4}\left\{\frac{\mathrm{d}(n^2)}{\mathrm{d}z}\right\}^2. \tag{15.26}$$

Equation (15.25) should be compared with (7.6) for horizontal polarisation, in which (15.23) appears instead of $M^2 - S^2$, and E_y instead of V. Thus the refractive index n in (7.6) is now replaced by M which is therefore called the 'effective' refractive index in (15.25). Similarly let

$$Q^2 = M^2 - S^2. \tag{15.27}$$

Then Q is called the effective value of q.

On entering the ionosphere with z increasing, suppose that $N(z)$ increases monotonically from zero, and neglect the effect of collisions. Then at first n and Q are positive, but they decrease until Q is zero where n^2 is still positive. Just above this level Q^2 is negative and the waves are evanescent. When z and $N(z)$ increase still further, a level is reached where $n^2 = 0$ and Q^2 and M^2 are infinite. This is the new feature of the equations for vertical polarisation. It is this infinity of the effective refractive index, sometimes called a resonance, that we mainly wish to study. It is on the real z axis if $Z = 0$, and near the real axis if Z is small.

By making a suitable choice of the function $N(z)$ it is possible to arrange that the differential equation (15.25) can be transformed to a standard form whose solutions can be expressed in terms of known functions. See for example Wait (1962, ch. VI), Burman and Gould (1964), Heading (1969), Westcott (1962b, c, 1969a, 1970). For some of these choices of $N(z)$ the refractive index n is never zero at any point on or near the real z axis, so the effect of the infinity of Q is not clearly evident. A possible method that does not have this disadvantage was suggested by van Duin and Sluijter (1980), who proposed using a transformation that converts (15.25) into Heun's equation; see § 15.12.

Now suppose that $N(z)$ is linear and Z is constant, as in (15.1), and let

$$\xi = z - h_0 - (1 - \mathrm{i}Z)/p \tag{15.28}$$

so that

$$n^2 = 1 - \frac{p(z-h_0)}{1-\mathrm{i}Z} = \frac{-p\xi}{1-\mathrm{i}Z}, \quad Q^2 = q^2 - \frac{3}{4k^2\xi^2}. \tag{15.29}$$

Then Q^2 is infinite where $\xi = 0$ and this is a regular singular point of the differential equation (15.25). It can in some cases be close to the point where $Q^2 = 0$, and then the function $\{Q(z)\}^2$ is not even approximately linear there. The presence of the singularity means that, at the reflection level, (15.25) does not approximate to the Stokes equation, so that the formula (7.152) for the reflection coefficient cannot be used. The reflection process is more complicated for vertical than for horizontal polarisation.

Equation (15.26) is not the only possible form for the square of the 'effective' refractive index. By using other dependent variables, for example nE_x/q, a different form is obtained. But for vertical polarisation it is never possible to find a dependent variable that makes $M = n$, except in the special case $S = 0$, that is for vertical incidence.

Now let

$$\zeta = \left(\frac{k^2 p}{1-\mathrm{i}Z}\right)^{\frac{1}{3}}\xi, \quad B = S^2\left\{\frac{(1-\mathrm{i}Z)k}{p}\right\}^{\frac{2}{3}}. \tag{15.30}$$

Then with ζ as independent variable, the differential equation (7.68) is

$$\frac{\mathrm{d}^2\mathscr{H}_y}{\mathrm{d}\zeta^2} - \frac{1}{\zeta}\frac{\mathrm{d}\mathscr{H}_y}{\mathrm{d}\zeta} - (\zeta + B)\mathscr{H}_y = 0. \tag{15.31}$$

This is not one of the standard differential equations of theoretical physics, but it must now be studied. In the special case of vertical incidence, $B = 0$, it becomes

$$\frac{\mathrm{d}^2\mathscr{H}_y}{\mathrm{d}\zeta^2} - \frac{1}{\zeta}\frac{\mathrm{d}\mathscr{H}_y}{\mathrm{d}\zeta} - \zeta\mathscr{H}_y = 0. \tag{15.32}$$

Now let $\mathscr{S}(\zeta)$ be any solution of the Stokes equation (8.7) so that $\mathscr{S}'' = \zeta\mathscr{S}$. Then it is easy to show that $\mathscr{H}_y = \mathscr{S}'(\zeta)$ is a solution of (15.32). For example one solution is $\mathrm{Ai}'(\zeta)$. Reflection at vertical incidence can equally well be studied for waves polarised with $\boldsymbol{E}$ parallel to the y axis. The solution is then as in § 15.2 where it was shown that the magnetic field is proportional to $\mathrm{Ai}'(\zeta)$, (15.7). Hence the solution $\mathrm{Ai}'(\zeta)$ of (15.32) satisfies the physical conditions for waves vertically incident from below.

It can be shown by the method of § 8.16 that the Stokes multiplier of (15.32) is $-\mathrm{i}$. For vertical polarisation the reflection coefficient of the ionosphere is R_{11} and (11.16) shows that it is the ratio of the magnetic fields of the downgoing and upgoing waves. Thus the formula for R_{11} analogous to (7.152) is

$$R_{11} = -\mathrm{i}\exp\left\{-2\mathrm{i}k\int_0^{z_0} n\,\mathrm{d}z\right\}. \tag{15.33}$$

If a reflection coefficient was defined as the ratio of the electric fields, a change of sign would be necessary; see end of § 11.8. Then (15.33) would give the same formula (7.152) as was derived for horizontal polarisation in §§ 7.19, 8.20.

15.6. Vertical polarisation 2. Fields near zero of refractive index

The study of the reflection of vertically polarised waves really requires a full treatment of the differential equation (15.31) similar to that given in ch. 8. for the Stokes equation. The following account is an outline only. It is based in part on a treatment given by Försterling and Wüster (1951). The equation (15.31) has also been studied by Denisov (1957), Hirsch and Shmoys (1965), and White and Chen (1974). The object of this section is to study the behaviour of the fields near the point $\zeta = 0$ where the refractive index n is zero. If $B \neq 0$, this is not the same as the 'reflection' point.

A solution of (15.31) can be found as a series in ascending powers of ζ, thus

$$v_1(\zeta) = \zeta^\beta(1 + a_1\zeta + a_2\zeta^2 + \cdots). \tag{15.34}$$

This is substituted in (15.31) and the coefficients of successive powers of ζ are equated to zero. This shows that if $B \neq 0$ the only possible value of β is 2, and the coefficients $a_1, a_2 \cdots$ can be found. Further $a_1 = 0$, so that

$$v_1(\zeta) = \zeta^2 + a_2\zeta^4 + a_3\zeta^5 + \cdots. \tag{15.35}$$

A second solution may be found (see, for example, Whittaker and Watson, 1927, § 10.32) as follows. Let

$$v_2(\zeta) = Kv_1(\zeta)\ln\zeta + 1 + b_1\zeta + b_2\zeta^2 + \cdots. \tag{15.36}$$

Substitute in (15.31) and equate coefficients of powers of ζ to zero. This gives

$$b_1 = 0, \quad K = \tfrac{1}{2}B, \quad b_3 = 1/3. \tag{15.37}$$

The value of b_2 is arbitrary and may be taken as zero. If it is not zero, the effect is simply to add a multiple of $v_1(\zeta)$ on to $v_2(\zeta)$. Then b_4, b_5 etc. may be found. For vertical incidence B is zero. Then K is zero and the logarithm term disappears from (15.36). The solutions $v_1(\zeta)$, $v_2(\zeta)$ are then simply the derivatives of the two series in (8.8).

The general solution of (15.31) is

$$\mathscr{H}_y = A_1 v_1(\zeta) + A_2 v_2(\zeta). \tag{15.38}$$

For vertical incidence the required solution is proportional to $\mathrm{Ai}'(\zeta)$ and (8.9) shows that

$$A_1 = \tfrac{1}{2}3^{-\frac{2}{3}}/(-\tfrac{1}{3})!, \quad A_2 = -3^{-\frac{1}{3}}/(-\tfrac{2}{3})!, \quad \text{for } B = 0. \tag{15.39}$$

Hence A_2 is not zero when $B = 0$ and it cannot drop discontinuously to zero when $B \neq 0$. The solution (15.38) for oblique incidence must contain a multiple of $v_2(\zeta)$ including the logarithm term in (15.36)

The field components E_x and E_z may be found by applying Maxwell's equations (7.4) to (15.38). This shows that E_x contains a term $A_2 B \ln\zeta$ and E_z contains a term A_2/ζ. These are both infinite when $\zeta = 0$, that is when $n = 0$. This is usually above the reflection level where $q = 0$.

In the ionosphere the electrons always make some collisions so that Z is never exactly zero. Consequently we cannot have $n = 0$ at any real height z. For example (15.29) shows that z is complex when $n^2 = 0$. But if Z is small the condition $n^2 = 0$ may hold at a point very close to the real axis in the complex z plane. At real heights near this point the terms $\ln\zeta$ and $1/\zeta$ in E_x, E_z respectively may be very large. This can happen only for vertical polarisation at oblique incidence. It does not occur for horizontal polarisation. Computed curves showing how $\mathscr{H}_y$, E_x and E_z depend on height near $\zeta = 0$ were given by Booker, Fejer and Lee (1968), for vertical polarisation and small Z.

The large field components E_x, E_z would mean that the simple linear form (3.12) of the equation of motion of the electrons cannot be used. A more complicated non-linear equation would be needed, leading to a modification of the foregoing treatment to allow for non-linear effects in strong fields. The following results might be expected.

The field component E_z imparts vertical motions to the electrons. When E_z is large these motions are large, and, within one cycle of oscillation of the field, any one electron moves to different levels where E_z is different. Thus one electron encounters very different fields E_z within one cycle. The vertical force to which it is subjected does not, therefore, vary harmonically with time and the electron motion is not simple harmonic. The motion could be Fourier analysed into a series of frequencies equal to the wave frequency and its harmonics. The theory of this process has been developed by Feinstein (1950) and by Försterling and Wüster (1951) who showed that harmonics of the wave frequency can be generated in this way near a level where $n = 0$. This process would be most marked when Z is very small, for then the point where $n=0$ is near the real axis in the complex z plane. Some energy must go into the harmonics generated, and this must come from the original wave which is therefore attenuated. The effect is thus similar to that of additional damping of the electrons' motions. Even if $Z=0$, some energy would be removed from the wave as harmonics, and this should lead to a reflection coefficient with modulus less than unity. This is discussed in the following section. When the earth's magnetic field is allowed for the theory must be modified, but there is still a point in the complex z plane where the governing differential equations have a singularity analogous to the one at $n=0$, in this section. This topic is discussed in § 19.5. Near this singularity the wave fields are very large and should still lead to the generation of harmonics.

The author does not know of any observations of harmonics of the wave frequency that could be attributed to a mechanism of this kind. This is in spite of the fact that in recent years many experiments have been done with radio waves of very large power, sufficient to cause heating and substantial artificial modification of the ionospheric plasma (see for example *Radio Science*, special issue Nov. 1974; Fejer, 1975, 1979).

15.7. Vertical polarisation 3. Reflection coefficient

In § 7.19 the phase integral form (7.152) was given for the reflection coefficient of the ionosphere. It was based on solutions of the differential equation (7.6) for horizontal polarisation, so that the reflection coefficient R in (7.152) is R_{22}, defined by (11.17). It was justified by an application of the principle of uniform approximation in Langer's (1937) form, given in § 8.20. It can be used if the function $q^2(z)$ in (7.6) has an isolated zero at the 'reflection level' where $q=0$. This is the same as the requirement that $q^2(z)$ shall be sufficiently near linear at this level. An approximate upper limit to the curvature was given by the criterion (8.75). The first factor i in (7.152) came from the second term of (8.73), which was derived from the asymptotic approximation for $\mathrm{Ai}(\zeta)$. Thus this factor is the Stokes multiplier for the Stokes equation (8.7).

We must now enquire whether the same method can be used for vertical polarisation to give R_{11} (11.16). The differential equation (15.25) is of the same form as (7.6). It contains the function $Q^2(z)$ (15.27), instead of $q^2(z)$. Clearly the method would be expected to work if $Q^2(z)$ satisfies the criterion (8.75) at the 'reflection' level where $Q=0$. If n^2 and Q^2 are given by (15.29) near the reflection level, it can be shown that this criterion leads to

$$S^3 \gg (\tfrac{3}{2})^{\frac{3}{4}} \left| \frac{p}{k(1-\mathrm{i}Z)} \right| \quad \text{that is } |B| \gg 1.36 \tag{15.40}$$

where B is given by (15.30). This shows that for oblique incidence with large enough

S the separation of the zero of Q and the singular point where $n = 0$, $Q \to \infty$, in the complex z plane, is great enough for the zero of Q to be treated as 'isolated'. The reflection coefficient may then be written

$$R_{11} = \lambda(B) \exp\left(-2ik \int_0^{z_0} Q \, dz\right) \tag{15.41}$$

where z_0 is the point where $Q = 0$. Here $\lambda(B)$ is the Stokes multiplier, and if (15.40) is satisfied it is approximately $+\mathrm{i}$. It was shown at (15.33) that, when S and B are zero, $\lambda(B) = -\mathrm{i}$ which is the Stokes multiplier for (15.32). For intermediate values of B, (15.41) might still be used provided that $\lambda(B)$ is given a suitable value.

