19

Applications of full wave methods

19.1. Introduction

This chapter is concerned with various applications of the full wave methods discussed in ch. 18. The object is to illustrate general principles but not to give details of the results. The number of possible applications is very large and only a selection can be given here. The topics can be divided into two groups: (a) problems where the solutions can be expressed in terms of known functions, and (b) problems where computer integration of the differential equations, or an equivalent method as in §§ 18.2–18.11, is used. In nearly all applications of group (a) it is necessary to make substantial approximations. The group (a) can be further subdivided into those cases where the fourth order governing differential equations are separated into two independent equations each of the second order, and those where the full fourth order system must be used. For the separated second order equations the theory is an extension of ch. 15 which applied for an isotropic medium, and many of its results can be used here.

In all the examples of this chapter the ionosphere is assumed to be horizontally stratified, with the z axis vertical. The incident wave is taken to be a plane wave with its wave normal in the x–z plane. It is assumed that the only effective charges in the plasma are the electrons.

In §§ 19.2, 19.3 it is assumed that the earth's magnetic field is vertical and it is convenient to give here the form of the basic differential equations (7.80) for this case. Thus $l_x = l_y = 0$, $l_z = 1$ for the northern hemisphere. Let

$$a = 1/(U^2 - Y^2). \tag{19.1}$$

Then the susceptibility matrix (3.35) is

$$\boldsymbol{M} = -X \begin{pmatrix} Ua & -\mathrm{i}Ya & 0 \\ \mathrm{i}Ya & Ua & 0 \\ 0 & 0 & 1/U \end{pmatrix}. \tag{19.2}$$

If this is now used in (7.81) for **T**, the equations (7.80) are

$$\begin{aligned} E'_x &= \mathrm{i}\left(\frac{S^2}{1-X/U} - 1\right)\mathscr{H}_y, \\ E'_y &= \mathrm{i}\mathscr{H}_x, \\ \mathscr{H}'_x &= XYaE_x + \mathrm{i}(C^2 - XUa)E_y, \\ \mathscr{H}'_y &= -(1 - XUa)E_x + XYaE_y. \end{aligned} \tag{19.3}$$

19.2. Vertical incidence and vertical magnetic field

As a simple but important example where the basic equations are separable consider the case where both the vector $\boldsymbol{Y}$ and the wave normals are vertical. Then $S = 0$, $C = 1$ and when these are used in (19.3) they can be rearranged to give

$$(E_x + \mathrm{i}E_y)' = -(\mathscr{H}_x + \mathrm{i}\mathscr{H}_y), \quad (\mathscr{H}_x + \mathrm{i}\mathscr{H}_y)' = \{1 - X/(U+Y)\}(E_x + \mathrm{i}E_y), \tag{19.4}$$

and

$$(E_x - \mathrm{i}E_y)' = (\mathscr{H}_x - \mathrm{i}\mathscr{H}_y), \quad (\mathscr{H}_x - \mathrm{i}\mathscr{H}_y)' = \{1 - X/(U-Y)\}(E_x - \mathrm{i}E_y). \tag{19.5}$$

The equations have separated into these two independent pairs. By eliminating $\mathscr{H}_x \pm \mathrm{i}\mathscr{H}_y$ the equations are converted to the two independent second order equations

$$(E_x + \mathrm{i}E_y)'' + \{1 - X/(U+Y)\}(E_x + \mathrm{i}E_y) = 0, \tag{19.6}$$

$$(E_x - \mathrm{i}E_y)'' + \{1 - X/(U-Y)\}(E_x - \mathrm{i}E_y) = 0. \tag{19.7}$$

The expressions in brackets $\{\}$ are the two values of the squared refractive index n^2 as given by (4.95) for purely longitudinal propagation. Equations (19.6), (19.7) are a special case of Försterling's equations (16.90), (16.91). The coupling parameter ψ is zero in this case.

This separation is important for it shows that there are two waves that are propagated and reflected independently. For example suppose that a wave is present for which, at some level, $E_x - \mathrm{i}E_y$ and $\mathscr{H}_x - \mathrm{i}\mathscr{H}_y$ are both zero. Then (19.5) shows that they remain zero at all levels, so that the wave is circularly polarised and the rotation is clockwise to an observer looking upwards. Thus, where it is possible to speak of upgoing and downgoing waves, for example below the ionosphere, the upgoing wave has a right-handed sense and the downgoing wave a left-handed sense; see §§ 4.3, 11.6. The propagation is governed by (19.4) or (19.6) and they show that reflection would occur near where $X = U + Y$. Similarly if $E_x + \mathrm{i}E_y$ and $\mathscr{H}_x + \mathrm{i}\mathscr{H}_y$ are zero at some level, the wave is circularly polarised but now with the opposite sense. Its propagation is governed by (19.7) and reflection would occur near where $X = U - Y$. In practice the incident wave may be linearly polarised. It must

then be resolved into two circularly polarised components whose reflection coefficients are found separately. The reflected waves are then recombined below the ionosphere. An example of this process follows.

For radio waves of very low frequency it is of interest to study the exponential form

$$N(z) \propto X(z) = \mathrm{e}^{\alpha z} \tag{19.8}$$

for the electron height distribution low down in the ionosphere, for the reasons given in § 15.8. Thus the level $z = 0$ is where $X = 1$. As in that section we also let Z be independent of z. The equations (19.6), (19.7) can be considered together. Let

$$F_\mathrm{r} = E_x + \mathrm{i}E_y, \quad F_\mathrm{l} = E_x - \mathrm{i}E_y, \quad b_\mathrm{r} = 1/(U+Y), \quad b_\mathrm{l} = 1/(U-Y) \tag{19.9}$$

where the subscripts r and l indicate that the upgoing incident wave has right- or left-handed circular polarisation, respectively. Then both (19.6) and (19.7) may be written

$$\mathrm{d}^2F/\mathrm{d}z^2 + k^2(1 - b\mathrm{e}^{\alpha z})F = 0 \tag{19.10}$$

where F and b are to have the same subscript, r or l. This is of the same form as (15.46) and can be treated in the same way. Let

$$\nu = 2\mathrm{i}k/\alpha; \tag{19.11}$$

compare (15.48). This ν must not be confused with the ν used elsewhere for the electron collision frequency. The method of § 15.8 can now be applied to (19.10). The reflection coefficient is given by (15.59) with $C = 1$ and with $(1 - \mathrm{i}Z)^{-1}$ replaced by b, thus

$$R = -b^\nu \left(\frac{k}{\alpha}\right)^{2\nu} \frac{(-\nu)!}{\nu!} \exp(-2\mathrm{i}kh_1) \tag{19.12}$$

where h_1 is the distance from the reference level (§ 11.2) to the level where $X = 1$, $z = 0$. Since ν is purely imaginary, the two factorial functions are complex conjugates and their ratio has modulus unity. The factor $(k/\alpha)^{2\nu}$ also has modulus unity. Hence

$$|R| = |b^{2\mathrm{i}k/\alpha}| = \exp\left(-\frac{2k}{\alpha} \arg b\right). \tag{19.13}$$

Now (19.12) gives the elements (11.33) of the form $\boldsymbol{R}_0$ of the reflection coefficient matrix defined in 11.6. If the polarisation of the incident wave is right-handed circular, the reflected wave is left-handed circular and has no right-handed component. Hence $R_{\mathrm{rr}} = 0$, and similarly $R_{\mathrm{ll}} = 0$, so that

$$\boldsymbol{R}_0 = \begin{pmatrix} 0 & R_{\mathrm{rl}} \\ R_{\mathrm{lr}} & 0 \end{pmatrix}. \tag{19.14}$$

Now let b be b_r. Then (19.12) gives $R_{1\mathrm{r}}$. Similarly if b is b_l, (19.12) is $R_{\mathrm{r}1}$. Then (19.13)

gives

$$|R_{1r}| = \exp\left(-\frac{2k}{\alpha}\arctan\frac{Z}{1+Y}\right), \quad |R_{r1}| = \exp\left(-\frac{2k}{\alpha}\arctan\frac{Z}{1-Y}\right). \tag{19.15}$$

If electron collisions are neglected, $Z = 0$, then $|R_{1r}| = 1$ so the wave is totally reflected. For this case the refractive index is zero where $X = 1 + Y$, and according to ray theory of ch. 13 there should be total reflection at this level in a slowly varying medium. We have now shown that for exponential $N(z)$ when collisions are neglected there is total reflection of this component even when the medium is not slowly varying. For frequencies greater than the gyro-frequency, $Y < 1$, and (19.15) shows that if collisions are neglected $|R_{r1}| = 1$ so the other component also is totally reflected. Here the refractive index is zero where $X = 1 - Y$.

These formulae are more interesting, however, for low frequencies so that $Y > 1$. Now (19.9) shows that $\arg b_1 = \pi - \arctan\{Z/(Y-1)\}$ and the second equation (19.15) is

$$|R_{r1}| = \exp\left(-\frac{2\pi k}{\alpha} + \frac{2k}{\alpha}\arctan\frac{Z}{Y-1}\right) \tag{19.16}$$

where the arctan is in the range 0 to $\frac{1}{2}\pi$. If $Z = 0$, $|R_{r1}| = \exp(-2\pi k/\alpha)$. In a slowly varying medium α is small so that the reflection coefficient is small. Now the condition $X = 1 - Y$ is not obeyed at any real height and according to the ray theory of ch. 13 there should be no reflection of this component. The full wave theory shows that for exponential $N(z)$ there is some reflection. It is small in a slowly varying medium but becomes larger when the medium varies more quickly.

