# 16

## Coupled wave equations

### 16.1. Introduction

For wave propagation in a stratified medium, the idea of writing the governing differential equations as a set of coupled equations has already been used in §§ 7.8, 7.15, for the study of W.K.B. solutions. The term 'coupled equations' is usually given to a set of simultaneous ordinary linear differential equations with the following properties:

(1) There is one independent variable which in this book is the height $z$ or a linear function of it.

(2) The number of equations is the same as the number of dependent variables.

(3) In each equation one dependent variable appears in derivatives up to a higher order than any other. The terms in this variable are called 'principal' terms and the remaining terms are called 'coupling' terms.

(4) The principal terms contain a different dependent variable in each equation, so that each dependent variable appears in the principal terms of one and only one equation.

It is often possible to choose the dependent variables so that the coupling terms are small over some range of $z$. Then the equations may be solved by successive approximations. As a first approximation the coupling terms are neglected, and the resulting equations can then be solved. The values thus obtained for the dependent variables are substituted in the coupling terms and the resulting equations are solved to give a better approximation. Some examples of this process are given in § 16.13. Clearly it cannot be used when any of the coupling terms becomes large.

In most of the coupled equations studied in this chapter the coupling terms are large near the turning points in the complex $z$ plane, that is points where two roots $q$ of the Booker quartic equation, ch. 6, are equal. Generally, coupled equations are not recommended for computing, because the transformations that lead to them introduce, at the turning points, singularities that are not originally present. See end of § 16.13 for further discussion of this. The importance of coupled equations is that they help in the study of the processes of reflection, and of coupling between ordinary and extraordinary waves, that is 'mode conversion'. It is therefore important to study coupled equations in the neighbourhood of a turning point, as is done in §§ 16.3, 16.4.

The basic equations for the general anisotropic stratified plasma are the set of four equations (7.80). These are already in the form of coupled equations as defined above. The first of them is

$$\mathrm{d}E_x/\mathrm{d}z + \mathrm{i}k\mathsf{T}_{11}E_x = \mathrm{i}k(\mathsf{T}_{12}E_y - \mathsf{T}_{14}\mathscr{H}_y) \tag{16.1}$$

since (7.81) shows that $\mathsf{T}_{13} = 0$. Here the terms in $E_x$ on the left-hand side are the principal terms and the terms in $E_y$, $\mathscr{H}_y$ are the coupling terms. In these equations, however, the coupling terms are in general of comparable magnitude to the principal terms, even in a homogeneous medium, so that we could not use a method of successive approximation in which the coupling terms are at first neglected.

It has been assumed in earlier chapters, especially chs. 10, 13, that the ordinary and extraordinary waves can be considered to be propagated independently. This suggests that the dependent variables should be chosen so that one refers to the ordinary wave only, and the other to the extraordinary wave only. The coupling terms should then be very small if the assumption is justified. Equations of this kind were given first by Försterling (1942) for the case of vertical incidence; see § 16.11. One of Försterling's variables $\mathscr{F}_{\mathrm{O}}$ gave the field of the total ordinary wave including both upgoing and downgoing waves. The other $\mathscr{F}_{\mathrm{E}}$, similarly, gave the total extraordinary wave. The terms 'ordinary' and 'extraordinary' are used here only for convenience of description. They may be ambiguous, as explained in §4.16. Thus Försterling did not separate the four characteristic waves completely. If the coupling terms in his equations are neglected, the resulting equations are two second order equations (16.92) governing the independent propagation of the ordinary and extraordinary waves.

The theory was extended to the general case by Clemmow and Heading (1954), who used four independent variables $\mathsf{f}_1$ to $\mathsf{f}_4$, (6.53), for the four characteristic waves; §§ 6.10, 7.15, 7.16. Their form of the coupled equations is (7.109). Where the coupling terms can be neglected, the resulting four equations (7.110) lead to the four W.K.B. solutions (7.112).

The theory of coupling and mode conversion in plasmas, presented from a somewhat different point of view from that used here, has been given by Fuchs, Ko and Bers (1981). They use a conformal mapping of the complex $z$ plane into the complex plane of $\omega n(z)/c$ ( $= k$ in their notation), and the mapping of the real $z$ axis is particularly important. They introduce branch cuts in the $z$ plane similar to those used here in §§ 16.5, 16.6. They give references to applications of the theory, including many from the literature of plasma physics.