The required $\lambda(B)$ is the Stokes multiplier for the differential equation (15.31). There is no closed expression for it, but it can be computed by the following method. The W.K.B. solutions of (15.31) can be found by the method of § 7.12. Equation (7.74) shows that they contain the factors $\exp\{\pm \int^{\zeta} (\zeta + B)^{\frac{1}{2}} \, d\zeta\}$. They are to be used where $\zeta \gg B$ so the square root is expanded and only positive powers of ζ need be retained. Hence the W.K.B. solutions are

$$\mathscr{H}_y \sim \zeta^{\frac{1}{4}} \exp\{\pm(\tfrac{2}{3}\zeta^{\frac{3}{2}} + B\zeta^{\frac{1}{2}})\}. \tag{15.42}$$

When $|\zeta|$ is very large the factor $\exp(\pm\frac{2}{3}\zeta^{\frac{3}{2}})$ predominates. It occurs also in the asymptotic approximations to the solutions of the Stokes equation. Hence the Stokes lines and anti-Stokes lines for (15.31) are the same as for the Stokes equation; see fig. 8.6.

At great heights $|\zeta|$ is large and (15.29), (15.30) show that $0 \leqslant \arg \zeta < \frac{1}{6}\pi$. Here the required solution contains only the subdominant term, representing either an upward travelling wave, or a field whose amplitude decreases as z increases. Well below the level of reflection $\frac{2}{3}\pi < \arg \zeta \leqslant \pi$. In going from great to small heights the Stokes line at $\arg \zeta = \frac{2}{3}\pi$ is crossed and there are two terms in the asymptotic approximation below the level of reflection. We now require a connection formula between the two asymptotic approximations. This is of the form (8.64) and may be written

$$\zeta^{\frac{1}{4}}\{\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}} - B\zeta^{\frac{1}{2}}) + \lambda(B)\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}} + B\zeta^{\frac{1}{2}})\} \leftrightarrow \zeta^{\frac{1}{4}} \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}} - B\zeta^{\frac{1}{2}}) \tag{15.43}$$

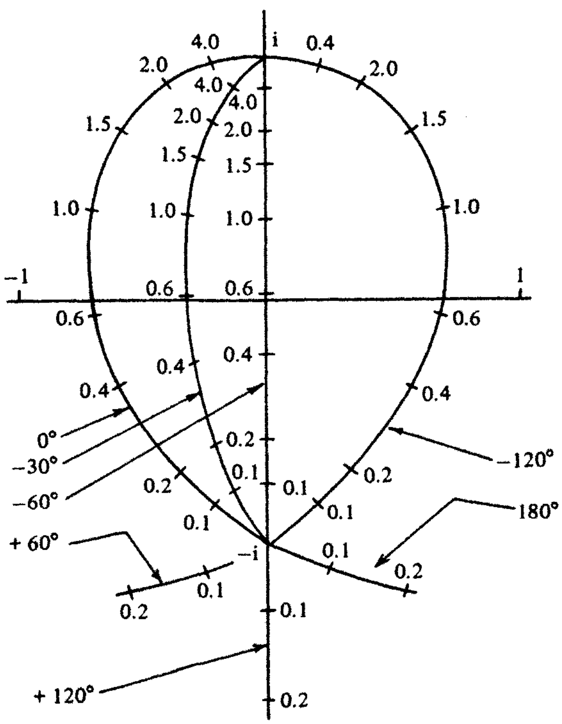
where $\lambda(B)$ is the Stokes multiplier for (15.31). It is equal to $-\mathrm{i}$ when $B = 0$. For other values of B it must be computed. This was done by a numerical integration of (15.31). For details of the computing methods see Budden (1961a, § 16.16 and Appendix).

Some results are given in fig. 15.2. They show how $\lambda(B)$ goes from $-\mathrm{i}$ to $+\mathrm{i}$ when $|B|$ goes from 0 to large values. If electron collisions are neglected, B is real and the integral in (15.41) is also real so that the exponential has modulus unity. But fig. 15.2 shows that $|\lambda(B)|$ is in general less than unity and hence $|R_{11}| < 1$. The smallest value is about 0.72 and occurs when $B \approx 0.45$. This means that not all the incident energy is reflected from the ionosphere, even when there is no physical mechanism for converting the energy of the waves into heat. This might be because some energy is converted into waves of the harmonic frequencies, as mentioned at the end of § 15.6. A more likely explanation, however, is that this problem cannot be treated as a 'steady state' problem. The incident wave must first be switched on, and after that the energy accumulates near the level where $n = 0$. If this energy cannot be absorbed, it continues to accumulate near the level where $n = 0$, and equilibrium is never attained. See § 19.5 for further discussion.

In practice there must be some collisions so that the plasma is most strongly heated where the electric field is very large, § 15.6, that is near where $n^2 = 0$. The possibility of heating a plasma in this way was considered by Freidberg, Mitchell, Morse and Rudsinski (1972).

In the lower ionosphere the dependence of X on height z is roughly exponential so that the gradient $p = \mathrm{d}X/\mathrm{d}z$ (15.1), near where $X = 1$, is not strongly dependent

Fig. 15.2. The complex plane showing values of the Stokes multiplier λ for the differential equation (7.68). Numbers by the curves show the values of $|B|$ marked with bars, and of arg B in degrees marked with arrows. Only values of arg B from $-60°$ to $0°$ are of interest when finding reflection coefficients, but some results for arg $B = \pm 120°$ are included for mathematical interest.



on frequency. If we take $p \approx 0.6\,\text{km}^{-1}$ as typical for all frequencies, this gives the following values of B, from (15.30) with $Z = 0$:

Frequency	16 kHz	2 MHz	6 MHz
Approx. value of B	$1.6\,S^2$	$40\,S^2$	$80\,S^2$

The value $B = 0.45$ therefore corresponds to an angle of incidence of about 30° at 16 kHz, or 6° at 2 MHz.

The above theory is modified when the earth's magnetic field is allowed for; see §§ 19.4–19.6.

15.8. Exponential electron height distribution

Radio waves of very low frequency, less than about 500 kHz, are reflected low down in the ionosphere usually between 70 and 100 km. Here the electron concentration $N(z)$ increases with height z. If the E-layer is a Chapman layer, § 1.5, then in its lower part $N(z)$ is roughly exponential. The actual $N(z)$ is more complicated than this, see fig. 1.3, but still an exponential $N(z)$ is very useful for studying some general features of the reflection of low frequency radio waves. It also has the advantage that there are no discontinuities of gradient $\mathrm{d}N/\mathrm{d}z$ that can affect the solution. For horizontal polarisation and constant collision frequency the reflection coefficient R_{22} can be expressed in closed form (15.59). For vertical polarisation there is no corresponding simple solution. The exponential distribution has been studied by Westcott (1962b, d), Wait (1962) and others.

If N increases exponentially with height z, then X, proportional to N, is given by

$$X = \mathrm{e}^{\alpha z}. \tag{15.44}$$

It is convenient to choose the origin of z where $X = 1$. It is important to notice that this origin depends on frequency f, since X is inversely proportional to f^2 (3.6). It is assumed that Z is a constant. From (4.9), (7.2)

$$q^2 = C^2 - \frac{\mathrm{e}^{\alpha z}}{1 - \mathrm{i}Z}. \tag{15.45}$$

Then the differential equation (7.6) for E_y is

$$\frac{\mathrm{d}^2 E_y}{\mathrm{d}z^2} + k^2\left(C^2 - \frac{\mathrm{e}^{\alpha z}}{1 - \mathrm{i}Z}\right)E_y = 0. \tag{15.46}$$

Now let

$$\zeta = \frac{2\mathrm{i}k}{\alpha}(1 - \mathrm{i}Z)^{-\frac{1}{2}}\exp\left(\tfrac{1}{2}\alpha z\right) \tag{15.47}$$

where the factor $(1 - \mathrm{i}Z)^{-\frac{1}{2}}$ is chosen so that its argument is positive and less than $\frac{1}{2}\pi$, and let

$$\nu = 2\mathrm{i}kC/\alpha. \tag{15.48}$$

Throughout this section ν has this meaning and must not be confused with the ν used elsewhere for the electron collision frequency. Then (15.46) becomes

$$\frac{d^2E_y}{d\zeta^2}+\frac{1}{\zeta}\frac{dE_y}{d\zeta}+\left(1-\frac{\nu^2}{\zeta^2}\right)E_y=0 \tag{15.49}$$

which is Bessel's equation of order ν. A solution must now be found that represents an upgoing wave only, at great heights. When z is very large (15.45) gives

$$q\approx -\mathrm{i}(1-\mathrm{i}Z)^{-\frac{1}{2}}\exp(\tfrac{1}{2}\alpha z) \tag{15.50}$$

where the minus sign has been chosen to ensure that q has a positive real part and a negative imaginary part. Then

$$k\int_0^z q\,\mathrm{d}z=-\zeta. \tag{15.51}$$

At a great height the upgoing W.K.B. solution of (15.46) contains the exponential of $-\mathrm{i}$ times (15.51), that is $\mathrm{e}^{+\mathrm{i}\zeta}$. Similarly the downgoing W.K.B. solution contains the factor $\mathrm{e}^{-\mathrm{i}\zeta}$ and this cannot appear in the required solution.

Two independent solutions of (15.49) are $H_\nu^{(1)}(\zeta)$ and $H_\nu^{(2)}(\zeta)$ where Watson's (1944) notation is used for Bessel functions of the third kind, sometimes called Hankel functions. The order ν of these functions here is purely imaginary (15.48). The appearance of a complex order is perhaps unusual in physical problems, but it does not affect the definition of the Hankel functions as given by Watson. Now (15.47) shows that when z is positive

$$\tfrac{1}{2}\pi\leqslant\arg\zeta<\pi \tag{15.52}$$

and for this range the asymptotic forms, based on formulae given by Watson (1944, p. 201) are

$$H_\nu^{(1)}(\zeta)\sim(2/\pi\zeta)^{\frac{1}{2}}\exp\{-\mathrm{i}(\tfrac{1}{2}\nu\pi+\tfrac{1}{4}\pi)\}\mathrm{e}^{\mathrm{i}\zeta}, \tag{15.53}$$

$$H_\nu^{(2)}(\zeta)\sim(2/\pi\zeta)^{\frac{1}{2}}\exp\{\mathrm{i}(\tfrac{1}{2}\nu\pi+\tfrac{1}{4}\pi)\}(\mathrm{e}^{-\mathrm{i}\zeta}+2\mathrm{i}\,\mathrm{e}^{\mathrm{i}\zeta}\cos\nu\pi). \tag{15.54}$$

These apply for

$$-\tfrac{1}{2}\pi\leqslant\arg\zeta\leqslant\tfrac{3}{2}\pi. \tag{15.55}$$

Bessel's equation (15.49) has anti-Stokes lines where $\arg\zeta=0,\ \pm\pi$, and Stokes lines where $\arg\zeta=\pm\frac{1}{2}\pi$. Watson gives asymptotic approximations in the Poincaré sense, § 8.11, so that his version of (15.55) is $-\pi<\arg\zeta<2\pi$. The range of $\arg\zeta$ that he uses therefore extends nearly to anti-Stokes lines. It was explained in § 8.17 how this may lead to errors when $|\zeta|$ is not indefinitely large. The range given in (15.55) ends on Stokes lines. For (15.53) only the subdominant term is present when $0<\arg\zeta<\pi$ and this includes the range (15.52). For (15.54) both dominant and subdominant terms are present in this range. The reader may find it instructive to draw the Stokes diagram for Bessel's equation, and to prove that the Stokes multiplier is $2\mathrm{i}\cos\nu\pi$. Since Bessel functions are not in general single valued, the form of diagram used in fig. 8.6 is not suitable. An alternative form suggested by

Heading (1957) may be used in which $\arg\zeta$ is plotted as abscissa and the circle of fig. 8.6 is replaced by a horizontal line.

Hence the required solution is $H_\nu^{(1)}(\zeta)$ (15.53). This function must now be examined at very low levels, where it is to be separated into upgoing and downgoing waves. Bessel's equation (15.49) may be solved by assuming that there are solutions of the form $\zeta^\beta(a_0 + a_1\zeta + \cdots)$. This is substituted in (15.49) and coefficients of successive powers of ζ are equated to zero. In this way two solutions are found that are independent when ν is not an integer. In the present problem ν is never an integer. The two solutions are

$$\begin{aligned} J_\nu(\zeta) &= (\tfrac{1}{2}\zeta)^\nu(\nu!)^{-1}(1 + a_2\zeta^2 + \cdots), \\ J_{-\nu}(\zeta) &= (\tfrac{1}{2}\zeta)^{-\nu}\{(-\nu)!\}^{-1}(1 + b_2\zeta^2 + \cdots). \end{aligned} \tag{15.56}$$

The coefficients a_2, b_2, etc. can be found but are of no interest in the present problem. The solution $H_\nu^{(1)}(\zeta)$ must be a linear combination of (15.56) and it is shown by Watson (1944, p. 74) that

$$H_\nu^{(1)}(\zeta) = \{J_{-\nu}(\zeta) - e^{-\nu\pi i}J_\nu(\zeta)\}/(i\sin\nu\pi). \tag{15.57}$$

This provides the connection formula in the present problem.