If the incident wave is linearly polarised, we may take its plane of polarisation to be the x–z plane. Then the reflecting properties of the ionosphere are conveniently specified by the elements R_{11}, R_{21} of the matrix $\boldsymbol{R}$. This is found from the transformation (11.33) that uses the unitary matrix (11.30) and gives

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{R}_0\boldsymbol{U}^{-1} = \tfrac{1}{2}\begin{pmatrix} -R_{r1} - R_{1r} & \mathrm{i}(R_{r1} - R_{1r}) \\ \mathrm{i}(R_{r1} - R_{1r}) & R_{r1} + R_{1r} \end{pmatrix} \tag{19.17}$$

whence, from (19.12):

$$\begin{aligned} R_{11} &= -R_{22} = \tfrac{1}{2}\left(\frac{k}{\alpha}\right)^{2\nu}\frac{(-\nu)!}{\nu!}(b_r^\nu + b_1^\nu), \\ R_{21} &= R_{12} = \tfrac{1}{2}\mathrm{i}\left(\frac{k}{\alpha}\right)^{2\nu}\frac{(-\nu)!}{\nu!}(b_r^\nu - b_1^\nu). \end{aligned} \tag{19.18}$$

These formulae are of interest in the special case of very low frequencies when $Y \gg 1$; for example at 16 kHz, $Y \approx 80$. If the 1 is neglected compared with Y in (19.9)

$$b_r = 1/(Y - \mathrm{i}Z), \quad b_1 = -1/(Y + \mathrm{i}Z), \tag{19.19}$$

so that $|b_r| = |b_1|$. Then b_r^ν, b_1^ν have the same argument since ν is purely imaginary, and b_r^ν has the greater modulus. Then (19.18) shows that R_{11}, R_{21} are in quadrature, so that when the incident wave is linearly polarised, the polarisation ellipse of the reflected wave has its major axis in the plane of polarisation of the incident wave. It can be shown that in the northern hemisphere the rotation in the reflected wave has a left-handed sense. Further from (19.18)

$$|R_{11}| = \tfrac{1}{2}\left[\exp\left(-\frac{2k}{\alpha}\arctan\frac{Z}{Y}\right) + \exp\left\{-\frac{2k}{\alpha}\left(\pi - \arctan\frac{Z}{Y}\right)\right\}\right]$$
$$= \mathrm{e}^{-\pi k/\alpha}\cosh\left\{\frac{k}{\alpha}\left(\pi - 2\arctan\frac{Z}{Y}\right)\right\} \tag{19.20}$$
$$|R_{21}| = \mathrm{e}^{-\pi k/\alpha}\sinh\left\{\frac{k}{\alpha}\left(\pi - 2\arctan\frac{Z}{Y}\right)\right\}.$$

These results were given by Stanley (1950) who derived them in a different way as a limiting case of the Epstein distribution, § 15.14. They are important because in observations with a frequency of 16 kHz at nearly vertical incidence (Bracewell *et al.*, 1951) it was found that the reflected wave was elliptically polarised with a left-handed sense and with the major axis of the ellipse parallel to the plane of polarisation of the incident wave.

The two differential equations (19.6), (19.7) are of the same form as (7.6) which was solved for various electron height distributions in ch. 15 where the earth's magnetic field was neglected. Thus all the distributions $N(z)$ used in ch. 15 can also be used for vertical incidence when the earth's magnetic field also is vertical. For example with the parabolic $N(z)$ used in § 15.9, the differential equation (15.66) is replaced by two equations obtained by putting F_r, F_l (19.9) for E_y, $1 \pm Y - \mathrm{i}Z$ for $1 - \mathrm{i}Z$, and $C^2 = 1$. The two reflection coefficients R_{lr} and R_{rl} are then found, by analogy with (15.86), and combined as in (19.17) to give the elements of $\boldsymbol{R}$. Some results for this case were given by Pfister (1949). Wilkes (1940) discussed the case where $N(z)$ is proportional to the square of the height; see § 15.18. He also discussed the case where N is independent of z, and $Z \propto 1/z$, which can be solved in terms of confluent hypergeometric functions. For a similar model see problem 15.1.

19.3. Oblique incidence and vertical magnetic field

When the waves are obliquely incident on the ionosphere, or when the earth's magnetic field is oblique, the equations (7.80) cannot in general be separated into two second order equations. They are equivalent to a single fourth order equation. There are very few cases where the properties of fourth order equations have been studied; most of the functions of theoretical physics are solutions of second order equations. Hence some approximations must be made before progress is possible.

A very important solution was given by Heading and Whipple (1952), for oblique

incidence and vertical field, and another similar solution was given by Heading (1955) for vertical incidence when the earth's magnetic field is oblique and in the plane of incidence. The principles of the two methods are similar and only the first is described here in outline. Heading and Whipple made the following assumptions:

(1) $|1 - iZ| \ll Y$. Thus the electron collision frequency is very small compared with the gyro-frequency.

(2) X increases monotonically with the height z.

(3) At the level where $X \approx |1 - iZ|$, Z is constant and X is proportional to $e^{\alpha z}$ where α is constant.

(4) At the level where $X \approx Y$, X is proportional to $e^{\gamma z}$ where γ is another constant. Here Z is negligible because of (1).

The assumption (1) may not be true in the D-region, but still the method is useful as an illustration of some physical principles.

Let the ionosphere be divided into five regions as follows. The rather unusual numbering is used to agree with Heading and Whipple who speak only of regions I and II.

Region II(a)	$X \gg Y$.
Region II	$\|1 - iZ\| \ll X \approx Y$ (partially reflecting)
Region I(a)	$\|1 - iZ\| \ll X \ll Y$
Region I	$X \approx \|1 - iZ\|$ (partially reflecting)
Region 0	$X \ll \|1 - iZ\| \ll Y$ (like free space).

The lowest region, 0, is like free space because X is very small. In the highest region II(a), X is very large and both refractive indices are large, of order $X^{\frac{1}{2}}$. Here the W.K.B. solutions are good approximations. This is above all reflecting levels so the solution here must contain only upgoing waves. Some reflection occurs in regions I and II. In region I(a) there is no reflection but the waves have some interesting features.

In the three lowest regions the conditions $X \ll Y$ and $|U| \equiv |1 - iZ| \ll Y$ both hold. Then a (19.1) is negligible, and the equations (19.3) become

$$E'_x = -i\frac{C^2 - X/U}{1 - X/U}\mathscr{H}_y, \quad \mathscr{H}'_y = -iE_x, \tag{19.22}$$

$$E'_y = i\mathscr{H}_x, \quad \mathscr{H}'_x = iC^2 E_y. \tag{19.23}$$

They have separated into two independent pairs. The pair (19.23) refers to linearly polarised waves with the electric vector horizontal, and shows that these waves are propagated just as though regions 0, I, I(a) are free space.

The pair (19.22) refers to linearly polarised waves with the electric vector in the plane of incidence. It is shown below that these can undergo partial reflection in region I. In region I(a) $|X/U| \gg 1$ so the C^2 and the 1 in the first equation (19.22) are

negligible, and the equations take the very simple form

$$E_x' = -\mathrm{i}\mathscr{H}_y, \quad \mathscr{H}_y' = -\mathrm{i}E_x. \tag{19.24}$$

In formulating the original equations (7.80) it was assumed, § 7.13, that all field quantities include a factor $\exp(-\mathrm{i}kSx)$. If this is restored, one solution of (19.24) is

$$E_x = \mathscr{H}_y = \exp\{-\mathrm{i}k(Sx+z)\} \tag{19.25}$$

which represents a progressive wave travelling obliquely upwards with phase velocity $c(1+S^2)^{-\frac{1}{2}}$ so that the refractive index is $(1+S^2)^{\frac{1}{2}}$. Its wave normal is at an angle ψ to the vertical where $\tan\psi = S = \sin\theta$. The vertical component E_z can be found from (7.78) which gives $E_z = -S\mathscr{H}_y/(1-X/U)$, and this is negligible because $|X/U| \gg 1$. Hence the electric field $\boldsymbol{E}$ is horizontal and in the plane of incidence. There is a second solution of (19.24) namely

$$E_x = -\mathscr{H}_y = \exp\{-\mathrm{i}k(Sx-z)\} \tag{19.26}$$

that represents a wave travelling obliquely downwards.

The above results may be summarised in physical terms as follows. The assumption $Y \gg X$ means that the earth's magnetic field is extremely large so that electrons are prevented from moving across it. Hence the electrons move only vertically, and will not respond to forces tending to move them horizontally. This explains why a wave with its electric vector horizontal is unaffected by the electrons, so the three lower regions behave like free space for this wave. In region II and above, however, X is comparable with Y and the electrons are now so numerous that they have some effect on the wave.

For a wave with its electric vector in the plane of incidence there is a force eE_z tending to move each electron vertically. But in region I(a) where $X \gg |1-\mathrm{i}Z|$ the electrons are so numerous that the medium effectively has infinite conductivity in the z direction. Hence it cannot sustain a vertical component of $\boldsymbol{E}$ and the only electric field is horizontal, as in (19.25), (19.26). In region 0 there are so few electrons that the medium is like free space. The transition occurs in region I which is a region of finite conductivity, increasing as the height increases.

In a medium with indefinitely large Y, (3.50) shows that the principal axis elements of the electric permittivity tensor $\boldsymbol{\varepsilon}$ are $\varepsilon_1 = \varepsilon_2 = 1, \varepsilon_3 = 1 - X/U$. The form (3.53) of $\boldsymbol{\varepsilon}$ applies for the real axes x, y, z, and this is diagonal with elements $1, 1, 1-X/U$. Thus for regions 0, I, I(a) the medium is uniaxial. This term is explained near the end of § 5.5. See also problem 4.7.