## 16.2. First order coupled equations

The coupled equations studied in this chapter all apply for waves with the constant values $S_1 = S = \sin\theta$, $S_2 = 0$; § 6.2. Thus they apply, for example, when the incident wave below the ionosphere is an obliquely incident plane wave. The form of any set of coupled equations depends on how we define the partial waves that give the principal terms. In §§ 6.10, 7.14 the four partial waves were defined as follows. At any level $z$ consider a fictitious homogeneous medium that has all the properties of the actual medium; see end of § 6.10. Then for each partial wave the ratios of the four field components $E_x$, $-E_y$, $\mathscr{H}_x$, $\mathscr{H}_y$ are to be the same as for one of the four possible progressive waves, § 6.2, with the given value of $S$ in the fictitious medium. It was this form that led to the four coupled equations (7.109). In each equation the principal terms include a first order derivative of one of the field variables, and the coupling terms contain no derivatives of the field variables.

To study these first order coupled equations it is useful first to give explicit expressions for the elements $\Gamma_{ij}$ of the coupling matrix $\mathbf{\Gamma} = -\mathbf{S}^{-1}\mathbf{S}'$ (7.108). The prime here means $k^{-1}\mathrm{d}/\mathrm{d}z$.

Now $\mathbf{S}^{-1}\mathbf{S}$ is a constant, unity, so that $\mathbf{S}^{-1}\mathbf{S}' + (\mathbf{S}^{-1})'\mathbf{S} = 0$. Hence

$$\mathbf{\Gamma} = \tfrac{1}{2}\{(\mathbf{S}^{-1})'\mathbf{S} - \mathbf{S}^{-1}\mathbf{S}'\}. \tag{16.2}$$

Substitution from (7.134), (7.135) now gives $\Gamma_{ij}$. The relation

$$\begin{aligned}&[\{(A_iF_i)^{-\frac{1}{2}}\}'(A_jF_j)^{-\frac{1}{2}} - (A_iF_i)^{-\frac{1}{2}}\{(A_jF_j)^{-\frac{1}{2}}\}']\\ &\quad\times[(\bar{a}_5q_i + \bar{a}_6)(a_3q_j + a_4) + A_iA_j(q_i + q_j) + (\bar{a}_3q_i + \bar{a}_4)(a_5q_j + a_6)] = 0\end{aligned} \tag{16.3}$$

holds for all $i, j$ since the first bracket is zero when $i = j$, and the second is zero when $i \neq j$ because $\mathbf{S}^{-1}\mathbf{S}$ is diagonal. The result is therefore

$$\begin{aligned}2\Gamma_{ij} = (A_iF_iA_jF_j)^{-\frac{1}{2}}\{&(\bar{a}_5q_i + \bar{a}_6)'(a_3q_j + a_4) - (\bar{a}_5q_i + \bar{a}_6)(a_3q_j + a_4)'\\ &+ (q_iA_i)'A_j - q_iA_iA_j' + q_jA_jA_i' - (q_jA_j)'A_i\\ &+ (\bar{a}_3q_i + \bar{a}_4)'(a_5q_j + a_6) - (\bar{a}_3q_i + \bar{a}_4)(a_5q_j + a_6)'\}.\end{aligned} \tag{16.4}$$

Finally, with the help of (7.131), a rearrangement gives

$$\begin{aligned}2\Gamma_{ii} = (A_iF_i)^{-1}\{&q_i^2(a_3\bar{a}_5' - a_3'\bar{a}_5 + a_5\bar{a}_3' - a_5'\bar{a}_3) + a_4\bar{a}_6' - a_4'\bar{a}_6 + a_6\bar{a}_4' - a_6'\bar{a}_4\\ &+ q_i(a_3\bar{a}_6' - a_3'\bar{a}_6 + a_4\bar{a}_5' - a_4'\bar{a}_5 + a_5\bar{a}_4' - a_5'\bar{a}_4 + a_6\bar{a}_3' - a_6'\bar{a}_3)\},\end{aligned} \tag{16.5}$$

$$\begin{aligned}2\Gamma_{ij} = (A_iF_iA_jF_j)^{-\frac{1}{2}}\{&(q_i'q_j - q_j'q_i)(a_3\bar{a}_5 + a_5\bar{a}_3)\\ &+ q_iq_j(a_3\bar{a}_5' - a_3'\bar{a}_5 + a_5\bar{a}_3' - a_5'\bar{a}_3) + q_i(a_4\bar{a}_5' - a_4'\bar{a}_5 + a_6\bar{a}_3' - a_6'\bar{a}_3)\\ &+ q_j(a_3\bar{a}_6' - a_3'\bar{a}_6 + a_5\bar{a}_4' - a_5'\bar{a}_4) + (q_i + q_j)(A_i'A_j - A_iA_j')\\ &+ (q_i' - q_j')(a_4\bar{a}_5 + a_6\bar{a}_3 + A_iA_j) + a_4\bar{a}_6' - a_4'\bar{a}_6 + a_6\bar{a}_4' - a_6'\bar{a}_4\}.\end{aligned} \tag{16.6}$$