At the ground z is very large and negative, so that $|\zeta|$, from (15.47), is very small. Hence only the first terms of the series (15.56) need be retained, and (15.57) shows that the solution near the ground is proportional to

$$\frac{(\frac{1}{2}\zeta)^{-\nu}}{(-\nu)!} - e^{-\nu\pi i}\frac{(\frac{1}{2}\zeta)^\nu}{\nu!} \tag{15.58}$$

since the denominator of (15.57) is a non-zero constant, and can therefore be omitted. The factor $\zeta^{-\nu}$ in the first term is, from (15.47), (15.48), proportional to e^{-ikCz} which is the upgoing wave. Similarly the second term contains a factor e^{ikCz} which is the downgoing wave. The ratio of the two terms is therefore the reflection coefficient

$$R_{22} = -\left(\frac{k}{\alpha}\right)^{2\nu}(1 - iZ)^{-\nu}\frac{(-\nu)!}{\nu!}\exp(-2ikCh_1) \tag{15.59}$$

where h_1 is the height above the ground where $X = 1$, $z = 0$. The reference level for R_{22} is at the ground. Since ν is purely imaginary all terms in (15.59) except $(1 - iZ)^{-\nu}$ have modulus unity and hence

$$|R_{22}| = \exp\left\{-\frac{2kC}{\alpha}\arctan Z\right\}, \tag{15.60}$$

$$\arg R_{22} = \pi + \frac{4kC}{\alpha}\ln\left(\frac{k}{\alpha}\right) - \frac{kC}{\alpha}\ln(1 + Z^2) + 2\arg\{(-2ikC/\alpha)!\} - 2kCh_1. \tag{15.61}$$

Table 15.1

kC/α	Frequency for normal incidence when $\alpha = 0.6\,\mathrm{km}^{-1}$	Phase difference (15.64) (degrees)
0	0	90
0.05	1.4 kHz	43
0.1	2.86 kHz	26
1.0	28.6 kHz	10
4.0	1.14 MHz	2.3

When electron collisions are neglected, $Z = 0$, (15.60) shows that $|R_{22}| = 1$. This was to be expected since no energy can be absorbed from the wave.

The approximate phase integral formula (7.152) may be used to find R_{22} for the exponential ionosphere (15.44). Let z_0 be the value of z where $q = 0$ in (15.45). Then the phase integral $\Phi = 2k\int_{-h_1}^{z_0} q\,\mathrm{d}z$ may be evaluated. The resulting expression contains $\exp(-\alpha h_1)$ but this is extremely small and may be neglected, so that

$$\Phi = 2kCh_1 + \frac{4kC}{\alpha}\{\ln(2C) - 1\} + \frac{kC}{\alpha}\ln(1 + Z^2) - \mathrm{i}\frac{2kC}{\alpha}\arctan Z. \quad (15.62)$$

Then (7.152) gives $R_{22} = \mathrm{i}e^{-\mathrm{i}\Phi}$ whence $|R_{22}|$ is the same as (15.60) and

$$\arg R_{22} = \tfrac{1}{2}\pi - \frac{4kC}{\alpha}\{\ln(2C) - 1\} - \frac{kC}{\alpha}\ln(1 + Z^2) - 2kCh_1. \quad (15.63)$$

The phase integral formula thus gives the correct $|R_{22}|$ for all values of α. The difference between the values (15.61), (15.63) for $\arg R_{22}$ is

$$\tfrac{1}{2}\pi + \frac{4kC}{\alpha}\left\{\ln\left(\frac{2kC}{\alpha}\right) - 1\right\} + 2\arg\left\{\left(\frac{-2\mathrm{i}kC}{\alpha}\right)!\right\}. \quad (15.64)$$

For a 'slowly varying' ionosphere α is small and $2kC/\alpha$ is large. The first term of Stirling's formula may then be used for the factorial function, and this makes (15.64) zero. Some values of (15.64) for larger α are given in table 15.1. The value $\alpha = 0.6\,\mathrm{km}^{-1}$ used in the second column is roughly typical for the lower ionosphere. Hence in this example the phase integral formula (7.152) is remarkably accurate, since it always gives the correct value of $|R_{22}|$ and the error in phase never exceeds 90°.

15.9 Parabolic electron height distribution 1. Phase integrals

The parabolic model of the ionosphere (12.17) was discussed in §§ 12.4, 12.7 – 12.10 by the methods of ray theory. This model is important because, near the maximum of any ionospheric layer, $N(z)$ can be approximated by the square law function (12.17). It has the disadvantage that at the upper and lower boundaries, $z = z_\mathrm{m} \pm a$ in (12.17), there is a discontinuity of gradient that does not occur in the actual ionosphere. The parabolic layer will now be studied with the main object of seeing what happens where ray theory fails. It is assumed that the electron collision frequency is independent of height so that Z is a constant. Only horizontal polarisation is discussed here, for then the differential equation (7.6) can be converted into a standard form. For vertical polarisation it is not possible to reduce the differential equation (7.68) to a simple standard form.

Consider a parabolic ionospheric layer with $N(z)$ given by (12.17). Let f_{p} be the penetration frequency for vertical incidence. Then with (12.18)

$$X = (1-\zeta^2) f_{\mathrm{p}}^2/f^2, \quad q^2 = C^2 - \left(\frac{1-\zeta^2}{1-\mathrm{i}Z}\right)\frac{f_{\mathrm{p}}^2}{f^2} \tag{15.65}$$

where $C = \cos\theta$, and θ is the angle of incidence. The differential equation (7.6) to be studied is then

$$\frac{\mathrm{d}^2 E_y}{\mathrm{d}z^2} + k^2\left\{C^2 - \left(\frac{1-\zeta^2}{1-\mathrm{i}Z}\right)\frac{f_{\mathrm{p}}^2}{f^2}\right\}E_y = 0. \tag{15.66}$$

In the rest of this section the subscript y will be omitted. Now let

$$\xi = \left(\frac{kf_{\mathrm{p}}}{af}\right)^{\frac{1}{2}}(1-\mathrm{i}Z)^{-\frac{1}{4}}(z - z_{\mathrm{m}}), \quad A^2 = ka\left\{\frac{f_{\mathrm{p}}}{f}(1-\mathrm{i}Z)^{-\frac{1}{2}} - C^2\frac{f}{f_{\mathrm{p}}}(1-\mathrm{i}Z)^{\frac{1}{2}}\right\}, \tag{15.67}$$

where the fractional powers of $1-\mathrm{i}Z$ are chosen so that they tend to $+1$ with argument zero, when $Z \to 0$. Then (15.66) is

$$\frac{\mathrm{d}^2 E}{\mathrm{d}\xi^2} + (\xi^2 - A^2)E = 0. \tag{15.68}$$

If the functions X and q^2 in (15.65) are not exactly quadratic functions of ζ, the factor $1-\zeta^2$ then contains added small correction terms in higher powers of ζ. It is then possible to use a 'uniform approximation' analogous to (8.72), (8.68) so that the differential equation is still converted to the form (15.68). Further details are given in §17.3; see (17.12), (17.13).

Alternative forms of (15.68) are often used. Thus let

$$\xi = 2^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{4}\mathrm{i}\pi}u, \quad A^2 = -2\mathrm{i}(n+\tfrac{1}{2}). \tag{15.69}$$

Then (15.68) becomes

$$\frac{\mathrm{d}^2 E}{\mathrm{d}u^2} + (n + \tfrac{1}{2} - \tfrac{1}{4}u^2)E = 0 \tag{15.70}$$

which is known as Weber's equation and is the form studied by Whittaker and Watson (1927). It was used for electromagnetic waves and a parabolic ionised layer by Rydbeck (1942a, 1943), Northover (1962), Heading (1962a). Its solutions are called parabolic cylinder functions. One of them, $D_n(u)$, corresponds to the solution $\mathrm{F}(x; A)$ of (15.68) used below, § 15.10. Other forms of the differential equation are given by Miller (1955) and by Abramowitz and Stegun (1965). The form (15.68) was used by Heading (1962a, § A.4). It has the advantage that the independent variable is simply a scaled value of the height $z - z_m$ and if $Z = 0$ the scaling factor is real and positive. The two values $\xi = \pm A$ are where $q = 0$ and are thus the turning points of (15.68) or 'reflection points'.

Before studying (15.68) in detail it is useful to summarise some of the results of ray theory. First let $Z = 0$. The penetration frequency for obliquely incident waves is f_p/C. If $f < f_p/C$ the turning points are on the real z axis, at P and P′ in fig. 15.3(a), which is a sketch of the complex z plane. If the lower turning point P at $z = z_0 < z_m$ is sufficiently isolated, the reflection coefficient R, for waves incident from below, can be found from the phase integral formula (7.152). Let R have reference level at the lower boundary $z = z_m - a$. Then

$$R = \mathrm{i}\exp\left(-2\mathrm{i}k\int_{z_m - a}^{z_0} q\,\mathrm{d}z\right), \quad |R| = 1,$$

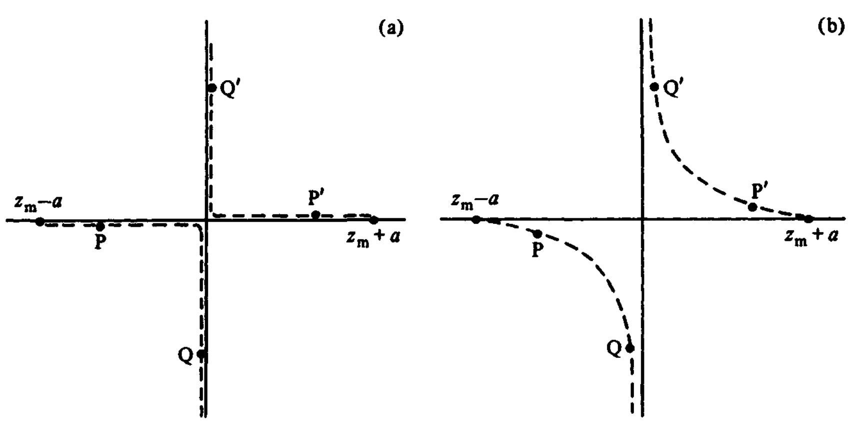
$$\arg R = \tfrac{1}{2}\pi - kaC\left\{1 - \frac{1}{2}\left(\frac{f_p}{Cf} - \frac{Cf}{f_p}\right)\ln\left(\frac{f_p + Cf}{f_p - Cf}\right)\right\}. \tag{15.71}$$

The integral here is of the same form as was used for (12.23). A criterion for P to be sufficiently isolated was given at (8.75), which can be shown to lead to

$$|A| \gg 0.84. \tag{15.72}$$

Now (15.67) shows that A is proportional to PP′, so that the criterion means that P and P′ must not be too close together, that is, f must not be too near the penetration frequency f_p/C. It also shows that A^2 is proportional to ka so the criterion is more easily satisfied for large ka, that is for thick layers. The phase integral formula (15.71) takes no account of the discontinuity of $\mathrm{d}N/\mathrm{d}z$ at the bottom of the layer. If the

Fig. 15.3. The complex z (height) plane showing the turning points PP′ or QQ′ of the differential equation (15.66). For (a) electron collisions are neglected, and for (b) they are included. PP′ are for a frequency less than the penetration frequency, and QQ′ are for a frequency greater than penetration.



reflection here is small, it is necessary that the W.K.B. solution should be a good approximation just above the lower boundary (see § 15.3). A condition for this is that (7.59) should be satisfied where $\zeta = -1$, and this gives

$$kaC^3 f^2/f_p^2 \gg 1. \tag{15.73}$$

This is the criterion that the lower turning point P, that is the 'reflection' level, shall not be too close to the lower boundary. This too is most easily satisfied for large ka.