In region I it is convenient to treat $\mathscr{H}_y$ as the dependent variable. Let $z = z_1$ be the level where $X = |1-\mathrm{i}Z|$. Then assumption (3) gives $X = |1-\mathrm{i}Z|\exp\{\alpha(z-z_1)\}$. Further let $\chi = \arctan Z = -\arg U$. Then

$$X/U = \exp\{\alpha(z-z_1) + \mathrm{i}\chi\}. \tag{19.27}$$

and (19.22) gives

$$\frac{\mathrm{d}^2\mathscr{H}_y}{\mathrm{d}z^2}+k^2\left[1-\frac{S^2}{1-\exp\{\alpha(z-z_1)+\mathrm{i}\chi\}}\right]\mathscr{H}_y=0. \tag{19.28}$$

This is of the form considered by Epstein (1930) and discussed in §§ 15.13–15.17. The expression in brackets may be called the effective value of q^2 (15.114), with which it is identical provided that we take

$$\eta_1=C^2,\quad \eta_2=1,\quad \eta_3=0,\quad \sigma=1/\alpha,\quad B=-\alpha z_1+\mathrm{i}\chi-\mathrm{i}\pi. \tag{19.29}$$

Equations (15.115)–(15.117) can now be used to find a, b, c. Then (15.132), (15.133) are used to find the reflection and transmission coefficients for waves incident from below, and similar expressions (Heading and Whipple, 1952, (6.5), (6.6); or Budden, 1961a, (17.108), (17.109)) are used for waves incident from above.

In regions II and II(a) X is comparable with or greater than Y so that XUa is of order $1/Y$ and still negligible. There may be some level above region II where XUa is of order unity, but this requires that X is of order Y^2. Now $XYa\approx -X/Y$ and this is no longer negligible. Since X is now large the factor $S^2/(1-X/U)$ is negligible and equations (19.3) become

$$\begin{aligned} E_x'&=-\mathrm{i}\mathscr{H}_y,\quad \mathscr{H}_y'=-\mathrm{i}E_x-\frac{X}{Y}E_y,\\ E_y'&=\mathrm{i}\mathscr{H}_x,\quad \mathscr{H}_x'=\mathrm{i}C^2E_y-\frac{X}{Y}E_x. \end{aligned} \tag{19.30}$$

Elimination of $\mathscr{H}_x$, $\mathscr{H}_y$ now gives

$$E_x''+E_x=\mathrm{i}\frac{X}{Y}E_y,\quad E_y''+C^2E_y=-\mathrm{i}\frac{X}{Y}E_x. \tag{19.31}$$

These two equations are equivalent to a single fourth order equation. They were studied by Wilkes (1947) for the case where X is a linear function of z, and he showed that the solution can then be expressed as the sum of a number of contour integrals. They were solved by Heading and Whipple (1952) for the case where X is proportional to $\mathrm{e}^{\gamma z}$.

The method, in outline, was as follows. It should be compared with § 15.13. Let $z=z_2$ be the level where $X/Y=4\gamma^2/k^2$. Introduce the new independent variable $w=\exp\{2\gamma(z-z_2)\}$. It is analogous to the ζ in (15.110). Eliminate E_x from (19.31). This gives a fourth order equation for E_y with w as independent variable. It has a regular singularity at $w=0$; in this respect it resembles Bessel's equation. It is found that there are four independent series solutions in ascending powers of w. These are like hypergeometric series but more complicated, and they are called generalised hypergeometric series, (Heading, 1980b). Below region II, w is very small so that only the first term of each series need be retained. For each of the four solutions E_y there is

a corresponding expression for E_x. Now these solutions can be identified with waves in region I(a). Two of them have $E_y = O(w) \to 0$ and E_x given by (19.25) (19.26). The other two have $E_x = O(w) \to 0$. They are solutions of (19.23) and are upgoing and downgoing waves travelling as though in free space. The possible ratios of the amplitudes of these four solutions must now be found. For this it is necessary to find circuit relations. In region II(a) the waves must be upgoing and there are two independent solutions with this property. Each of these can be expressed as a contour integral of Barnes type; see § 15.13 and (15.107) (Barnes, 1908). These two integrals are such that when w is large they can be evaluated by steepest descents (ch. 9) and this gives the two W.K.B. solutions for the upgoing waves. Then in region I(a) each Barnes integral is expressed as four series of the residues of poles of the integrand, and these are the generalised hypergeometric series already mentioned. This process gives the four elements of the reflection coefficient matrix of region II as measured below it in region I(a), with reference level $z = z_2$. They are very complicated expressions involving factorial functions containing γ, k and C.

The final step is to combine these reflection coefficients of region II with the reflection and transmission coefficients for region I. In region I(a) there are both upgoing and downgoing waves, whose relative phases must be allowed for. This depends on the height range $z_2 - z_1$, between the reference levels for regions II and I. The details and the final result are given by Heading and Whipple (1952, § 10). See also Budden (1961a, § 21.12).

Curves were given by Heading and Whipple to show how the overall reflection coefficients and the separate reflection coefficients for regions I and II depend on angle of incidence, for frequencies 16 kHz and 80 kHz. The formulae for the overall reflection coefficients were used by Budden (1955b) who showed that they agreed well with computer integrations of the differential equations. In the paper by Budden (1955b) there were errors in the signs of l_x and l_y so that results given for east to west propagation apply for west to east, and vice versa.

19.4. Resonance and barriers

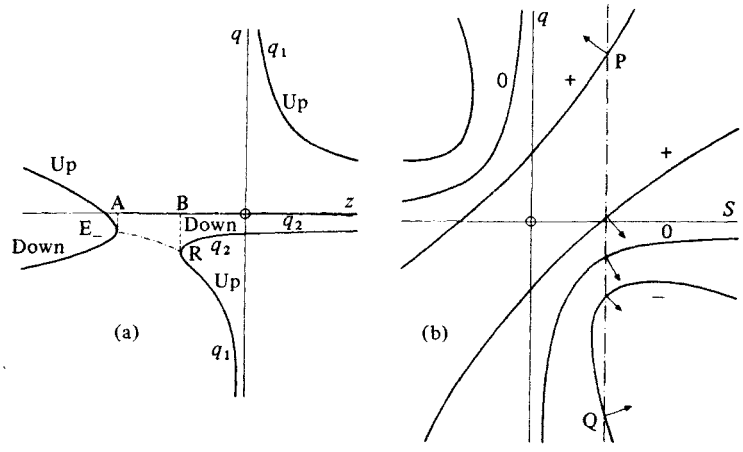
One root q of the Booker quartic (6.15) is infinite when the coefficient α of q^4 is zero, and we need to study the effect of this infinity on wave propagation. At first collisions will be neglected, so that α is given by (6.24) and is zero for one real value of X satisfying (4.76) with $U = 1$. In this section the q that is infinite is denoted by q_1. It is assumed that the zero of α is simple and at some real height, since this is the most important case in practice. The origin $z = 0$ is chosen at this zero. For the lower side of an ionospheric layer X increases and α decreases when z increases, so that α is negative where $z > 0$. For no collisions the coefficients $\alpha, \beta, \ldots$ of the quartic are all real, and near the infinity

$$q_1 \approx -\beta/\alpha \tag{19.32}$$

so that q_1 is real on both sides of $z = 0$, when z is real. In this respect it differs from the infinity studied in §§ 15.5–15.7 where Q, the effective q, is purely imaginary when collisions are neglected. The behaviour of q near an infinity can be seen in figs. 6.4, 6.5, 6.7 and in figures given by Booker (1939), Smith, M.S. (1974a). Fig. 19.1(a) is a sketch of the typical behaviour. It would apply for the extraordinary wave in the Z-mode when $Y < 1$. It often happens that near the infinity there is a coupling point, $z = b < 0$ at B in fig. 19.1(a), where q_1 is equal to another root say q_2. An example is the reflection point R in fig. 6.5. Then for $z < b$, q_1 and q_2 are complex conjugates. There may also be another coupling point, $z = a < b$ at A in fig. 19.1(a), where the same two roots are again equal, so that they are again real when $z < a$; an example is the reflection point E_ of fig. 6.5. Then the range $a < z < b$ is similar to a conventional barrier because the waves here are inhomogeneous, but its properties are modified by the proximity of the resonance. The real values of q can be found by Poeverlein's construction, § 10.8, with refractive index surfaces for a sequence of values of z. Fig. 19.1(b) (compare fig. 10.7) is a sketch of the form these surfaces would take to give the two right-hand curves of fig. 19.1(a). The arrows indicate the directions of the ray. This shows that both the branches marked q_1 in fig. 19.1(a) refer to upgoing waves.

Consider now the special case of vertical incidence, so that the coefficient β of the quartic (6.15) is zero. Then, where $\alpha = 0$, two roots of the quartic are infinite. The reflection point R, fig. 6.5, has moved to coincidence with the infinity, so that $b = 0$, in fig. 19.1(a). This is therefore a degenerate case where a reflection point and a

Fig. 19.1. (a) is an example of how q may depend on $s = kz$ in and near a barrier when collisions are neglected. The chain curve shows $\mathrm{Re}(q_1) = \mathrm{Re}(q_2)$ in the barrier region where $(q_1 - q_2)^2$ is negative. (b) shows a family of refractive index surfaces for various values of the height z. The symbols $+$, 0, $-$ by the curves refer to the values of z. The broken line is used for Poeverlein's construction.



resonance point coincide. The qs are now the same as the refractive indices n, and the quartic is just the quadratic equation (4.65) for n^2. It is now convenient to plot n^2 against z, and the form of the curve is sketched in fig. 19.2(b). There is still a barrier with $n^2 < 0$, where $a \leqslant z \leqslant 0$. At its left-hand boundary $n^2 = 0$ as with a conventional barrier, but at the other boundary $n^2 \to \infty$. The case where $n^2 \to \infty$ at both boundaries of a barrier can occur and some aspects of it have been studied (Tang, Wong and Caron, 1975) but it is not considered here.