With these elements of $\mathbf{\Gamma}$ the coupled equations (7.109) are

$$f_i' + \mathrm{i}q_if_i = \Gamma_{i1}f_1 + \Gamma_{i2}f_2 + \Gamma_{i3}f_3 + \Gamma_{i4}f_4 \quad (i = 1, 2, 3, 4) \tag{16.7}$$

where, from (6.53) second equation, and (7.135)

$$f_i = (A_iF_i)^{-\frac{1}{2}}\{(\bar{a}_5q_i + \bar{a}_6)E_x - q_iA_iE_y + A_i\mathscr{H}_x + (\bar{a}_3q_i + \bar{a}_4)\mathscr{H}_y\}. \tag{16.8}$$

All the terms of the coupling coefficients $\Gamma_{ij}$ (16.6) contain derivatives with respect to $z$. They are therefore small in a sufficiently slowly varying medium, provided that the denominators $(A_iF_iA_jF_j)^{-\frac{1}{2}}$ are not small. At points in the complex $z$ plane where they can be neglected, the remaining terms in (7.109) are four independent differential equations whose solutions are the W.K.B. approximations, as defined and derived in § 7.15. These four solutions are associated with the four roots $q_1$ to $q_4$ of the Booker quartic, ch. 6, and will be referred to as wave 1 to wave 4 respectively.

As an illustration of the properties of the coupling coefficients consider the special case of $\Gamma_{12}$ and $\Gamma_{21}$ from (16.6). Their denominators $(A_1F_1A_2F_2)^{-\frac{1}{2}}$ contain a factor $(q_1 - q_2)^{-1}$, so that they are both infinite at a point $z = z_p$ in the complex $z$ plane where $q_1 = q_2$. Such a point is called a coupling point. It may be a reflection point. In the general treatment of this section there is no distinction between reflection and coupling, and mathematically they are the same phenomena. Near $z_p$ waves 1 and 2 are not independently propagated and are said to be 'strongly coupled'. But (16.6) also shows that $\Gamma_{13}, \Gamma_{14}, \Gamma_{31}, \Gamma_{41}, \Gamma_{23}, \Gamma_{24}, \Gamma_{32}, \Gamma_{42}$ all contain a factor $(q_1 - q_2)^{-\frac{1}{2}}$ so that they too are infinite at $z_p$, even through the power is $-\frac{1}{2}$ giving an infinity of lower order than the power $-1$ in $\Gamma_{12}, \Gamma_{21}$. This might still suggest that there is strong coupling of wave 3 and wave 4 with waves 1 and 2. But this is not so. Provided that $q_1 = q_2 \neq q_3$ and $\neq q_4$, it can be shown that wave 3 and wave 4 are propagated independently of waves 1 and 2. The proof of this was first given by Booker (1936). Another proof follows from the theory in § 16.3, based on the original method of Heading (1961). An alternative proof for vertically incident waves is given by using second order coupled equations; see §§ 16.10–16.12.

The expression (16.6) for $\Gamma_{ij}$ has a slightly different form if the matrix **S** in (6.53), (7.108) is normalised differently. For example if the normalisation factor $(A_jF_j)^{-\frac{1}{2}}$ is omitted from (7.134), the coefficient $(A_iF_iA_jF_j)^{-\frac{1}{2}}$ in (16.6) must be replaced by $(A_jF_j)^{-1}$. Then in the example just discussed $\Gamma_{13}, \Gamma_{14}, \Gamma_{23}, \Gamma_{24}$ are not now infinite at $z_p$ where $q_1 = q_2$, but $\Gamma_{31}, \Gamma_{41}, \Gamma_{32}, \Gamma_{42}$ now contain a factor $(q_1 - q_2)^{-1}$ and so they are infinite to a higher order. The difficulty cannot be removed by a change in the normalisation of **S**. Thus care is needed when using the first order coupled equations near a reflection or coupling point. Their main function is to illustrate physical principles, and they are not recommended for computing.