The range of z between P and P′ in fig. 15.3(a) is where q is purely imaginary, and the waves are evanescent. This is an exact analogy of a potential barrier in wave mechanics. Within the barrier the 'upgoing' wave decays as z increases, but if the barrier is not too wide the wave energy that reaches P′ may not be negligible. The waves can tunnel through the barrier and this is an example of the simplest form of tunnelling in which q^2 is zero at both ends of the barrier. These physical considerations suggest that if T is the transmission coefficient of the ionosphere then

$$|T| = \exp\left(-k\int_{(\mathrm{P})}^{(\mathrm{P}')} |q|\,\mathrm{d}z\right) = \exp\left\{-\frac{1}{2}\pi ak\left(\frac{f_p}{f} - \frac{C^2 f}{f_p}\right)\right\}. \tag{15.74}$$

Now suppose that $Z \neq 0$. The values z_0 of z at the turning points P, P′ are found from (15.65) and are now complex, as shown in fig. 15.3(b). But the phase integral formula (7.152) can still be used, as suggested in § 7.19. The integral is now a contour integral with a complex upper limit z_0. Equation (15.65) shows that in the first equation (15.71), in q, f_p is to be replaced by $f_p(1 - \mathrm{i}Z)^{-\frac{1}{2}} = g$ say. Then

$$R = \mathrm{i}\exp\left[-\mathrm{i}kaC\left\{1 - \frac{1}{2}\left(\frac{g}{Cf} - \frac{Cf}{g}\right)\ln\left(\frac{g + Cf}{g - Cf}\right)\right\}\right]. \tag{15.75}$$

If $f \ll f_p/C$, the point P is near the real z axis. When f increases, P and P′ move to the left and right respectively along the dotted curves in fig. 15.3(b). Now PP′ is never zero but it may be small when f is near to f_p/C, and then (15.75) may be unreliable. But for still larger f, P and P′ move apart again to positions such as Q and Q′ in fig. 15.3(b). The turning point now at Q is again 'isolated' and so the approximation (15.75) can be good even when $f > f_p/C$. If now $Z = 0$ and $f > f_p/C$ the point P in fig. 15.3(a) has moved to the position Q. The formula (15.75) then gives

$$|R| = \exp\left\{-\frac{1}{2}\pi kzC\left(\frac{Cf}{f_p} - \frac{f_p}{Cf}\right)\right\}. \tag{15.76}$$

When $Z \neq 0$, q is never zero when z is real and for the upgoing wave Re(q) is positive, though it may be small within a barrier. The transmission coefficient is given by

$$T = \exp\left(-\mathrm{i}k\int_{z_m - a}^{z_m + a} q\,\mathrm{d}z\right) = \exp\left[-\mathrm{i}kaC\left\{1 - \frac{1}{2}\left(\frac{g}{Cf} - \frac{Cf}{g}\right)\ln\left(\frac{Cf + g}{Cf - g}\right)\right\}\right]. \tag{15.77}$$

Note the similarity of this and (15.75).

The approximate formulae (15.71)–(15.77) are based on the phase integral method, which is a generalisation of ray theory. They are compared below with results from full wave theory.

15.10. Parabolic electron height distribution 2. Full wave solutions

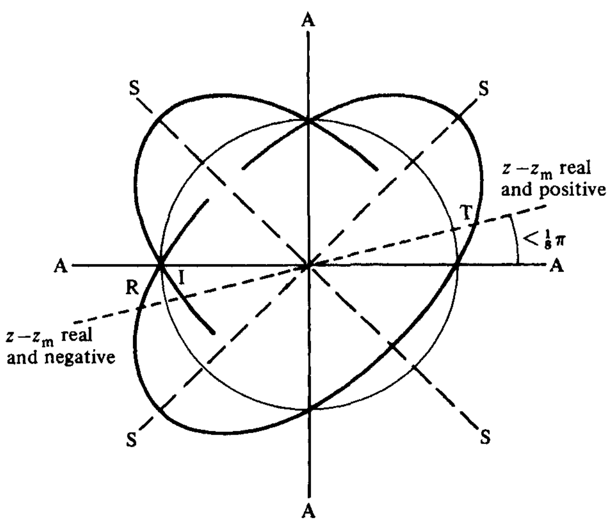
The differential equation (15.68) must now be studied. It has no singularities for bounded ξ. In this respect it resembles the Stokes differential equation (8.7). The only singularity is at infinity. In other words every point in the complex ξ plane, including the turning points $\xi = \pm A$, is an ordinary point. It means that two independent series solutions in ascending integer powers of ξ can be found and they are convergent for all ξ. Hence any solution of (15.68) is bounded and single valued for all ξ. It has no singularities except at infinity. Such a function is called an 'integral function', or 'entire function'. The form of (15.68) shows that for one of the series the powers must be all even, and for the other all odd. It proves more convenient to find series for the two functions $E\exp(\pm\frac{1}{2}\mathrm{i}\xi^2)$. They are not needed here but are given by Heading (1962a, § A.4).

The asymptotic approximations for the solutions of (15.68) may be found by the W.K.B. formula (7.26) which gives

$$E \sim (\xi^2 - A^2)^{-\frac{1}{4}}\exp\left\{\pm\mathrm{i}\int^{\xi}(\xi^2 - A^2)^{\frac{1}{2}}\,\mathrm{d}\xi\right\}. \tag{15.78}$$

This is to be used for large ξ, and in particular for $|\xi| > |A|$, to give an asymptotic

Fig. 15.4. Stokes diagram for the function $\mathrm{F}(\xi; A)$. S denotes Stokes lines and A anti-Stokes lines. The branch I represents the incident wave, R the reflected wave and T the transmitted wave.



form multiplied by an asymptotic series in descending powers of ξ (see § 8.11). We first find the asymptotic form. The series is given later. The fractional powers in (15.78) are expanded by the binomial theorem. For the first factor only one term need be retained since later terms affect only the second and later terms of the asymptotic expansion. For the integrand, however, two terms must be retained since the second gives a logarithm on integration. The result is

$$E \sim \xi^{-\frac{1}{2}}\xi^{\frac{1}{2}\mathrm{i}A^2}\mathrm{e}^{-\frac{1}{2}\mathrm{i}\xi^2} \text{ or } \xi^{-\frac{1}{2}}\xi^{-\frac{1}{2}\mathrm{i}A^2}\mathrm{e}^{\frac{1}{2}\mathrm{i}\xi^2}. \tag{15.79}$$

When $|\xi|$ is very large, the behaviour of these functions is determined mainly by the last exponential. The exponent $\frac{1}{2}\mathrm{i}\xi^2$ is real when $\arg\xi = \pm\frac{1}{4}\pi, \pm\frac{3}{4}\pi$ and these are therefore the Stokes lines (see §8.13) for (15.68). Similarly $\frac{1}{2}\mathrm{i}\xi^2$ is purely imaginary when $\arg\xi = 0, \pm\frac{1}{2}\pi, \pi$ and these are the anti-Stokes lines; see fig. 15.4.

The fractional powers of ξ in (15.79) may be written $\exp(\pm\frac{1}{2}\mathrm{i}A^2\ln\xi)$. In general these functions have infinitely many values. We choose the values for which $-\frac{5}{4}\pi \leqslant \arg\xi < \frac{3}{4}\pi$. This is equivalent to the use of a branch cut in the complex ξ plane where $\arg\xi = \frac{3}{4}\pi$. Let the asymptotic series that must multiply the first of the functions (15.79) be

$$S(\xi) = 1 + C_1\xi^{-2} + C_2\xi^{-4} + \cdots. \tag{15.80}$$

The differential equation (15.68) shows that the powers must all be even. Substitute (15.80) in (15.68) and equate coefficients of successively decreasing powers of ξ. This gives

$$C_1 = \frac{\mathrm{i}}{16}(1-\mathrm{i}A^2)(3-\mathrm{i}A^2), \quad C_r = \frac{\mathrm{i}^r}{16^r r!}(1-\mathrm{i}A^2)\cdots(4r-1-\mathrm{i}A^2). \tag{15.81}$$

The series for the second function (15.79) is obtained from (15.81) by reversing the sign of i.

Suppose now that conditions are such that there is no appreciable reflection at the discontinuity of $\mathrm{d}N/\mathrm{d}z$ at the upper boundary. Equation (15.67) shows that when z–z_m is large and positive, $|\xi|$ is large and $0 \leqslant \arg\xi < \frac{1}{8}\pi$. Thus $\mathrm{Re}(\frac{1}{2}\xi^2)$ increases as z increases. This shows that the term with the factor $\exp(-\frac{1}{2}\mathrm{i}\xi^2)$ in (15.79) is the one that represents an upgoing wave. For radio waves incident from below, we seek a solution in which only this term is present when z is large and positive. This solution will be denoted by $\mathrm{F}(\xi; A)$. Fig. 15.4 shows the Stokes diagram for this function. It can be shown that

$$\mathrm{F}(\xi; A) = 2^{\frac{1}{4}(1-\mathrm{i}A^2)}\exp\{\tfrac{1}{8}\pi(\mathrm{i}+A^2)\}D_n(u) \tag{15.82}$$

where u and n are given by (15.69).

For large real z the term with the factor $\exp(-\frac{1}{2}\mathrm{i}\xi^2)$ is the dominant term if $Z \neq 0$, and $\mathrm{Re}(-\frac{1}{2}\mathrm{i}\xi^2)$ is then positive, so that the exponential increases with increasing z. This suggests that the wave increases in amplitude as it travels, which would be physically unacceptable. But the solution is to be used only within the ionosphere

$z \leqslant z_m + a$, and here the other factors in (15.79) must ensure that the amplitude does not increase with increasing z.

When $z - z_m$ is real, large and negative, $\mathrm{Re}(\frac{1}{2}\xi^2)$ decreases when z increases. Hence the term with the factor $\exp(-\frac{1}{2}\mathrm{i}\xi^2)$ is now the downgoing wave, that is the reflected wave. But here the second term of (15.79) is also present as shown by the Stokes diagram fig. 15.4, and this is the incident wave. The required asymptotic approximations are

$$\mathrm{F}(\xi;A) \sim \begin{cases} \xi^{-\frac{1}{2}}\xi^{\frac{1}{2}\mathrm{i}A^2}\mathrm{e}^{-\frac{1}{2}\mathrm{i}\xi^2} \quad \text{for} \quad -\dfrac{3\pi}{4} \leqslant \arg\xi \leqslant \dfrac{\pi}{4}, & (15.83) \\ \xi^{-\frac{1}{2}}\xi^{\frac{1}{2}\mathrm{i}A^2}\mathrm{e}^{-\frac{1}{2}\mathrm{i}\xi^2} - \dfrac{\mathrm{i}(2\pi)^{\frac{1}{2}}}{\Gamma(\frac{1}{2}-\frac{1}{2}\mathrm{i}A^2)}\mathrm{e}^{\frac{1}{2}\pi A^2}2^{-\frac{1}{2}\mathrm{i}A^2}\xi^{-\frac{1}{2}}\xi^{-\frac{1}{2}\mathrm{i}A^2}\mathrm{e}^{\frac{1}{2}\mathrm{i}\xi^2} & \end{cases}$$

$$\text{for} \quad -\frac{5\pi}{4} \leqslant \arg\xi \leqslant -\frac{3\pi}{4}. \tag{15.84}$$

The last term can be derived from a knowledge of the Stokes multiplier for the Stokes line at $\arg\xi = -\frac{3}{4}\pi$. The Stokes multipliers for all four Stokes lines are given by Heading (1962a). Alternatively it can be found from (15.82) by using the asymptotic forms for $D_n(u)$ given by Whittaker and Watson (1935, p. 347).

In (15.83), (15.84) the ranges given for $\arg\xi$ end at Stokes lines; see § 8.17. In (15.83) the range could be given as $-\pi < \arg\xi < \pi/2$ ending just within anti-Stokes lines and this would be correct in the Poincaré sense. But then this range would include the value of $\arg\xi$ where $z - z_m$ is real and negative and it might be inferred that only the one term (15.83) is needed at both the top and bottom of the ionosphere. The use of the ranges given in (15.83), (15.84) ensures that at the bottom of the ionosphere the second term of (15.84) must be present.

When $z = z_m \pm a$ let $\xi = \xi_+$ and ξ_-, respectively, so that

$$\xi_- = \mathrm{e}^{-\mathrm{i}\pi}\xi_+. \tag{15.85}$$

In (15.84) the first term is the downgoing wave and the second is the upgoing wave, and their ratio is the reflection coefficient

$$R_{22} = \mathrm{i}(2\pi)^{-\frac{1}{2}}\Gamma(\tfrac{1}{2} - \tfrac{1}{2}\mathrm{i}A^2)\mathrm{e}^{\frac{1}{2}\pi A^2}2^{\frac{1}{2}\mathrm{i}A^2}\xi_+^{\mathrm{i}A^2}\mathrm{e}^{-\mathrm{i}\xi^2} \tag{15.86}$$

with reference level at the base of the ionosphere. Similarly (15.83) is the upgoing wave at the top of the ionosphere and its ratio to the second term of (15.84) is the transmission coefficient

$$T_{22} = R_{22}\exp\{-\tfrac{1}{2}\pi(A^2 + \mathrm{i})\} \tag{15.87}$$

where (15.85) has been used. For (15.86), (15.87) to be good approximations it is necessary that the condition (15.73) shall be satisfied. This means that the thickness of the parabolic layer must be several or many free space wavelengths.