The objective now is to study the effect of the infinity of q_1 on the tunnelling through the barrier. A full theory that covers most of the important cases was given by Budden (1979). It is very long and only a few of the examples are given here. It is found that, by making approximations, the differential equations can be separated so that the tunnelling phenomenon is studied by using differential equations of only second order. Thus this theory is a continuation of the topics in ch. 15.

A point of resonance is a singular point of the governing differential equations. This was discussed in § 18.13 where it was explained that a branch cut must be inserted in the complex z plane running from the resonance point to infinity without crossing the real z axis. If there is a very small collision frequency, the resonance occurs where $\mathrm{Im}(X) < 0$. The case discussed here is for the lower ionosphere where X increases with increasing z as in fig. 19.1(a), so the resonance is where $\mathrm{Im}(z) < 0$. The branch cut runs to infinity in the negative imaginary half of the z plane. In the limit of no collisions this branch cut must be retained, and it now runs from $z = 0$ to $|z| \to \infty$ with $\mathrm{Im}(z) < 0$.

19.5. Isolated resonance

Suppose now that the incidence is oblique and that X increases very slowly with height z so that the reflection point at $z = b$ is at a great distance from the resonance. Then there is a large domain of the z plane, containing $z = 0$, where the solution q_1 of the quartic refers to a wave for which there are no reflection or coupling points. To study this wave we now examine the coupled equations (7.109) or (16.7). In the domain, (19.32) gives

$$q_1 \approx \Delta/s, \quad s = kz \tag{19.33}$$

where Δ is real and large. We take Δ to be positive as in fig. 19.1(a). If there are coupling points not involving q_1 in the domain, this can be easily dealt with. For simplicity it will be assumed that there are none. At the resonance some elements of $\mathbf{T}$ in (7.80), (7.81) have simple poles because of the factor ε_{zz}, proportional to α, in the denominator. But $\varepsilon_{zz}\mathbf{T}$ (7.82) is analytic and bounded at and near the resonance, and therefore its eigen values $\varepsilon_{zz}q$ and eigen columns are bounded and analytic. The eigen columns of $\varepsilon_{zz}\mathbf{T}$ are also eigen columns of $\mathbf{T}$. The matrix $\mathbf{S}$ formed from the four eigen columns is therefore bounded, non-singular and analytic. The coupling matrix

$\mathbf{\Gamma} = -\mathbf{S}^{-1}\mathbf{S}'$ (7.108) exists throughout the domain. It is very small because of the derivative $\mathbf{S}'$, since the medium is very slowly varying and there are no coupling points in the domain. If $\mathbf{\Gamma}$ is neglected, the first of the four coupled equations (7.109) is

$$\mathrm{f}_1' = -\mathrm{i}\mathrm{f}_1\Delta/s \tag{19.34}$$

with the solution

$$\mathrm{f}_1 = P\exp(-\mathrm{i}\Delta\ln s) \tag{19.35}$$

where P is a constant. The other three of the four coupled equations are associated with the remaining three values of q. These are bounded, and the solutions are not affected by the resonance. These three waves are propagated independently and we may assume that they are absent, so that $\mathrm{f}_2 = \mathrm{f}_3 = \mathrm{f}_4 = 0$.

Since Δ is positive, the phase propagation of the wave (19.35) is upwards when s is real and positive and downwards when s is real and negative. But in fig. 19.1(b) the corresponding points are P, Q respectively and the arrows show that the ray, that is the direction of group propagation, is obliquely upwards in both cases. They also show that the ray velocities have large horizontal components in opposite directions for positive and negative s.

Now the vertical component Π_z of the Poynting vector is to be found for real s above and below the resonance. For this we use some of the matrix theory of § 7.14. The following version is a summary. For the full theory see Budden (1979, §§ 3, 9).

The adjoint $\bar{\mathbf{e}}$ of the column $\mathbf{e}$ of the field variables was introduced in § 7.14(4). It satisfies the adjoint differential equation

$$\bar{\mathbf{e}}' = \mathrm{i}\bar{\mathbf{T}}\bar{\mathbf{e}} \tag{19.36}$$

from (7.95). Now from § 7.14(2) the eigen columns of the adjoint $\bar{\mathbf{T}}$ are the rows of $\mathbf{S}^{-1}$ with the order of its columns reversed, that is the rows of $\mathbf{S}^{-1}\mathbf{B}$ which are the same as the columns of $\mathbf{B}(\mathbf{S}^{\mathrm{T}})^{-1}$, where $\mathbf{B}$ is given by (7.92). The first equation (6.53) is $\mathbf{e} = \mathbf{S}\mathbf{f}$ and its adjoint is

$$\bar{\mathbf{e}} = \mathbf{B}(\mathbf{S}^{\mathrm{T}})^{-1}\bar{\mathbf{f}} \quad \text{whence} \quad \bar{\mathbf{e}}^{\mathrm{T}} = \bar{\mathbf{f}}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{B}. \tag{19.37}$$

The adjoint of the coupled equations may now be found. The coupling matrix $\bar{\mathbf{\Gamma}}$, like $\mathbf{\Gamma}$, is negligible. The adjoint of (19.34) is then, from (19.36)

$$\bar{\mathrm{f}}_1' = \mathrm{i}\bar{\mathrm{f}}_1\Delta/s \tag{19.38}$$

with the solution

$$\bar{\mathrm{f}}_1 = \bar{P}\exp(\mathrm{i}\Delta\ln s). \tag{19.39}$$

The bilinear concomitant W was given in 7.14(5), and if (19.37) is used, (7.97) shows that

$$W = \bar{\mathbf{e}}^{\mathrm{T}}\mathbf{B}\mathbf{e} = \bar{\mathbf{f}}^{\mathrm{T}}\mathbf{f} = \bar{\mathrm{f}}_1\mathrm{f}_1 = \bar{P}P. \tag{19.40}$$

The solution (19.39) contains the arbitrary constant $\bar{P}$. It was shown in § 7.14(8) that for any loss-free medium where s is real, $\bar{P}$ can be chosen so that $W = 4Z_0\Pi_z$. We apply this where s is real and positive, indicated by the + sign, so that

$$W = 4Z_0\Pi_z(+) \tag{19.41}$$

from (7.104). Now Π_z must also be real and positive so we take $\bar{P} = P^*$, $4Z_0\Pi_z = PP^*$, and then $\mathsf{f}_1, \bar{\mathsf{f}}_1$ are complex conjugates. This is true only when s is real and positive. If $\mathrm{Im}(s) \neq 0$, then $W \neq 4Z_0\Pi_z$ and $\bar{\mathsf{f}}_1 \neq \mathsf{f}_1^*$. Now from (7.87) W is independent of z, that is of s, and is therefore still given by (19.40) where s is real and negative. In going from positive to negative real s, the branch cut must not be crossed so the path is where $\mathrm{Im}(s) > 0$, and $\arg s$ increases from 0 to π. Then for s real and negative, from (19.35), (19.39)

$$\mathsf{f}_1 = P|s|^{-\mathrm{i}\Delta}\mathrm{e}^{\pi\Delta}, \quad \bar{\mathsf{f}}_1 = P^*|s|^{\mathrm{i}\Delta}\mathrm{e}^{-\pi\Delta}. \tag{19.42}$$

The product (19.40) of these is unchanged and W is the same as for positive s. But (19.42) are not now complex conjugates, and (19.41) is not now true.

To find $\Pi_z(-)$ where s is real and negative the product $\mathsf{f}_1\bar{\mathsf{f}}_1$ in (19.40) must be replaced by $\mathsf{f}_1\mathsf{f}_1^*$. This means that the arbitrary constant $\bar{P}$ in (19.39) must now have a new value $P^*\mathrm{e}^{2\pi\Delta}$, and then

$$\Pi_z(-) = \mathrm{e}^{2\pi\Delta}PP^* = \mathrm{e}^{2\pi\Delta}\Pi_z(+). \tag{19.43}$$

The upward energy flux below the resonance level is greater than that above it by a factor $\mathrm{e}^{2\pi\Delta}$. There must be a disappearance of energy at the resonance, in the loss-free case, if the solution (19.35), with a singularity at $s = 0$, is present. Another example of this is given in the following section.

The question as to what happens to the wave energy that disappears near a resonance is an intriguing physical problem. At least four different view points may be taken.

First it may be argued that the assumption of a loss-free medium is unrealistic. There is no medium, except for a vacuum, for which all forms of wave attenuation are absent. Therefore the refractive index n always has a non-zero imaginary part when $s = kz$ is real. In a plasma this could be caused by collision damping, Landau damping or other forms of damping. $\mathrm{Im}(n)$ may be negligibly small for most real values of z. But the pole at the resonance is not exactly on the real z axis. Near it both real and imaginary parts of n are very large, so that the energy absorption rate is large. The lost energy is simply converted to heat or other forms of energy in the region of the real z axis close to the pole. In this book non-linear effects have been ignored. The solutions for E_x, E_y, proportional to (19.35), are bounded near the resonance, so that for these components, the assumption of linearity may be justified for small wave amplitudes. But it can be shown from the last equation of (7.78) that E_z is not bounded at the resonance. Thus non-linear effects must occur. One result

would be the generation of harmonics of the wave frequency. This provides another way by which energy could be lost.