When (15.73) is not adequately satisfied, it cannot be assumed that the solution of (15.68) just below the upper boundary is $F(\xi; A)$ alone, and it would be necessary to add to it a multiple of a second solution. This subject was studied by Rydbeck (1943). The reflection coefficient would then be expressed in terms of these two functions and their derivatives at $\xi = \pm A$. It would be much more complicated than (15.86). For this general case it proves easier to find R_{22} and T_{22} by a numerical integration of (15.68), for example by the method of §§ 18.2, 18.6, 18.7. This was used for the results of figs. 15.5, 15.6.

If collisions are neglected, (15.67) shows that A^2 is real and ξ_+ is real and positive. Then the modulus of the gamma function in (15.86) can be found from the 'reflection formula' for gamma functions: $\Gamma(x)\Gamma(1-x) = \pi \operatorname{cosec}(\pi x)$ (Abramowitz and Stegun, 1965, 6.1.17), whence

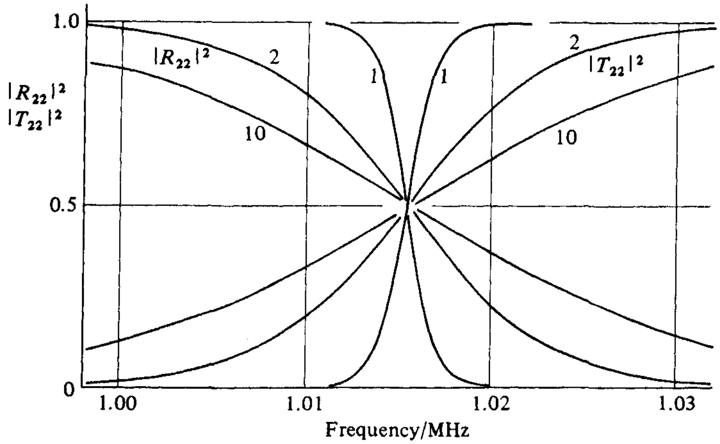
$$|\Gamma(\tfrac{1}{2} - \tfrac{1}{2}\mathrm{i}A^2)|^2 = \Gamma(\tfrac{1}{2} - \tfrac{1}{2}\mathrm{i}A^2)\Gamma(\tfrac{1}{2} + \tfrac{1}{2}\mathrm{i}A^2) = \pi \operatorname{sech}(\tfrac{1}{2}\pi A^2) \tag{15.88}$$

and (15.86), (15.87) give

$$|R_{22}|^2 = (1 + \mathrm{e}^{-\pi A^2})^{-1}, \quad |T_{22}|^2 = (1 + \mathrm{e}^{\pi A^2})^{-1}, \quad |R_{22}|^2 + |T_{22}|^2 = 1. \tag{15.89}$$

This shows that no energy is lost from the waves. The ray theory formula (15.71) suggests that if $f < f_p/C$, that is $A^2 > 0$, the waves cannot penetrate the layer and $|R_{22}| = 1$. But (15.89) shows that $|T_{22}|$ is not zero and $|R_{22}| < 1$. Similarly if $f > f_p/C$, the ray theory result (15.77) with $Z = 0$ suggests that $|T_{22}| = 1$, but (15.89) shows that $|T_{22}| < 1$. For $f = f_p/C$ (15.89) shows that $|R_{22}|^2 = |T_{22}|^2 = \tfrac{1}{2}$. For a range of

Fig. 15.5. Dependence of $|R_{22}|^2$ and $|T_{22}|^2$ on frequency for a parabolic layer when collisions and the earth's magnetic field are neglected. In all these examples the vertical incidence penetration frequency f_p is 1 MHz, and the angle of incidence θ is 10°. The numbers by the curves are the half thickness a in km. The ordinate scale is linear.



frequencies on either side of the penetration frequency f_p/C there is partial penetration and partial reflection. This range gets large when ka is made smaller. Fig. 15.5 shows how $|R_{22}|^2$ and $|T_{22}|^2$ depend on frequency for various values of the semi-thickness a, when collisions are neglected.

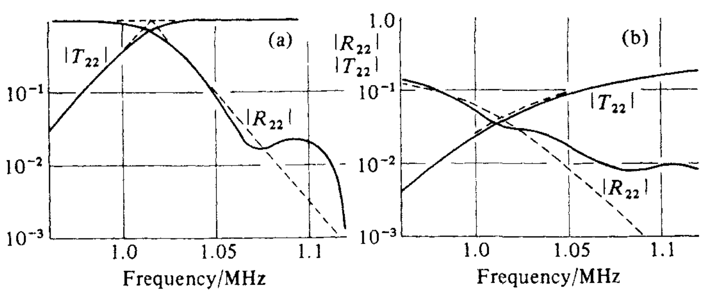
Fig. 15.6 shows further examples of the frequency dependence of $|R_{22}|$ and $|T_{22}|$. The continuous curves in both figs. 15.5, 15.6 were calculated by numerical integration of the differential equations with the method of §§ 18.2, 18.6, 18.7. The broken curves were calculated from the phase integral formulae (15.75), (15.77). In fig. 15.6(a) collisions are neglected. The values for $|R_{22}|$ from (15.89) are indistinguishable from the continuous curve where $f < 1.03$ MHz and indistinguishable from the broken curve when $f > 1.04$ MHz. The values for $|T_{22}|$ are indistinguishable from the continuous curve for the whole frequency range. In both figs. 15.6(a, b) the curves for $|R_{22}|$ show oscillations resembling an interference pattern, and suggesting that there are two contributions to the reflected wave, one from the discontinuity of dN/dz at the bottom of the layer and the other associated with the zero of q at the lower turning point.

15.11. Parabolic electron height distribution 3. Equivalent height of reflection

In § 12.4 ray theory was used to find the equivalent height of reflection $h'(f)$ (12.22), for radio waves vertically incident on a parabolic layer from below. This theory fails for frequencies near the penetration frequency and the full wave theory of § 15.10 will therefore now be used. Only vertical incidence, $C = 1$, is discussed, and electron collisions are neglected, so that $Z = 0$. The effect of discontinuities of dN/dz at the boundaries of the layer is ignored so that the reflection coefficient is given by (15.86) with $C = 1$, $Z = 0$. The theory taking account of collisions was given by Rydbeck (1942, 1943). Some results with collisions were given by Budden and Cooper (1962).

For a wave that travels over a given path the total change of phase is $2kP$ where P

Fig. 15.6. Two examples of the dependence of $|R_{22}|$ and $|T_{22}|$ on frequency f. The half thickness $a = 1$ km and θ, f_p are the same as in fig. 15.5. The continuous curves were calculated by numerical integration of the differential equations, and the broken curves from the phase integral formulae (15.75), (15.77). The ordinate scales are logarithmic. In (a) electron collisions are neglected, and in (b) $\nu = 5 \times 10^{-5}\,\mathrm{s}^{-1}$.



is the phase path, defined in § 10.16. It is related to the group path P' by (10.40). For vertical incidence $P' = 2h'(f)$ and $P = 2h(f)$ where $h(f)$ is called the phase height of reflection. In general h and h' may be complex. The amplitude of the wave is changed by a factor $\exp(-\mathrm{i}kP)$, so that

$$R_{22} = \exp(-2\mathrm{i}kh_{\mathrm{L}}) \tag{15.90}$$

where h_{L} is the contribution to h from within the layer and R_{22} is the reflection coefficient (15.86), which is referred to the level $z = z_{\mathrm{m}} - a$. Hence, for a transmitter and receiver at the ground

$$h(f) = z_{\mathrm{m}} - a + \mathrm{i}(\ln R_{22})/2k \tag{15.91}$$

and on applying (10.40)

$$\begin{aligned} h'(f) &= z_{\mathrm{m}} - a + \tfrac{1}{2}\mathrm{i}\,\mathrm{d}(\ln R_{22})/\mathrm{d}k \\ &= z_{\mathrm{m}} - a - \mathrm{i}a\frac{f}{f_{\mathrm{p}}}\frac{\mathrm{d}(\ln R_{22})}{\mathrm{d}(A^2)} \end{aligned} \tag{15.92}$$

where (15.67) for A^2 has been used with $C = 1$, $Z = 0$. Let k_{p} be the value of k for the penetration frequency $f = f_{\mathrm{p}}$. Then (15.67) shows that $\xi_+ = (ak_{\mathrm{p}})^{\frac{1}{2}}$ and (15.86) gives

$$\ln R_{22} = \ln\Gamma(\tfrac{1}{2} - \tfrac{1}{2}\mathrm{i}A^2) + \tfrac{1}{4}\pi A^2 + \mathrm{i}A^2(\tfrac{1}{2}\ln 2 + \ln\xi_+) + \text{constant}. \tag{15.93}$$

It was shown in § 11.16 that the time of travel of a wave packet is given approximately by $\mathrm{Re}\{h'(f)\}$ so only the imaginary part of (15.93) is needed in (15.92). The derivative of the logarithm of the gamma function is the psi function $\psi(x) = \mathrm{d}\{\ln\Gamma(x)\}/\mathrm{d}x$. Its properties are given by Abramowitz and Stegun (1965, § 6.3). From (15.93), (15.92)

$$h'(f) = z_{\mathrm{m}} - a + \tfrac{1}{2}a\frac{f}{f_{\mathrm{p}}}\{\ln(2ak_{\mathrm{p}}) - \psi(\tfrac{1}{2} - \tfrac{1}{2}\mathrm{i}A^2) - \tfrac{1}{2}\mathrm{i}\pi\}. \tag{15.94}$$

Since A^2 is real, the real part of the psi function is symmetric about $A^2 = 0$, $f = f_{\mathrm{p}}$, and $\mathrm{Re}\{h'(f)\}$ here has a maximum value

$$z_{\mathrm{m}} - a + \tfrac{1}{2}a\{\ln(2ak_{\mathrm{p}}) + \gamma + \ln 4\} = z_{\mathrm{m}} + 0.3283a + \tfrac{1}{2}a\ln(ak_{\mathrm{p}}) \tag{15.95}$$

where γ is Euler's constant and the formula $\psi(\tfrac{1}{2}) = -\gamma - \ln 4$ has been used.

For waves that travel right through the parabolic layer, the contribution to $h'(f)$ from the region between the ground and the top of the layer can be found as was done for ray theory at (12.24). The contribution is

$$h'(f) = z_{\mathrm{m}} - a + \mathrm{i}\,\mathrm{d}(\ln T_{22})/\mathrm{d}k \tag{15.96}$$

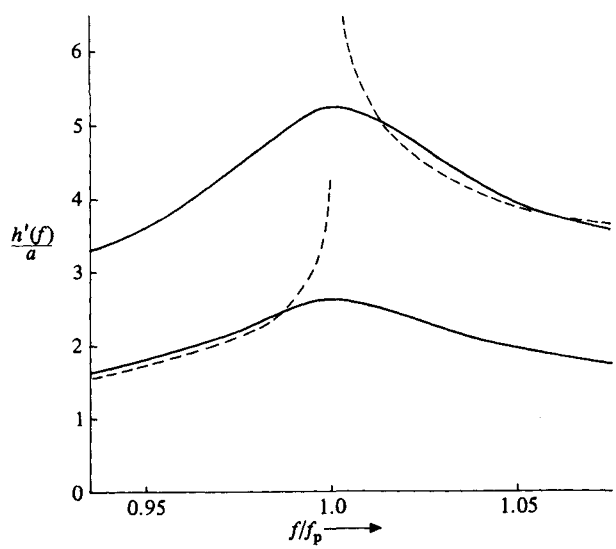
where T_{22} is the transmission coefficient. Now (15.87) shows that the real part of the last term of (15.96) is simply twice the real part of the last term of (15.91). Hence the contribution from the layer to $\mathrm{Re}\{h'(f)\}$ is twice the last term of (15.94).

Fig. 15.7 shows how these values of $\mathrm{Re}\{h'(f)\}$ depend on frequency and the corresponding curves for the ray theory, (12.22), (12.24), are shown for comparison. The difference is small when $|A|^2 \gg 0$, but near the penetration frequency the ray theory curves approach infinity whereas the full wave theory shows that there is a bounded upper limit. The range of frequency where the difference is appreciable depends on the thickness $2a$ of the layer. It is greatest when a is small.

15.12. The differential equations of theoretical physics

In the study of the reflection and transmission of radio waves by a stratified isotropic ionosphere, the electron height distributions $N(z)$ discussed earlier in this chapter have been deliberately chosen so that the governing differential equation can be transformed to one of the standard differential equations of theoretical physics. The number of suitable standard equations is very large. Lists of suitable equations are given by Brekhovskikh (1960), Ginzburg (1970), Heading (1977a), and these authors give further references. The general problem of what functions $N(z)$ can be studied in this way has been examined by Heading (1965). Solutions for various functions $N(z)$, including many that are not described in this book, have been given by Försterling and Wüster (1950), Brekhovskikh (1960), Wait (1962, ch. III), Gould and Burman (1964), Ginzburg (1970, ch. IV). Solutions for a spherically stratified isotropic medium were studied by Westcott (1968), Sharaf (1969), and for cylindrically stratified media by Westcott (1969b). Discussion of all the possibilities is not the object of this book. The purpose here is to set out the basic ideas, so that the interested reader can then follow up the numerous references for further examples.