Second, the illustrations in this book apply for a cold plasma whose dielectric constant is (3.55). If the plasma is warm its dielectric constant is not given exactly by (3.33). The result is that when ε_{zz} approaches zero, the solution q_1 gets very large, but it does not become infinite when $\varepsilon_{zz} = 0$. Instead, the characteristics of the wave go over continuously to those of a plasma wave. Curves showing how this happens have been given by Ginzburg (1970, figs. 12.2–12.4, 12.9–12.13). See also Golant and Piliya (1972). On this view the resonance at $z = 0$ does not exist. The waves have simply gone over to a mode of propagation governed by a modification of the basic equation, and they carry the 'lost' energy with them. Some authors call this 'mode conversion'. But the transition is continuous and it is only the name of the wave type that changes. If the medium is extremely slowly varying there is no coupling point near $z = 0$ and the transition is not a mode conversion. In an actual medium there might be some mode conversion near the transition, but this would be a separate phenomenon associated with a coupling point.

Third, for the theory of this chapter the incident wave is a plane wave and therefore extends indefinitely in the x and y directions. Any actual incident wave must be a laterally bounded beam. It was shown earlier in this section, from fig. 19.1(b), that the ray direction, that is the direction of energy propagation, gets more nearly horizontal as the resonance level is approached. Thus the upgoing energy is deviated sideways right out of the beam.

If any of these views is taken, the lost energy can be accounted for and there is nothing more to be said. But it is still of interest to pursue the enquiry for a fictitious medium which is a cold plasma and really has no damping mechanism at all. The following suggestion is offered. In any wave system some energy is stored in the medium. The time average of the stored energy per unit volume is constant in the steady state conditions assumed here, and is proportional to the dielectric constant and to the square of the electric field. Near a resonance at $z = 0$ the dielectric constant (3.55) is bounded, but E_z tends to infinity like $1/z$. Consequently if the stored energy is integrated through a volume near the resonance, the integral gets indefinitely large as the edge of the volume moves up to the resonance. The total energy stored in the part of the medium containing the resonance is infinite. But this state of affairs cannot be attained in a finite time. After the wave is first switched on, a state of physical equilibrium is never reached. The medium has an infinite capacity for storing energy. The energy that is apparently lost is simply going into storage in the medium.

19.6. Resonance tunnelling

The effect of a resonance on the penetration of a barrier is now to be studied. The theory is given here only for vertical incidence so that the reflection point R of fig.

19.1(a) coincides with the resonance, which is one boundary of the barrier. The full theory allowing for oblique incidence is given by Budden (1979). It is similar in form and leads to differential equations of the same type, but the algebra is more complicated. The theory is also of interest in the realm of thermo-nuclear plasmas and this aspect has been studied, for example, by White and Chen (1974) and Batchelor (1980).

When a vertically incident radio packet approaches a cut-off level where $n = 0$, and $X = X_0$, the vertical component $\mathscr{U}_z$ of the group velocity gets smaller. It is proportional to $(X_0 - X)^{\frac{1}{2}}$ from (5.78), (5.79). The wave packet slows down, reverses its direction and is reflected. Although $\mathscr{U}_z \to 0$ when $X \to X_0$, the group time of travel is bounded. It affects the calculation of equivalent height from the integral (13.8). When the wave packet approaches a resonance level where $n \to \infty$, $X = X_\infty$, again $\mathscr{U}_z$ gets smaller but it is now proportional to $(X - X_\infty)^{\frac{3}{2}}$ as shown by (5.82). The wave packet slows down, but because of the power 3/2, it would take an infinite time to reach the resonance level, and this crude ray type theory suggests that it cannot be reflected. It has sometimes been suggested that reflection from a resonance can occur, and the expression for $X = X_\infty$ has been called 'the fourth reflection condition' (the other three are $X = 1, 1 \pm Y$). The full wave theory that follows confirms that there is no reflection at a resonance. In those cases where reflection has been thought to occur at the level where $X = X_\infty$, it must have resulted from some other cause such as a nearby zero of n or a steep gradient of $N(z)$.

When an extraordinary wave travels vertically upwards into a collisionless ionosphere with $Y < 1$, it first encounters a cut-off level where $X = 1 - Y$ and here it is reflected. Above this is a barrier region where n_E^2 is negative. It extends up to where $X = X_\infty$ given by (4.76) with $U = 1$. Here n_E^2 is infinite, and above this it is positive and decreasing. This case will be discussed here and the subscript E will be omitted. It will be shown that for the upgoing wave there can be some penetration of the barrier, (19.63) below. Most of the energy is reflected at the cut-off level but some energy apparently disappears. For a wave coming down from above the resonance there is no reflection. Part of the wave energy penetrates the barrier and the rest apparently disappears; (19.72) below.

For vertical incidence Försterling's coupled equations (16.90), (16.91) can be used. The coupling parameter ψ depends on $\mathrm{d}N(z)/\mathrm{d}z$ and is small in a slowly varying medium, except possibly near the levels of the coupling points, that is $X \approx 1$, but this is not near the resonance. We are not concerned with coupling and therefore ψ will be neglected. The two Försterling equations are then independent, and (16.91) for the extraordinary wave will be written, without the subscripts, thus

$$\frac{\mathrm{d}^2\mathscr{F}}{\mathrm{d}s^2} + n^2\mathscr{F} = 0, \quad s = kz. \tag{19.44}$$

This is similar to (7.6) for an isotropic medium and may be discussed in a similar way.

Approximate solutions are the two W.K.B. solutions

$$\mathscr{F} = n^{-\frac{1}{2}} \exp\left\{ \mp \mathrm{i} \int^{s} n \, \mathrm{d}s \right\} \tag{19.45}$$

and the approximation is good provided that (7.59), with q replaced by n, is satisfied.

In a slowly varying ionosphere it may be assumed that, for a small range of s, X varies linearly with height. The origin of s is chosen at the resonance where $X = X_\infty$ so that

$$X - X_\infty = \alpha s \tag{19.46}$$

where α is a constant. Then it can be shown from (4.67) that, for the n^2 that has the infinity,

$$n^2 = \frac{A}{\alpha s} + B + \mathrm{O}(s) \tag{19.47}$$

where A and B are real positive constants expressible in terms of Y and Θ. If $|s|$ is small all terms except the first are negligible and substitution into (7.59) gives

$$|s| \gg 3\alpha/8A \tag{19.48}$$

as the condition for the validity of the W.K.B. solutions (19.45). The condition fails near the resonance at $s = 0$ and a full wave solution of the differential equation must be sought. The solutions for two important cases will now be given.

Assume first that α is so small that only the first term of (19.47) need be retained for values of $|s|$ up to that for which (19.48) becomes valid. Then the differential equation (19.44) is

$$\frac{\mathrm{d}^2 \mathscr{F}}{\mathrm{d}s^2} + \frac{\beta}{s} \mathscr{F} = 0 \tag{19.49}$$

where β is a real positive constant. The squared refractive index is shown in fig. 19.2(a) (continuous curve). A solution is sought that gives the reflection coefficient for waves incident from the right. Thus when s is large and negative the solution must represent a wave travelling to the left or a field whose amplitude decreases as s gets more negative. If there is a small constant collision frequency, the infinity of n^2 occurs where $X = X_\infty$, and therefore also s, has a small negative imaginary part. The zero of s is then chosen to be where $X = \mathrm{Re}(X_\infty)$, and (19.49) becomes

$$\frac{\mathrm{d}^2 \mathscr{F}}{\mathrm{d}s^2} + \frac{\beta}{s + \mathrm{i}\gamma} \mathscr{F} = 0 \tag{19.50}$$

where γ is real and positive. The real part of n^2 is then as shown by the broken curve in fig. 19.2(a). Equations (19.49), (19.50) are of standard form (Watson, 1944, p. 97). A solution of (19.50) is

$$\mathscr{F} = (s + \mathrm{i}\gamma)^{\frac{1}{2}} \mathscr{C}\{2\beta^{\frac{1}{2}}(s + \mathrm{i}\gamma)^{\frac{1}{2}}\} \tag{19.51}$$

where $\mathscr{C}$ denotes a Bessel function. If, for real s, the sign of the square root is chosen

so that

$$0 < \arg(s + i\gamma)^{\frac{1}{2}} < \pi \tag{19.52}$$

then the Bessel function that satisfies the physical conditions is $\mathscr{C} = H_1^{(1)}$, a Bessel function of the third kind, or Hankel function. When $|s|$ is large, the asymptotic formulae for this function give the following results (Watson, 1944, p. 197). For s real and negative

$$\mathscr{F} \sim (s + i\gamma)^{\frac{1}{4}} \exp\left[2\beta^{\frac{1}{2}}\{-|s|^{\frac{1}{2}} + \tfrac{1}{2}i\gamma|s|^{-\frac{1}{2}}\}\right] \tag{19.53}$$

and for s real and positive

$$\mathscr{F} \sim (s + i\gamma)^{\frac{1}{4}} \exp\{2\beta^{\frac{1}{2}} i|s|^{\frac{1}{2}}\}. \tag{19.54}$$

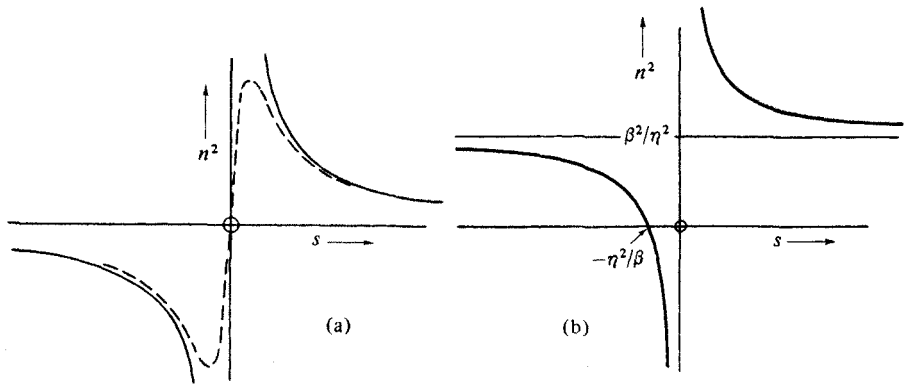
Here (19.54) represents a wave travelling to the left in fig. 19.2(a). This is the incident wave and may be identified with one of the W.K.B. solutions (19.45). There is no other term and therefore no reflected wave. The result applies in the limit $\gamma \to 0$ when the collision frequency is zero. The wave (19.53) is then evanescent and of negligible amplitude when s is large and negative. The energy in the incident wave has apparently disappeared. See end of § 19.5 for possible explanations.

To study the effect of the resonance on the penetration of the barrier, suppose now that B in (19.47) is comparable with $A/\alpha s$ when s has the smallest value that satisfies (19.48). Then the zero and infinity of n^2, (19.47), are so close together that we must seek a full wave solution that embraces both the zero and the infinity. If the electron collision frequency is zero, the differential equation (19.44) is now

$$\frac{d^2\mathscr{F}}{ds^2} + \left(\frac{\beta}{s} + \frac{\beta^2}{\eta^2}\right)\mathscr{F} = 0 \tag{19.55}$$

where β and η are constants. Fig. 19.2(b) shows how n^2 depends on s. It is zero where

Fig. 19.2. Height dependence of the squared refractive index n^2 near a resonance. In (a) the resonance is 'isolated'. The full curve is for $Z = 0$ (no collisions), and the broken curve is $\mathrm{Re}(n^2)$ when Z is small and non-zero. In (b) $Z = 0$ and there is a zero of n^2 close to the resonance. In both figures the abscissa is $s = kz$ where z is height.



$s = -\eta^2/\beta$ and infinite where $s = 0$. If there is a small constant collision frequency, s in the second term of (19.55) is replaced by $s + i\gamma$ as in (19.50), where γ is real and positive. Then the pole and zero of n^2 both lie where Im(s) is small and negative. The pole is a singular point of the differential equation and there must therefore be a branch cut running from it to infinity in the negative imaginary half of the s plane. If $\gamma = 0$, this cut runs from $s = 0$.

Let

$$\zeta = 2i\beta s/\eta. \tag{19.56}$$

Then (19.55) becomes

$$\frac{d^2\mathscr{F}}{d\zeta^2} + \left(-\tfrac{1}{4} - \frac{\frac{1}{2}i\eta}{\zeta}\right)\mathscr{F} = 0 \tag{19.57}$$

which is the form given by Whittaker and Watson (1927, p. 337) for the confluent hypergeometric function..

Case (i). Wave incident from below

One solution of (19.57) is

$$\mathscr{F} = W_{k,m}(\zeta) \tag{19.58}$$

where $m = \pm\frac{1}{2}$, $k = -\frac{1}{2}i\eta$. When s is real, positive and large, arg $s = 0$, arg $\zeta = \frac{1}{2}\pi$ and (Whittaker and Watson, 1927, p. 343):

$$\mathscr{F} \sim e^{-\frac{1}{2}\zeta}\zeta^k = \exp\{-is\beta/\eta - \tfrac{1}{2}i\eta\ln s - \tfrac{1}{2}i\eta\ln(2i\beta/\eta)\}. \tag{19.59}$$

This represents an upgoing wave above the resonance level. The solution (19.58) may therefore be used to find the reflection and transmission coefficients for a wave incident from below. To do this it is necessary to find the asymptotic form of (19.58) when s is real and negative. Here arg $s = \pi$, because in going from positive to negative s the branch cut, where Im(s) is negative, must not be crossed. Hence arg $\zeta = \frac{3}{2}\pi$, and the asymptotic form for (19.58), given by Heading (1962b), is

$$W_{k,m}(\zeta) \sim \zeta^k \exp(-\tfrac{1}{2}\zeta) + \frac{2\pi i \exp(2\pi i k)}{(-\frac{1}{2} - m - k)!\,(-\frac{1}{2} + m - k)!}\zeta^{-k}\exp(\tfrac{1}{2}\zeta) \tag{19.60}$$

for $\pi \leqslant \arg\zeta \leqslant 2\pi$. For the special values $m = \pm\frac{1}{2}$, $k = -\frac{1}{2}\eta$ this becomes

$$\exp\{-is\beta/\eta - \tfrac{1}{2}i\eta\ln|s| + \tfrac{1}{2}\pi\eta - \tfrac{1}{2}i\eta\ln(2i\beta/\eta)\}$$
$$+ \frac{2\pi i\exp(\pi\eta)}{(\frac{1}{2}i\eta)!\,(-1 + \frac{1}{2}i\eta)!}\exp\{is\beta/\eta + \tfrac{1}{2}i\eta\ln|s| - \tfrac{1}{2}\pi\eta + \tfrac{1}{2}i\eta\ln(2i\beta/\eta)\}. \tag{19.61}$$

The first term represents an upgoing wave, the incident wave, and the second a downgoing wave, the reflected wave. The ratio of the moduli of these terms gives the modulus of the reflection coefficient R:

$$|R| = \frac{2\pi\exp(-\frac{1}{2}\pi\eta)}{|(\frac{1}{2}i\eta)!\,(-1 + \frac{1}{2}i\eta)!|} = 1 - e^{-\pi\eta}. \tag{19.62}$$

Similarly the modulus of the transmission coefficient T is given, from (19.54), (19.61), by

$$|T| = \mathrm{e}^{-\frac{1}{2}\pi\eta}. \tag{19.63}$$

Case (ii). Wave incident from above

Instead of using (19.56), let

$$\xi = -2\mathrm{i}\beta s/\eta$$

so that (19.55) is

$$\frac{\mathrm{d}^2\mathscr{F}}{\mathrm{d}\xi^2} + \left(-\tfrac{1}{4} + \frac{\frac{1}{2}\mathrm{i}\eta}{\xi}\right)\mathscr{F} = 0. \tag{19.64}$$

One solution is

$$\mathscr{F} = W_{k,m}(\xi) \tag{19.65}$$

where $m = \pm\frac{1}{2}$, $k = +\frac{1}{2}\mathrm{i}\eta$. When s is real and positive $\arg\xi = -\frac{1}{2}\pi$ and for large $|s|$ (Heading, 1962b or Whittaker and Watson, 1927, p. 343):

$$\mathscr{F} \sim \mathrm{e}^{-\frac{1}{2}\xi}\xi^k = \exp\{\mathrm{i}s\beta/\eta + \tfrac{1}{2}\mathrm{i}\eta\ln s + \tfrac{1}{2}\mathrm{i}\eta\ln(-2\mathrm{i}\beta/\eta)\}. \tag{19.66}$$

This represents a downgoing wave from above, the incident wave. There is no upgoing wave and so the reflection coefficient is zero. When s is real and negative $\arg\xi = \frac{1}{2}\pi$. This is still within the range for which the expression (19.66) is valid, and for large $|s|$

$$\mathscr{F} \sim \exp\{\mathrm{i}s\beta/\eta + \tfrac{1}{2}\mathrm{i}n\ln|s| - \tfrac{1}{2}\pi\eta + \tfrac{1}{2}\mathrm{i}\eta\ln(-2\mathrm{i}\beta/\eta)\}. \tag{19.67}$$

This represents a downgoing wave at the bottom, the transmitted wave. The modulus of the transmission coefficient T is

$$|T| = \mathrm{e}^{-\frac{1}{2}\pi\eta}. \tag{19.68}$$

The R and T used here are ratios of the values of $\mathscr{F}$ as given by (16.89). We are considering the extraordinary wave whose polarisation is $\rho_\mathrm{E} = 1/\rho_\mathrm{O}$, so that $E_x = \rho_\mathrm{O}E_y$. For large $|s|$ the refractive index n is real and ρ_O is purely imaginary. Hence

$$\mathscr{F} = (\rho_\mathrm{O}^2 - 1)^{\frac{1}{2}}E_y, \quad \mathscr{F}\mathscr{F}^* = (1 - \rho_\mathrm{O}^2)E_yE_y^*. \tag{19.69}$$

Where the W.K.B. solutions are good approximations, (16.72) shows that $\mathscr{H}_x = -nE_y$, $\mathscr{H}_y = nE_x$. Hence from (2.64)

$$4Z_0\Pi_z = 2n(E_xE_x^* + E_yE_y^*) = 2n\mathscr{F}\mathscr{F}^*. \tag{19.70}$$

At very large s, for each W.K.B. solution, the vertical energy flux Π_z is independent of s. For $|s| \to \infty$ the two values of n are the same for positive and negative s. It follows that $|R|^2$ and $|T|^2$ are the ratios of the energy fluxes.