Fig. 15.7. The continuous curves show the contribution to the equivalent height of reflection from a parabolic layer according to full wave theory when collisions and the earth's magnetic field are neglected. In this example $k_p\, a = 12.5$. The lower curve is for reflection and the upper curve is the contribution for waves that have penetrated the layer. The broken curves show the corresponding results for ray theory.



All these methods depend on the use of circuit relations. A given solution must first be separated into upgoing and downgoing component waves, both at a great height, above the reflecting levels, and at a low height that is usually in the free space below the ionosphere. The circuit relations are the connection formulae that relate the amplitudes of these component waves at the top and at the bottom (see § 8.19). In all the cases so far discussed the independent variable at great heights tends to infinity, and the differential equations studied, the Stokes equation, ch. 8, Bessel's equation, § 15.8, Weber's equation or its equivalent, §§ 15.9–15.11, all had an irregular singularity at infinity; see § 8.11. This led to the need for asymptotic approximations and the study of the Stokes phenomenon, when deriving the circuit relations.

There is one class of second order differential equations known as Fuchsian equations for which there are no irregular singularities. The most important is the hypergeometric equation, which has three regular singularities and no other singularity. Its solutions include many of the functions of theoretical physics as special cases or as limiting cases. It was used by Epstein (1930) and by Eckhart (1930) to find reflection and transmission coefficients. It can be used for a wide range of electron height distributions $N(z)$. An even wider range can be covered by an extension of this idea. The differential equation with four regular singularities is known as Heun's equation (Heun, 1889) and it has been used in some recent studies of radio propagation problems (van Duin and Sluijter, 1980).

The theory given by Epstein is summarised here as an example of this type of method. Some special cases of it are discussed in §§ 15.16–15.17.

15.13. The hypergeometric equation and its circuit relations

The theory summarised here is given in full by Whittaker and Watson (1927), Copson (1935), Olver (1974) and in other mathematical text books. A fuller version than the one that follows was given by Budden (1961a).

The hypergeometric equation may be written

$$\zeta(1-\zeta)\frac{\mathrm{d}^2u}{\mathrm{d}\zeta^2}+\{c-(a+b+1)\zeta\}\frac{\mathrm{d}u}{\mathrm{d}\zeta}-abu=0. \tag{15.97}$$

This is the form used by the authors cited above. Epstein (1930) used a different form with independent variable $u=-\zeta$, and the meanings of his symbols a, b, are different. The symbol c here is a new constant, not the speed of light. Equation (15.97) has regular singularities at $\zeta=0$, 1 and ∞. First, for the neighbourhood of $\zeta=0$ we seek a solution in ascending powers of ζ:

$$u=\zeta^{\beta}(1+a_1\zeta+a_2\zeta^2+\cdots). \tag{15.98}$$

Substitute in (15.97) and equate to zero the coefficients of successively increasing powers of ζ. The lowest power $\zeta^{\beta-1}$ gives

$$\beta = 0 \quad \text{or} \quad 1 - c. \tag{15.99}$$

The others give the coefficients a_1, a_2,.... Hence for $\beta = 0$ a solution is

$$\begin{aligned} u \equiv F(a,b;c;\zeta) = 1 + \frac{ab}{c\,1!}\zeta + \frac{a(a+1)b(b+1)}{c(c+1)2!}\zeta^2 + \cdots \\ + \frac{a(a+1)\ldots(a+r-1)b(b+1)\ldots(b+r-1)}{c(c+1)\cdots(c+r-1)r!}\zeta^r + \cdots. \end{aligned} \tag{15.100}$$

This is called the hypergeometric series. Similarly for $\beta = 1 - c$ a solution is

$$\begin{aligned} u &\equiv (-\zeta)^{1-c}F(a-c+1, b-c+1; 2-c; \zeta) \\ &= (-\zeta)^{1-c}\left\{1 + \frac{(a-c+1)(b-c+1)}{(2-c)\,1!}\zeta \right. \\ &\quad \left. + \frac{(a-c+1)(a-c+2)(b-c+1)(b-c+2)}{(2-c)(3-c)2!}\zeta^2 + \cdots \right\}. \end{aligned} \tag{15.101}$$

A factor $(-1)^{1-c}$ has here been introduced for later convenience. Both (15.100) and (15.101) can be shown to be convergent when $|\zeta| < 1$.

Alternatively we may take as a trial solution a series in descending powers of ζ. This is possible because the singularity of (15.97) at infinity is regular, which ensures that there must be at least one convergent series solution. Thus let

$$u = \zeta^{\beta}\left(1 + \frac{b_1}{\zeta} + \frac{b_2}{\zeta^2} + \cdots\right). \tag{15.102}$$

A similar treatment then gives

$$\beta = -a \quad \text{or} \quad -b \tag{15.103}$$

and thence two possible solutions are

$$u = (-\zeta)^{-a}F(a, a-c+1; a-b+1; \zeta^{-1}), \tag{15.104}$$

$$u = (-\zeta)^{-b}F(b, b-c+1; b-a+1; \zeta^{-1}). \tag{15.105}$$

Both are convergent if $|\zeta| > 1$. Again, factors $(-1)^{-a}$, $(-1)^{-b}$ have been introduced for later convenience. These two functions may have branch points at infinity, but they give solutions valid for any range 2π of $\arg\zeta$ and they do not display the Stokes phenomenon.

The solutions (15.100), (15.101), (15.104), (15.105) or their analytic continuations can have branch points at the singularities $\zeta = 0, 1, \infty$. A cut is therefore introduced in the complex ζ plane extending along the positive real axis. The solutions are then single valued provided that the cut is not crossed, and in future the values used will be those for which

$$|\arg(-\zeta)| < \pi. \tag{15.106}$$

The above method for deriving (15.101) might fail if c is an integer. Then the series would have to be replaced by a more complicated expression containing a logarithm. A similar failure could occur for (15.104) or (15.105) if a–b is an integer. But these integer values do not occur in the radio wave problem; see (15.118), (15.119) below.

The function (15.100) is defined by the series only in the range $|\zeta| < 1$, but the process of analytic continuation is now used to extend the definition outside this range. For $|\zeta| < 1$ it can be shown that

$$F(a,b;c;\zeta) = \frac{(c-1)!}{2\pi\mathrm{i}(a-1)!(b-1)!}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\frac{(a+s-2)!(b+s-2)!(-s)!}{(c+s-2)!}(-\zeta)^{s-1}\,\mathrm{d}s \tag{15.107}$$

where the contour runs to the right of the poles of $(a+s-2)!(b+s-2)!$ and to the left of the poles of $(-s)!$. Integrals of this kind were used by Barnes (1908). For a diagram showing the contour see Budden (1961a, fig. 17.6). For $|\zeta| < 1$ the contour can be closed by a large semi-circle to the right and the series (15.100) is the sum of the residues of the poles of $(-s)!$. For $|\zeta| > 1$ the integral still defines a function $F(a,b;c;\zeta)$ that is now called the hypergeometric function. The integral therefore provides the required analytic continuation. The contour can now be closed by a large semi-circle to the left, and thus the integral is expressed as the sum of two series from the residues of the poles of $(a+s-2)!$ and $(b+s-2)!$. The solution (15.100) is therefore equal to a linear combination of the two solutions (15.104), (15.105). Similarly the series (15.101) can be expressed as a Barnes integral, continued analytically into the region $|\zeta| > 1$, and then expressed as another linear combination of (15.104), (15.105). These two linear combinations are the required circuit relations. From them (15.104) can be expressed as a linear combination of (15.100), (15.101), and so also can (15.105). The results are

$$\frac{(-\zeta)^{-a}}{(a-b)!}F(a,1-c+a;1-b+a;\zeta^{-1}) = \frac{(-c)!}{(-b)!(a-c)!}F(a,b;c;\zeta) + \frac{(c-2)!(-\zeta)^{1-c}}{(c-b-1)!(a-1)!}F(a-c+1,b-c+1;2-c;\zeta), \tag{15.108}$$

$$\frac{(-\zeta)^{-b}}{(b-a)!}F(b,1-c+b;1-a+b;\zeta^{-1}) = \frac{(-c)!}{(-a)!(b-c)!}F(a,b;c;\zeta) + \frac{(c-2)!(-\zeta)^{1-c}}{(c-a-1)!(b-1)!}F(a-c+1,b-c+1;2-c;\zeta). \tag{15.109}$$

For the applications that follow, the hypergeometric functions have now fulfilled their purpose and need not be used again. Those on the right of (15.108), (15.109) will only be used when $|\zeta|$ is very small so that only the first terms (unity) of the series in (15.100), (15.101) are appreciable. Similarly those on the left will only be used for very large $|\zeta|$ and again only the first terms (unity) of the series are appreciable. Hence in the following sections, the hypergeometric functions will be set equal to unity.

15.14. Epstein distributions

In § 15.8 Bessel's equation (15.49) was studied. If it is compared with the hypergeometric equation (15.97) it is seen that both have a regular singularity at $\zeta = 0$. For Bessel's equation the substitution (15.47) was used so that ζ is proportional to $e^{\frac{1}{2}\alpha z}$ where z is height. Thus the singularity at $\zeta = 0$ is now where $z \to -\infty$ so that it is associated with the region of free space below the ionosphere. This suggests that a similar substitution should be used for the hypergeometric equation. We therefore now take

$$-\zeta = e^{\xi}, \quad \xi = (z/\sigma) + B. \tag{15.110}$$

This is equivalent to the substitution used by Epstein (1930). The constant σ determines the scale of the vertical structure of the model ionosphere. The constant B allows the origin of z to be chosen conveniently. In terms of ξ as independent variable, (15.97) becomes

$$(1 + e^{\xi})\frac{d^2u}{d\xi^2} + \{c - 1 + (a + b)e^{\xi}\}\frac{du}{d\xi} + abe^{\xi}u = 0. \tag{15.111}$$

This is now transformed to a version without a first derivative term by changing the dependent variable u to E thus

$$u = E\exp\{\tfrac{1}{2}(1 - c)\xi\}(1 + e^{\xi})^{\frac{1}{2}(c-1-a-b)} \tag{15.112}$$

which gives

$$\frac{d^2E}{dz^2} + k^2q^2E = 0 \tag{15.113}$$

where

$$q^2 = \eta_1 + \frac{e^{\xi}}{(e^{\xi} + 1)^2}\{(\eta_2 - \eta_1)(e^{\xi} + 1) + \eta_3\} \tag{15.114}$$

and

$$\eta_1 = -\tfrac{1}{4}(c - 1)^2/\sigma^2k^2, \tag{15.115}$$

$$\eta_2 = -\tfrac{1}{4}(a - b)^2/\sigma^2k^2, \tag{15.116}$$

$$\eta_3 = \tfrac{1}{4}(a + b - c + 1)(a + b - c - 1)/\sigma^2k^2. \tag{15.117}$$

Equation (15.113) is the same as (7.6) and E may be regarded as the electric field of a horizontally polarised wave obliquely incident on the ionosphere. Hence (15.114) determines what electron height distributions $N(z)$ can be studied with this theory. By suitably assigning the constants σ, B, η_1, η_2, η_3 a very wide range of models of the ionosphere can be investigated.

Below the ionosphere z is large and negative and (15.114) gives $q^2 \approx \eta_1$. But here $q = C$ so that $\eta_1 = C^2$ where $C = \cos\theta$ and θ is the angle of incidence. Then (15.115) gives

$$c - 1 = -2ik\sigma C. \tag{15.118}$$

The choice of sign here is arbitrary. The opposite sign would lead to the same final result. At a great height in the ionosphere z is large and positive and (15.114) gives $q^2 \approx \eta_2$. Hence q tends to a constant value, q_2 say, and (15.116) gives

$$a - b = -2\mathrm{i}k\sigma q_2 \tag{15.119}$$

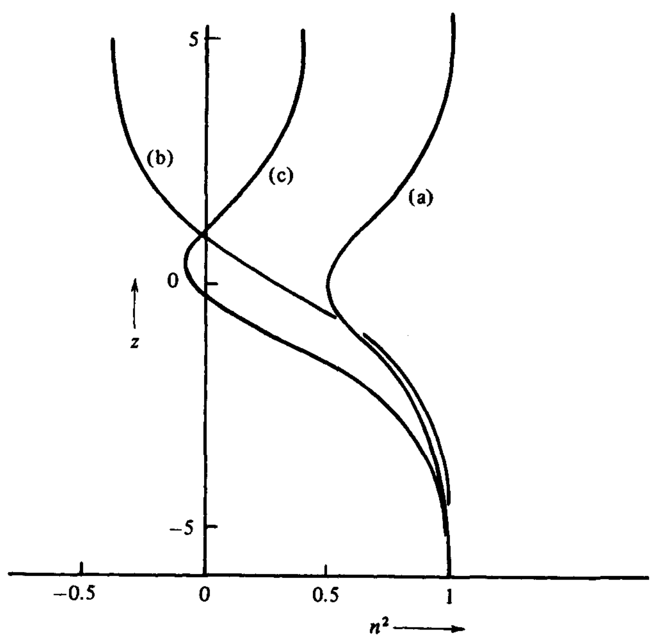
where again the choice of sign is arbitrary. At intermediate heights the dependence of q on z is determined by η_3 and (15.117) shows that

$$a + b - c = \pm(4k^2\sigma^2\eta_3 + 1)^{\frac{1}{2}} \tag{15.120}$$

where either sign may be chosen.