The results of this section may now be summarised as follows. For a wave incident from below

$$|R| = 1 - \mathrm{e}^{-\pi\eta}, \quad |T| = \mathrm{e}^{-\frac{1}{2}\pi\eta}, \quad |R^2| + |T^2| = 1 - \mathrm{e}^{-\pi\eta} + \mathrm{e}^{-2\pi\eta} < 1. \tag{19.71}$$

For a wave incident from above

$$|R| = 0, \quad |T| = e^{-\frac{1}{2}\pi\eta}, \quad |R^2| + |T^2| = e^{-\pi\eta} < 1. \tag{19.72}$$

In both cases the energy in the incident wave does not all reappear in the reflected and transmitted waves. Yet these results apply for the loss-free medium where there is no mechanism for absorbing energy. Possible explanations were discussed at the end of §19.5.

In this section it was assumed that near the resonance the height dependence of X was linear (19.46). An alternative method is to assume that X depends on an exponential function of height, as in (15.110), (15.114). Then instead of giving the confluent hypergeometric equation (19.57), (19.64), the equations are transformed to give the hypergeometric differential equation (15.97). Instead of a single pole and zero on the real axis, as in (19.55), there is an infinite number of each, on lines parallel to the imaginary s axis. This method was used by van Duin and Sluijter (1983). Their models had more assignable parameters than the one used here, so they were able to study more complicated cases. For example the squared refractive index n^2 could tend to different values, including negative values, above and below the resonance. One interesting result of this work is that for an incident wave approaching a resonance without first encountering a zero, the reflection coefficient does not now agree with (19.72) but has a small non-zero value.

19.7. Inversion of ionospheric reflection measurements

One of the main objects of using radio waves to probe the ionosphere from the ground is to find the height distribution $N(z)$ of electron concentration, and possibly also the electron collision frequency $\nu(z)$. At high frequencies, greater than about 1 MHz, this is done, for $N(z)$, by measuring the equivalent height function $h'(f)$ and using it to compute $N(z)$. Techniques for this were described in §§ 12.6, 13.6. This is an example of inversion of the data to give the required function. The method, however, gives very little information about the lowest parts of the ionosphere, that is the D-region. Here the refractive indices for the high frequency waves are close to unity and it has little effect on the function $h'(f)$. But radio waves in the very low frequency range, of order 10 to 100 kHz are reflected in the D-region and can be used to investigate its structure. Although radio pulses at frequencies as low as 50 kHz have been used (Watts 1957), the method is difficult and expensive because very large transmitting aerials are needed. The usual method is to measure the signals received from commercial transmitters. These emit morse or telex signals which can be received and measured just as if they were continuous unmodulated waves. They nearly always use vertical electric dipole aerials so that the waves are vertically polarised and the elements R_{11}, R_{21} of the reflection coefficient matrix $\boldsymbol{R}$ are

measured. The dependence of their amplitude and phase on frequency, time of day and season, and distance from transmitter to receiver has been thoroughly studied (Bracewell *et al.*, 1951; Bracewell, 1952; Weekes and Stuart, 1952a; Bain *et al.*, 1952; Bracewell *et al.*, 1954; Straker, 1955; Belrose, 1968). By using a mobile receiver in a vehicle or aircraft, the dependence on angle of incidence has been measured (Hollingworth, 1926; Weekes, 1950; Bracewell *et al.*, 1951; Weekes and Stuart, 1952b). In some measurements, transmitting aerials that emit horizontally polarised waves have been constructed by laying aerial wires on poorly conducting ground such as desert soil, or ice, so that measurements of R_{12}, R_{22} were also possible (Macmillan, Rusch and Golden, 1960; Raghuram, Smith and Bell, 1974).

The effect on the calculated $\boldsymbol{R}$ of making changes in a given ionospheric model, that is in $N(z)$ and $\nu(z)$, was studied by Piggott, Pitteway and Thrane (1965). They were thus able to suggest modifications that might be made to improve the agreement with observations.

There is no direct method of using results of observations to calculate $N(z)$ and $\nu(z)$. One approach, used very successfully by Deeks (1966a, b) is a systematic application of trial and error. The data he used came from observations with many different frequencies and transmitter sites, and many receiving sites including mobile receivers. His method was to set up a computer program for calculating $\boldsymbol{R}$ for given functions $N(z)$, $\nu(z)$. Then by making a large number of trials with modifications made to different sections of the $N(z)$ curve, and with various functions $\nu(z)$, he was able to find those functions $N(z)$, $\nu(z)$ that best fit the observations. He concluded that if $N(z)$, at most heights z, is altered by 20%, the agreement with observations would be upset. Others have used the same method. Bain and May (1967) had more observations available, and suggested some modifications to Deeks' results. Thomas and Harrison (1970) applied the method for the D-region at night. The functions $N(z)$ and $\nu(z)$ deduced by Deeks' method have been widely accepted and used, and, as far as the author knows, no better results have since been published.

A general theory of the technique of inverting data to give the required results directly, without a trial and error process, has been given by Backus and Gilbert (1967, 1968, 1970). It was intended for application in the field of solid earth geophysics but the principle is equally applicable for the radio ionosphere problem at very low frequencies. The mathematical techniques that it uses are fairly advanced and include some functional analysis. A summary in the context of the radio observations was given by Mobbs (1978) in an unpublished thesis. The author is indebted to Dr A.J. Mobbs and Professor T.B. Jones of Leicester University for the opportunity of studying this work. The full general technique has not yet been brought to the stage where it can be applied more simply than the method of Deeks, and further work in this field would be of the greatest interest. Some simpler

techniques have, however, been used, for example by Baybulatov and Krasnushkin (1966), Shellman (1970), Bailey and Jones (1974), Mambo, Nagano, Nakamura and Fukami (1983), Field, Warren and Warber (1983).

A very brief survey follows of the kinds of problem that have to be tackled in this study. We suppose that it is the electron height distribution $N(z)$ that is to be found, but the argument could be extended to include $\nu(z)$ also. Although $N(z)$ is a continuous function, it must be specified by a finite set of numbers. These might, for example, be tabulated values of N, or possibly $\ln N$, at equal intervals of height z, with the assumption that linear or higher order interpolation is used. When $N(z)$ is mentioned below it is this set of numbers that is referred to. They are to be found from a given set of data, such as $|R_{11}|$, $\arg R_{11}$, $|R_{21}|$, $\arg R_{21}$, etc. for various frequencies, angles of incidence, azimuths of propagation, etc. The following are some of the factors that have to be taken into account.

(a) All items of data are subject to error and those with least error must be given the greatest weight. The errors of different items may not be independent. In other words some items of data may be correlated and allowance can be made for this if the degree of correlation is known. For example, at 16 kHz, $|R_{11}|$ and $|R_{21}|$ may both vary from day to day but their ratio is found to be nearly constant.

(b) A given set of data is not enough to determine $N(z)$ uniquely. We try to find an $N(z)$ that minimises the difference between the calculated and observed values of the data.

(c) Some values of $N(z)$ may be 'insignificant' for some items of data. For example the values of N at a great height, z_2 say, cannot have any influence on $\boldsymbol{R}$ for low frequencies whose reflection level is well below z_2. The insignificant numbers in $N(z)$ are different for different items of data.

(d) The true function $N(z)$ may show some rapid variations. For example the work on partial reflections at high frequencies, § 11.13, shows that there are small scale irregularities in the D-region. For lower frequencies many of these variations of $N(z)$ would have a height scale very much less than one wavelength and could not appreciably influence the value of $\boldsymbol{R}$. The numbers that specify $N(z)$ give a running average over some height scale Δz. It is important to be able to find the greatest value Δz_0 of Δz that can be used without affecting the calculated $\boldsymbol{R}$. Then $\Delta z_0(z)$ is called the 'resolution' of the method, and it is usually dependent on z. If the tabulation interval for $N(z)$ is less than Δz_0, the table would contain redundant information that could not be found from the data.

It is item (b) above that is the main task of the inversion process. Baybulatov and Krasnushkin (1966) refer to it as the 'optimisation principle'. Shellman (1970) gives particular attention to the mapping of the standard errors of the items of data into error limits for the numbers that specify $N(z)$.

Suppose that a function $N(z)$ has been chosen and used to calculate the reflection

coefficient $\boldsymbol{R}$. We now ask: how much change can be made in $N(z)$ at various heights z, without changing the elements of $\boldsymbol{R}$ by more than the limits of observational error? The answer to this was supplied by Bailey and Jones (1974). Their method illustrates well how the basic equations can be extended, and a summary of it is therefore given below.

The basic equations (7.80) are to be solved by a downward integration starting at the top $z = z_1$ and using two independent solutions $\mathbf{e}$ that represent only upgoing waves at the start. But the equation has four independent solutions. The other two can be chosen, for example, so that there are only downgoing waves at the top. The four independent columns $\mathbf{e}$ are used as the four columns of a square matrix $\mathbf{E}$ that satisfies (7.80) so that

$$\mathbf{E}' = -\mathrm{i}\mathbf{T}\mathbf{E} \tag{19.73}$$

where, as usual, a prime means $k^{-1}\,\mathrm{d}/\mathrm{d}z$. The columns that represent upgoing waves at the top are the first two columns of $\mathbf{E}$. Now for a given function $N(z)$ we wish to study how the reflection coefficient $\boldsymbol{R}$ is affected by changes of N. At the bottom, $z = z_0$, $\boldsymbol{R}$ is found, as explained in §18.7, from the first two columns of $\mathbf{E}$. Thus we need to know how $\mathbf{E}(z_0)$ changes when $N(z)$ is changed. It is more convenient to use $\ln N$. Let the change of this be

$$\rho(z) = \delta(\ln N). \tag{19.74}$$

It is $\mathbf{T}$ that depends on $\ln N$ so let

$$\mathrm{d}\mathbf{T}/\mathrm{d}(\ln N) = \mathbf{J}(z). \tag{19.75}$$

Then if ρ is assumed to be very small, from (19.73):

$$\delta\mathbf{E}' = -\mathrm{i}(\mathbf{T}\delta\mathbf{E} + \mathbf{J}\mathbf{E}\rho). \tag{19.76}$$

Now

$$(\mathbf{E}^{-1}\delta\mathbf{E})' = \mathbf{E}^{-1}\delta\mathbf{E}' - \mathbf{E}^{-1}\mathbf{E}'\mathbf{E}^{-1}\delta\mathbf{E} = \mathbf{E}^{-1}(\delta\mathbf{E}' + \mathrm{i}\mathbf{T}\delta\mathbf{E}) = -\mathrm{i}\mathbf{E}^{-1}\mathbf{J}\mathbf{E}\rho \tag{19.77}$$

where (19.73), (19.76) have been used. At the top starting level $z = z_1$ it is assumed that $\rho(z_1) = 0$. Then the same starting value of $\mathbf{E}$ can be used for all functions $N(z)$ so that $\delta\mathbf{E}(z_1) = 0$. Now integrate with respect to z from z_0 to z_1. Then

$$\mathbf{E}^{-1}(z_0)\delta\mathbf{E}(z_0) = \mathrm{i}k\int_{z_0}^{z_1}\mathbf{E}^{-1}(\zeta)\mathbf{J}(\zeta)\mathbf{E}(\zeta)\rho(\zeta)\,\mathrm{d}\zeta. \tag{19.78}$$

This can be written

$$\delta\mathbf{E}(z_0) = \int_{z_0}^{z_1}\mathbf{K}_{\mathbf{E}}(\zeta)\rho(\zeta)\,\mathrm{d}\zeta, \text{ where } \mathbf{K}_{\mathbf{E}}(\zeta) = \mathrm{i}k\mathbf{E}(z_0)\mathbf{E}^{-1}(\zeta)\mathbf{J}(\zeta)\mathbf{E}(\zeta). \tag{19.79}$$

The four columns of $\mathbf{E}$ can be computed at the same time, so that $\mathbf{K}_{\mathbf{E}}(\zeta)$ can be found. A separate computation must be made for each different frequency, angle of incidence and azimuth of propagation used in the data.

When $\mathbf{E}(z_0)$ and $\mathbf{E}(z_0) + \delta\mathbf{E}(z_0)$ are known at the bottom of the ionosphere, they can be used to calculate the reflection coefficients $\boldsymbol{R}$ and $\boldsymbol{R} + \delta\boldsymbol{R}$ respectively, and thence the modulus and argument of any element. Some of these are the items used as data. Let γ_i be one such data item. Then if ρ is small $\delta\gamma_i$ is a linear combination of elements of the first two columns of $\delta\mathbf{E}(z_0)$ and

$$\delta\gamma_i = \int_{z_0}^{z_1} K_i(\zeta)\rho(\zeta)\,\mathrm{d}\zeta \tag{19.80}$$

where K_i is the same linear combination of the corresponding elements of $\mathbf{K}_E$. The functions $K_i(\zeta)$ are called 'data kernels'. Bailey and Jones (1974) give an example where eight data items γ_i were used. They give curves of the function $K_i(\zeta)$ for these eight items, for typical functions $N(z)$. These curves are now to be used to estimate the error limits and resolution for the function $N(z)$. Clearly a large value of $K_i(\zeta)$ at height $z = \zeta$ indicates that only a small value of ρ at that height can be tolerated by the observations of data item γ_i. Similarly where any $K_i(\zeta)$ shows rapid oscillations, the height range for one oscillation leads to an estimate of the resolution Δz_0 needed for specifying $N(z)$. In a full treatment it is necessary to consider all the data items together and Bailey and Jones use a matrix method for doing this. One important result is that where the resolution range Δz_0 is small, the error limits of $N(z)$ are large, and vice versa. A very similar method was used by Mambo *et al.* (1983). They started with a test function $N(z)$, either found from rocket data or a simple exponential function. To this they applied the method by using radio reflection data for frequencies in the range 10 to 40 kHz, to give $\delta N(z)$, and thus they obtained an improved estimate of $N(z)$.

All these methods entail a very large amount of computing. They are of the greatest theoretical interest, but there is no indication yet that they are likely to replace the direct trial and error method of Deeks for elucidating the structure of the D-region.

19.8. Full wave solutions at higher frequencies

Although the main use of computed full wave solutions has been in the study of radio waves of very low frequency of order 10 to 200 kHz, there have been a few applications for higher frequencies. They were mainly investigations of reflection and transmission for very thin but intensely ionised layers, such as are believed to account for one form of sporadic E-layer, (E_S), § 1.8. These have a thickness of order 1 km or less and a penetration frequency of a few MHz, and may give appreciable reflection for frequencies up to 7 to 10 MHz.

Measurements with rockets of the electron height distribution $N(z)$ in E_S layers have shown a great variety of shapes and thicknesses (Chessell 1971b). Some have very large $|\mathrm{d}N/\mathrm{d}z|$ at the top and bottom with almost constant N between, and

others have much sharper maxima of $N(z)$ near the middle. Most workers on the theory have used model functions $N(z)$ with free space below and above. Thus Chessell (1971b) has used the functions of fig. 15.1(b,c), the parabolic model, fig. 12.3, and others. Miller and Smith (1977) have used $N(z)$ curves derived by fitting them to rocket data. All these authors have given many computed curves to show how the reflection and transmission coefficients depend on frequency. Nearly all the published results apply for vertically incident waves. This is the most important case because nearly all observations of E_S are made with ionosondes, § 1.7, at vertical incidence.

For a normal slowly varying ionosphere at frequencies of about 1 MHz or more we would expect to be able to use ray theory, and to find that the ordinary and extraordinary waves are propagated independently, and some examples were given in ch. 13. But in some parts of an E_S layer the gradient dN/dz must be very large. Here the various coupling points, § 16.5, can be close together and close to the real z axis. Thus there is strong coupling between the ordinary and extraordinary waves, and it is for this reason that full wave solutions are needed. An incident ordinary wave then gives rise to a large amount of the extraordinary wave component in both the reflected and transmitted waves, and similarly for an incident extraordinary wave. To study this it is convenient to calculate the reflection and transmission coefficients in a form that gives the ratios of the amplitudes of the ordinary and extraordinary waves. A form $\mathscr{R}$ of the reflection coefficient with this property was given by (18.56). Chessell (1971a, b) and Miller and Smith (1977) used a slightly different but essentially equivalent form, and they also used a transmission coefficient that relates the amplitudes of the ordinary and extraordinary waves.

The various coupling points can occur in the region of the complex z plane near where $N(z)$ is increasing in the lower half of the layer. The same coupling points appear again in a region near the upper half of the layer, where $N(z)$ is decreasing. Thus the coupling processes occur twice and there is a repeated exchange of energy between the ordinary and extraordinary waves. One effect of this is that for frequencies on either side of the penetration frequency, the range of frequency where partial penetration and reflection occurs is greater than it would be according to a simple ray theory without mode coupling. The occurrence of this coupling process was demonstrated by the authors cited above, and has been confirmed by Jalonen, Nygrén and Turunen (1982).

In this type of calculation there must be points of resonance which, for some frequencies, are on or near the real z axis. They may be allowed for by one of the methods of § 18.13. If collisions are neglected the resonance condition is (4.76) with $U = 1$, and this gives

$$f^4 - f^2(f_N^2 + f_H^2) + f_H^2 f_N^2 \cos^2 \Theta = 0. \tag{19.81}$$

Hence at a point on the real z axis where the plasma frequency is f_N, a resonance will occur for two real frequencies f, one less than the gyro-frequency f_H and the other greater. Any given value of f_N, less than the value at the maximum of the layer, must occur twice, once in the lower half and once in the upper half of the layer. Thus for any E_S-layer with $N(z)$ going to zero at the boundaries, there must be two frequency ranges that give resonances on the real z axis. For any one frequency in these ranges there are two such resonances.

At a height near 115 km where E_S usually occurs, the electron collision frequency ν is of order $2 \times 10^4 s^{-1}$; see fig. 1.2. Thus at 1 MHz, $Z \approx 0.003$ and it is smaller still for higher frequencies. When collisions are allowed for the resonant points are no longer exactly on the real z axis. The path of integration can then be the real axis, but it must be remembered that near the resonance points one refractive index is very large so a very small step size must be used in the integration routine. An alternative method is to move the path of integration into the complex z plane away from the resonance points; see §18.13. This method was used by Chessell (1971a). For the resonance in the lower half of the layer the resonance point is on the negative imaginary side of the real z axis, and there must be a branch cut running from it to infinity in the negative imaginary half plane. Similarly for the upper half of the layer, the resonance point and its branch cut are on the positive imaginary side. Thus the path of integration must be given a positive imaginary displacement in the lower half of the layer and a negative imaginary displacement in the upper half. Chessell (1971a) studied cases both with and without collisions. He confirmed that when collisions are neglected there can be an apparent disappearance of up to 65% of the incident energy. For a discussion of this see end of § 19.5.

For further discussion of the theory and observations of sporadic E-layers see Reddy and Rao (1968), Reddy (1968) and the references cited in §1.8.