The ionospheric plasma is here assumed to be isotropic and (15.114) is equal to $n^2 - S^2 = C^2 - X/(1 - \mathrm{i}Z)$. Suppose now that σ is real and positive and $B = -z_{\mathrm{m}}/\sigma$ which is real. Then ξ is simply proportional to $(z - z_{\mathrm{m}})$, from (15.110). Suppose further that η_1, η_2, η_3 are all real. Then q^2 and n^2 are real when z is real, and $Z = 0$. Fig. 15.8 shows some typical curves of $n^2 = 1 - X$ vs ξ. In all the examples n^2 and thence X tend to constant values at small heights and at great heights. In between, the change may be monotonic or there may be one maximum, or one minimum. Two special cases are of particular interest. The first, discussed in § 15.16, is for $\eta_3 = 0$, so that n^2 makes a continuous monotonic change from unity below the ionosphere to a

Fig. 15.8. Typical Epstein distributions. All curves are for vertical incidence, $\theta = 0$. Curve (a) is a sech^2 distribution with $\eta_1 = \eta_2 = 1$, $\eta_3 = -3$. For curve (b) $\eta_1 = 1$, $\eta_2 = -0.4$, $\eta_3 = 0$. For curve (c) $\eta_1 = 1$, $\eta_2 = 0.4$, $\eta_3 = -3$.



constant value at great heights, fig. 15.8(b), thus

$$n^2 = 1 + (\eta_2 - C^2)e^{\xi}/(1 + e^{\xi}) = 1 + \tfrac{1}{2}(\eta_2 - C^2)\{1 + \tanh(\tfrac{1}{2}\xi)\}. \quad (15.121)$$

The second case is when $\eta_1 = \eta_2 = C^2$ so that there is free space above and below the ionosphere, and

$$n^2 = 1 + \tfrac{1}{4}\eta_3 \operatorname{sech}^2(\tfrac{1}{2}\xi). \quad (15.122)$$

This is the sech^2 distribution discussed in §§ 12.4, 15.17.

If B is complex a further group of distributions can be obtained from (15.114). Examples will be found in § 19.3 and in problem 15.1. Still other distributions can be obtained by letting some of the variables tend to infinity. For example let

$$B = i\pi - \ln\{\eta_2(1 - iZ)\}, \quad \eta_1 = C^2, \quad \sigma = 1/\alpha \quad (15.123)$$

and then let $\eta_2 \to \infty$. Then (15.114) gives $q^2 = C^2 - e^{\alpha z}/(1 - iZ)$ which is simply the exponential distribution (15.45). The reader should verify that the reflection coefficient for this case, (15.132) below, is the same as (15.59).

Epstein's transformation is about the simplest possible. When it is replaced by other transformations, a very wide range of height distributions $N(z)$ can be studied. The published papers on this topic are very numerous. Some of the most important are Rawer (1939), Burman and Gould (1954), Heading (1965).

15.15. Reflection and transmission coefficients for Epstein layers

In § 15.13 four independent solutions (15.100), (15.101), (15.104), (15.105) were found for the hypergeometric equation. The field variable E associated with each of these must now be examined. First, when z is large and negative, $|e^{\xi}| \ll 1$ and (15.112) becomes

$$E = u \exp\{\tfrac{1}{2}(c - 1)\xi\}. \quad (15.124)$$

In (15.100), (15.101) the hypergeometric functions are unity and when (15.110), (15.118) are used these solutions give

$$(15.100): \quad u = 1, \quad E = \exp\{\tfrac{1}{2}(c-1)\xi\} = \exp\{-ikC(z + \sigma B)\}, \quad (15.125)$$

$$(15.101): \quad u = (-\zeta)^{1-c}, \quad E = \exp\{\tfrac{1}{2}(1-c)\xi\} = \exp\{ikC(z + \sigma B)\}. \quad (15.126)$$

Hence (15.100) and (15.125) represent the upgoing or incident wave below the ionosphere and (15.101) and (15.126) represent the downgoing or reflected wave.

Next, when z is large and positive $|e^{z}| \gg 1$ and (15.112) becomes

$$E = u \exp\{\tfrac{1}{2}(a + b)\xi\}. \quad (15.127)$$

In (15.104), (15.105) the hypergeometric functions are unity when z is large and positive and when (15.110), (15.119) are used they give

$$(15.104): \quad u = (-\zeta)^{-a}, \quad E = \exp\{\tfrac{1}{2}(b-a)\xi\} = \exp\{ikq_2(z + \sigma B)\}, \quad (15.128)$$

$$(15.105): \quad u = (-\zeta)^{-b}, \quad E = \exp\{\tfrac{1}{2}(a-b)\xi\} = \exp\{-ikq_2(z + \sigma B)\}. \quad (15.129)$$

Hence (15.105) and (15.129) represent the upgoing wave or transmitted wave above the ionosphere, and (15.104) and (15.128) represent a downgoing wave that cannot be present when the only source of energy is below the ionosphere.

To find the reflection and transmission coefficients we select a solution in which (15.104) that is (15.128) is absent. The required solution is (15.105) and this is a linear combination of the solutions (15.100), (15.101) given by the circuit relation (15.109). Hence the solution below the ionosphere is, from (15.109), (15.124),

$$E=\frac{(-c)!}{(-a)!(b-c)!}\exp\{-\mathrm{i}kC(z+\sigma B)\}+\frac{(c-2)!}{(c-a-1)!(b-1)!}\exp\{\mathrm{i}kC(z+\sigma B)\} \tag{15.130}$$

and above the ionosphere it is

$$E=\frac{1}{(b-a)!}\exp\{-\mathrm{i}kq_2(z+\sigma B)\}. \tag{15.131}$$

The reflection coefficient R_{22} is the ratio of the second to the first term in (15.130) and for some reference level $z=z_1$ it is

$$R_{22}=\frac{(c-2)!(-a)!(b-c)!}{(c-a-1)!(b-1)!(-c)!}\exp\{2\mathrm{i}kC(z_1+\sigma B)\}. \tag{15.132}$$

The transmission coefficient T_{22} is the ratio of (15.131) to the first term of (15.130). If the transmitted wave is observed at $z=z_2$ and the incident wave at $z=z_1$, then

$$T_{22}=\frac{(-a)!(b-c)!}{(b-a)!(-c)!}\exp[-\mathrm{i}k\{q_2(z_2+\sigma B)-C(z_1+\sigma B)\}]. \tag{15.133}$$

15.16. Ionosphere with gradual boundary

As an illustration of the use of these formula consider the example of fig. 15.8 curve b, in which $\eta_3=0$ and n^2 is given by (15.121). Below the ionosphere where z is large and negative, $n^2=1$. At great heights where z is large and positive, n^2 is a constant equal to

$$n^2=\eta_2+S^2=1-X_2/(1-\mathrm{i}Z_2) \tag{15.134}$$

so that the medium is ultimately homogeneous and there is a continuous transition from this to free space as the height decreases. Let $B=-z_{\mathrm{m}}/\sigma$, $\eta_2=q_2^2$. Then the reflection coefficient (15.132) referred to the level $z_1=z_{\mathrm{m}}$ is

$$R_{22}=\frac{C-q_2}{C+q_2}\frac{(-2\mathrm{i}k\sigma C)!}{(2\mathrm{i}k\sigma C)!}\left[\frac{\{\mathrm{i}k\sigma(q_2+C)\}!}{\{\mathrm{i}k\sigma(q_2-C)\}!}\right]^2. \tag{15.135}$$

If the scale factor σ is very small, the transition from $q=C$ to $q=q_2$ occurs sharply where $z=0$. The factorials in (15.135) are then all unity and the reflection coefficient reduces to $(C-q_2)/(C+q_2)$ which is simply the Fresnel formula (11.57) for reflection at a sharp boundary.

The factor $(-2\mathrm{i}k\sigma C)!/(2\mathrm{i}k\sigma C)!$ in (15.135) is the ratio of two complex conjugates and has modulus unity. It therefore affects only the phase change at reflection. If electron collisons are neglected, n^2 is real and q_2 is either real and positive or purely imaginary. When it is purely imaginary the angle of incidence θ exceeds the critical angle. The last term of (15.135) is then the ratio of complex conjugates, and $|R_{22}| = 1$, so there is total reflection whether or not the boundary is sharp. When q_2 is real the moduli of the factorial functions may be found from the formula $x!(-x)! = \pi x/\sin(\pi x)$ whence it is easily shown that

$$|R_{22}| = \frac{\sinh\{\pi k\sigma(C - q_2)\}}{\sinh\{\pi k\sigma(C + q_2)\}}. \tag{15.136}$$

When $\sigma \to 0$ this reduces to the Fresnel formula. When σ is large it gives

$$|R_{22}| = \exp(-2\pi k\sigma q_2) \tag{15.137}$$

which tends to zero as $\sigma \to \infty$.

For large σ the medium varies very slowly with z, and the arguments of § 7.19 show that it ought then to be possible to find R_{22} from the phase integral formula (7.152). The upper limit z_0 of the integral in this formula is where $q = 0$, and (15.121) shows that then

$$e^{\xi} = -C^2/\eta_2, \text{ that is } z = z_m + \sigma \ln(-C^2/\eta_2) + 2\pi i\sigma r \tag{15.138}$$

where r is a positive, negative or zero integer. Thus q has an infinite number of zeros and it must be decided which is to be used for z_0 in (7.152). Now (15.134) shows that $-C^2/\eta_2$ always has a negative imaginary part and we choose the logarithm in (15.138) so that its imaginary part is in the range $-\pi$ to 0.

Suppose that $Z_2 = 0$ and $\eta_2 = q_2^2$ is real and negative. Then q_2 is imaginary and $|R_{22}| = 1$ as shown above. In this case one of the zeros of q is on the real axis. This is the point where, in the language of ray theory, the wave would be reflected. It must be the point to be chosen for z_0, because it is the only one for which (7.152) gives $|R_{22}| = 1$. Thus in (15.138) we use $r = 0$. If now Z_2 takes a small positive value, $\ln(-C^2/\eta_2)$ acquires a negative imaginary part. The point z_0 moves away from the real z axis to the negative imaginary side. A similar situation occurred for the parabolic distribution; see § 15.9 and fig. 15.3.

Now let Z_2 and X_2 in (15.134) be changed continuously. The zero of q at $z = a_0$, with $r = 0$, changes position but it can never cross the real z axis and is always the nearest zero to the real z axis on the negative imaginary side. In particular if $Z_2 = 0$ and $\eta^2 = q_2^2$ is real and negative, this zero is where $\mathrm{Im}(z_0) = -\pi\sigma$. It is the analogue of the point Q in fig. 15.3(a).

Evaluation of the phase integral for this problem is rather complicated. A method of doing it was given by Budden (1961a, § 20.5) for the special case $Z = 0$, $\theta = 0$, $C = 1$. If σ is large, the factorials in (15.135) can be given their asymptotic values from Stirling's formula. The values of R_{22} found from these two methods were the same. This confirms that the choice of the zero of q with $r = 0$ in (15.138) is correct.

The phase integral formula may therefore be used when the model ionosphere varies slowly enough, that is when σ is large. For smaller σ the more accurate formula (15.135) must be used, and when σ is infinitesimally small this reduces to the Fresnel formula (11.57) for the reflection coefficient of a sharp boundary. There is a continuous transition from reflection in a slowly varying medium at a zero of q, to reflection at a sharp boundary.

15.17. The 'sech^2' distribution

The distribution (15.122) is, for $B = -z_m/\sigma$,

$$n^2 = 1 + \tfrac{1}{4}\eta_3 \operatorname{sech}^2\{\tfrac{1}{2}(z - z_{\mathrm{m}})/\sigma\}. \tag{15.139}$$

There is free space both below and above the ionosphere where $|z - z_{\mathrm{m}}|$ is large. The ionosphere is a symmetrical layer with its centre at $z = z_{\mathrm{m}}$, where $n^2 = 1 + \frac{1}{4}\eta_3$. This distribution was used by Rawer (1939) to study partial penetration and reflection. It has also been studied by Westcott (1962c, d). It is in some respects better than the parabolic layer studied in §§ 15.9–15.11 since there is no discontinuity of $\mathrm{d}N/\mathrm{d}z$. Hence the 'sech^2' distribution can be used to study very thin layers and is not subject to the limitation (15.73).

Let

$$\eta_3 = -4X_{\mathrm{m}}/(1 - \mathrm{i}Z) \tag{15.140}$$

so that the collision frequency is independent of height and the electron concentration is given from

$$X = X_{\mathrm{m}} \operatorname{sech}^2\{\tfrac{1}{2}(z - z_{\mathrm{m}})/\sigma\}. \tag{15.141}$$

From (15.118), (15.119) it follows that $a - b = c - 1 = -2\mathrm{i}k\sigma C$. Let $4k^2\sigma^2\eta_3 + 1 = 4\gamma^2$ so that $a + b - c = 2\gamma$ from (15.120). Then the reflection coefficient (15.132) referred to the level $z_1 = z_{\mathrm{m}}$ is

$$R_{22} = -\frac{(-2\mathrm{i}k\sigma C)!\,(2\mathrm{i}k\sigma C - \gamma - \frac{1}{2})!\,(2\mathrm{i}k\sigma C + \gamma - \frac{1}{2})!}{(2\mathrm{i}k\sigma C)!\,(-\gamma - \frac{1}{2})!\,(\gamma - \frac{1}{2})!} \tag{15.142}$$

and the transmission coefficient (15.133) when both incident and transmitted waves are referred to the same level, $z_1 = z_2$, is

$$T_{22} = 2\mathrm{i}kC\frac{(2\mathrm{i}k\sigma C - \gamma - \frac{1}{2})!\,(2\mathrm{i}k\sigma C + \gamma - \frac{1}{2})!}{\{(2\mathrm{i}k\sigma C)!\}^2}. \tag{15.143}$$

It can be verified that $R_{22} \to 0$ when $\sigma \to 0$, that is when the layer becomes indefinitely thin.

If $Z = 0$, η_3 is $-4k_{\mathrm{p}}^2/k^2$ from (15.140) where k_{p} is the value of k for the penetration frequency f_{p}, and

$$4\gamma^2 = 1 - 16\sigma^2 k_{\mathrm{p}}^2. \tag{15.144}$$

Thus γ is independent of frequency, and is either real or purely imaginary. In both cases the moduli of the factorial functions can be found from the formula $x!(-x)! = \pi x/\sin(\pi x)$, and (15.142), (15.143) give

$$|R_{22}|^2 = \frac{\cos 2\pi\gamma + 1}{\cos 2\pi\gamma + \cosh 4\pi k\sigma C}, \quad |T_{22}|^2 = \frac{\cosh 4\pi k\sigma C - 1}{\cos 2\pi\gamma + \cosh 4\pi k\sigma C}. \tag{15.145}$$

These formulae can be used to study the partial penetration and reflection near the penetration frequency of the layer.

It is interesting to compare the 'sech^2' distribution with a parabolic distribution having the same penetration frequency and the same curvature at its maximum; see fig. 12.1. It can be shown that the parabola then has half thickness $a = 2\sigma$. For frequencies greater than about 2 MHz most ionospheric layers have $a/\lambda \geqslant 2$ so that in the cases of greatest interest $16\sigma^2 k_\mathrm{p}^2 \gg 1$ and $\mathrm{i}\gamma \approx 2\sigma k_\mathrm{p} \gg 1$. Thus $\cos 2\pi\gamma \gg 1$ so that the 1 may be neglected in the first numerator of (15.145) and we may take $\cos 2\pi\gamma \approx \frac{1}{2}\exp(4\pi\sigma k_\mathrm{p})$. Then, for frequencies close to penetration, $k \approx k_\mathrm{p}/C$, the first formula (15.145) reduces to (15.89) for the parabolic distribution.

When $Z \neq 0$ the formulae (15.142), (15.143) cannot be so easily simplified and the factorial functions with complex argument have to be computed. Rawer (1939) has computed $|R_{22}|^2$ and $|T_{22}|^2$ for various values of Z and curves are given in his paper.

Equations (15.145) show that $|T_{22}|/|R_{22}| = \sinh 2\pi k\sigma C/\cos \pi\gamma$. At high enough frequencies $\sinh 2\pi k\sigma C \approx \frac{1}{2}\exp(2\pi k\sigma C)$ so that

$$\ln|T_{22}/R_{22}| \approx 2\pi k\sigma C + \text{constant}. \tag{15.146}$$

If this is plotted against frequency, proportional to k, it gives a straight line whose slope gives σ, which is a measure of the thickness of the ionospheric layer. This method has been applied to observations at vertical incidence by Briggs (1951). Even if electron collisions are included, it may be assumed that $Z \ll 1$ at high frequencies so that γ is still nearly constant and the method can still be used.

The equivalent height of reflection for radio waves vertically incident on a 'sech^2' layer can be found by the method used in §15.11 for the parabolic layer. Since the centre of the 'sech^2' layer is at a height z_m above the ground

$$h'(f) = z_\mathrm{m} + \tfrac{1}{2}\mathrm{i}\,\mathrm{d}(\ln R_{22})/\mathrm{d}k \tag{15.147}$$

where R_{22} is given by (15.142) with $C = 1$; compare (15.92). Formulae derived from this, with collisions included, were given by Rawer (1939) who also gave curves of $h'(f)$ versus f for various values of Z and σ.

15.18. Other electron height distributions

Two other distributions $N(z)$ that come within the subject of this chapter should be briefly mentioned. The first is the sinusoidal distribution

$$X = \begin{cases} \frac{1}{2}X_\mathrm{m}[1 + \cos\{\pi(z - z_\mathrm{m})/a\}] & \text{for } |z - z_\mathrm{m}| \leqslant a, \\ 0 & \text{for } |z - z_\mathrm{m}| \geqslant a. \end{cases} \tag{15.148}$$

It has no discontinuities of dN/dz but there are discontinuities of curvature at the top and bottom of the layer. If Z is independent of height, the differential equation (7.6) for horizontal polarisation can be transformed into one of the standard forms of Mathieu's equation (see, for example, Whittaker and Watson, 1927, ch. XIX), by a change of the independent variable. But the reflection and transmission coefficients, R, T, for this problem can be found very easily by one of the numerical methods described in ch. 18. Some results for $Z = 0$ and vertical incidence were given by Budden (1961a, § 17.8), showing how $|R|$ depends on frequency for frequencies extending on both sides of penetration. They show the phenomenon of partial penetration and reflection very similar to that for the parabolic layer, fig. 15.5.

The second distribution $N(z)$ has the square law dependence

$$X = \begin{cases} \beta(z-h_0)^2 & \text{for } z \geqslant h_0, \\ 0 & \text{for } z \leqslant h_0. \end{cases} \tag{15.149}$$

It is of interest because there is a discontinuity of curvature d^2N/dz^2 at $z = h_0$. The equation (7.6) for horizontal polarisation can be transformed to Weber's equation (15.70). The reflection coefficient for this case was given by Hartree (1931a) and Rydbeck (1944). A closely related problem, with some allowance for the earth's magnetic field, was discussed by Wilkes (1940). Further details are given by Budden (1961a, § 17.7).

15.19. Collisions. Booker's theorem

The effect of electron collisions has been allowed for in several of the functions $n^2(z)$ studied in this chapter. The following discussion gives another way of dealing with collisions in certain types of isotropic ionosphere.

Suppose that for a loss-free isotropic ionosphere, in which $X(z)$ is a given analytic function of height z, the reflection coefficient R for some real reference level $z = z_1$ has been found for an angle of incidence θ. It is assumed that there is free space in a range of height below z_1. This will be called the 'original system'. The following results apply for any θ and for either horizontal or vertical polarisation. Define a new variable

$$\check{z} = z - \mathrm{i}b \tag{15.150}$$

where b is a real positive constant. Let $\check{z}$ now be treated as the height variable instead of z in a new model of the ionosphere so that real $\check{z}$ refers to real heights. This will be called the 'displaced system', because it is formed by a displacement b in the imaginary direction of the complex z plane. The real $\check{z}$ axis is where $\mathrm{Im}(z) = \mathrm{i}b$. Let $\check{R}$ be the reflection coefficient for reference level $\check{z} = z_1$ in the displaced system. Let F be the field component used to define R, so that F is $\mathscr{H}_y$ for vertical polarisation and E_y for horizontal polarisation. Then in the free space below the ionosphere the z

dependence of F is given by

$$\begin{aligned} F &= \exp\{-ik\cos\theta(z-z_1)\} + R\exp\{ik\cos\theta(z-z_1)\} \\ &= \exp(kb\cos\theta)[\exp\{-ik\cos\theta(\check{z}-z_1)\} \\ &\quad + R\exp(-2kb\cos\theta)\exp\{ik\cos\theta(\check{z}-z_1)\}], \end{aligned} \tag{15.151}$$

whence

$$\check{R} = R\exp(-2kb\cos\theta). \tag{15.152}$$

This new reflection coefficient $\check{R}$ is the same as if the reference level in the original system was changed to the complex value $z_1 + ib$. When $\check{z}$ is real $X(z)$ is complex and given by

$$X(\check{z}+ib) = C(\check{z}) + iD(\check{z}) = \mathscr{X}/(1+\mathscr{Z}^2) + i\mathscr{X}\mathscr{Z}/(1+\mathscr{Z}^2) \tag{15.153}$$

where $C, D, \mathscr{X}, \mathscr{Z}$ are functions of $\check{z}$ that are real when $\check{z}$ is real. The displaced model is a new model of the ionosphere in which X is $\mathscr{X}(\check{z})$ and Z is $\mathscr{Z}(\check{z})$, and the reflection coefficient is given by (15.152). This result was first given by Booker, Fejer and Lee (1968). Later writers have referred to it as 'Booker's theorem'. By choosing various values of b a whole range of new models can be derived from the original loss-free system.

For this method to work it is necessary first that $\mathscr{X}(\check{z})$ and $\mathscr{Z}(\check{z})$ shall be physically realisable functions, and second that at great heights in the displaced system the only wave present is an upgoing wave. As a simple example, given by Booker *et al.* (1968), suppose that for the original system $X(z) = e^{\alpha z}$ as in (15.44). Then (15.153) gives

$$\mathscr{X} = \exp(\alpha\check{z})\sec(\alpha b), \quad \mathscr{Z} = \tan(\alpha b) \tag{15.154}$$

so that $\mathscr{Z}$ is independent of height $\check{z}$. This is the same as in the example of § 15.8. It is easily checked that (15.152) is in agreement with (15.59), (15.60).

An essential requirement is that the original function $X(z)$ shall be analytic and shall not have any singularities in the region $0 < \mathrm{Im}(z) \leqslant b$. The theorem cannot be applied if $X(z)$ is only piecewise analytic. For example in the linear model of § 15.2, $X(z)$ has a discontinuity of gradient where $z = h_0$. When $X(z)$ and the solutions (15.8), (15.7) are continued analytically to use complex z, the two real regions $z < h_0$ and $z > h_0$ are continued separately to give two different complexions of the complex z plane. The two solutions (15.8), (15.7) were made to match at the real point $z = h_0$. They do not match at any other point in the two complexions of the z plane. For similar reasons the theorem cannot be applied to the parabolic model, §§ 15.9–15.11. It was shown by von Roos (1970) that if the earth's magnetic field is allowed for, the functions $\mathscr{X}(z)$, $\mathscr{Z}(z)$ that result from a displacement are not physically realisable.

The most useful application of the theorem is for an isotropic ionosphere where the original $X(z)$ is some combination of exponential functions. The Epstein functions, §15.14, are important examples. Heading (1972) has shown that the theorem can be generalised as follows. If the original real function $X(z)$ is multiplied by a complex constant A, it may still be possible to find the reflection coefficient R, even though this modification is physically unrealisable. But if a displacement b is now imposed as in (15.150), new functions $\mathscr{X}(z)$, $\mathscr{Z}(z)$ can be found that are physically realisable. This extends the range of usefulness of the theorem. Heading has used this to derive three different examples of ionospheric models with collisions, using loss-free Epstein distributions as the original system.

PROBLEMS 15

15.1. Consider a model isotropic ionosphere in which the electron concentration is the same at all heights so that X is a constant, and the collision frequency is height dependent so that $Z = Z_g \exp(-z/\sigma)$. For z large and negative, Z is so large that $n^2 = 1 - X/(1 - \mathrm{i}Z) \approx 1$ and the medium is like free space. At great heights Z is negligible and the medium is homogeneous with refractive index $n_2 = (1 - X)^{\frac{1}{2}}$. Show that the reflection coefficient R for vertical incidence can be found by using an Epstein distribution (15.110)–(15.117) with $\eta_1 = 1$, $\eta_2 = 1 - X$, $\eta_3 = 0$, $B = \frac{1}{2}\mathrm{i}\pi - \ln Z_g$. Show that if n_2^2 is positive

$$|R| = \frac{\sinh\{\pi k\sigma |n_2 - 1|\}}{\sinh\{\pi k\sigma(n_2 + 1)\}} \exp(-\pi k\sigma)$$

and if n_2^2 is negative, $|R| = \exp(-\pi k\sigma)$. For an example of the use of a model of this type see Greifinger and Greifinger (1965).

For the solution see Budden (1961a, §17.17).