The expressions (16.5), (16.6) for the elements of $\mathbf{\Gamma}$ take simpler forms in some special cases.

(a) *Vertical incidence*

When $\theta = 0, S = 0, C = 1$ the numbers $a$ and $\bar{a}$ take the simple forms (7.137) and some of them are zero. Then (16.5) shows that $\Gamma_{ii} = 0$ and (16.6) gives

$$
\begin{aligned}
2\Gamma_{ij} &= (A_iF_iA_jF_j)^{-\frac{1}{2}}\{(A_i'A_j - A_iA_j')(q_i + q_j) + (A_iA_j - a_4^2)(q_i' - q_j')\} \\
&= \left(\frac{A_i}{F_i}\frac{A_j}{F_j}\right)^{\frac{1}{2}}\left\{\left(\frac{A_i'}{A_i} - \frac{A_j'}{A_j}\right)(q_i + q_j) + \left(1 - \frac{a_4^2}{A_iA_j}\right)(q_i' - q_j')\right\} \qquad (16.9)
\end{aligned}
$$

which is antisymmetric. The $q$s are now given by (7.139) and it can be shown from (7.140)–(7.144) that

$$\frac{A_1}{F_1} = -\frac{A_2}{F_2} = \frac{\rho_O^2}{2n_O(\rho_O^2 - 1)}, \quad -\frac{A_3}{F_3} = \frac{A_4}{F_4} = \frac{1}{2n_E(\rho_O^2 - 1)} \tag{16.10}$$

and from (7.143), (7.144)

$$\frac{A_1'}{A_1} = \frac{A_2'}{A_2} = \frac{a_4'}{a_4} + \frac{\rho_O'}{\rho_O}, \quad \frac{A_3'}{A_3} = \frac{A_4'}{A_4} = \frac{a_4'}{a_4} - \frac{\rho_O'}{\rho_O}. \tag{16.11}$$

For the square root in (16.9) a sign convention must be adopted and we take

$$(-n_O)^{-\frac{1}{2}} = \mathrm{i}n_O^{-\frac{1}{2}}, \; (-n_E)^{-\frac{1}{2}} = \mathrm{i}n_E^{-\frac{1}{2}}, \tag{16.12}$$

where $n_O^{-\frac{1}{2}}$, $n_E^{-\frac{1}{2}}$ have positive real parts. Consider now the case $i = 1$, $j = 3$. Then $1 - a_4^2/A_iA_j$ is zero from (7.143), (7.144), (4.29) and the remaining term in (16.9) gives

$$\Gamma_{13} = \tfrac{1}{2}\mathrm{i}\psi(n_O + n_E)(n_On_E)^{-\frac{1}{2}} \tag{16.13}$$

where

$$\psi = \rho_O'/(\rho_O^2 - 1) = \frac{1}{2k}\frac{\mathrm{d}}{\mathrm{d}z}\ln\left(\frac{\rho_O - 1}{\rho_O + 1}\right). \tag{16.14}$$

The other elements of $\mathbf{\Gamma}$ can similarly be found and when they are inserted in (7.109) they give the following four first order coupled equations

$$\left.\begin{aligned}
\mathrm{f}_1' + \mathrm{i}n_O\mathrm{f}_1 &= -(n_O'/2\mathrm{i}n_O)\mathrm{f}_2 + \tfrac{1}{2}\mathrm{i}\psi(n_O + n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_3 + \tfrac{1}{2}\psi(n_O - n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_4,\\
\mathrm{f}_2' - \mathrm{i}n_O\mathrm{f}_2 &= (n_O'/2\mathrm{i}n_O)\mathrm{f}_1 + \tfrac{1}{2}\psi(n_O - n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_3 - \tfrac{1}{2}\mathrm{i}\psi(n_O + n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_4,\\
\mathrm{f}_3' + \mathrm{i}n_E\mathrm{f}_3 &= -\tfrac{1}{2}\psi(n_O + n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_1 - \tfrac{1}{2}\psi(n_O - n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_2 + (n_E'/2\mathrm{i}n_E)\mathrm{f}_4,\\
\mathrm{f}_4' - \mathrm{i}n_E\mathrm{f}_4 &= -\tfrac{1}{2}\psi(n_O - n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_1 + \tfrac{1}{2}\mathrm{i}\psi(n_O + n_E)(n_On_E)^{-\frac{1}{2}}\mathrm{f}_2 - (n_E'/2\mathrm{i}n_E)\mathrm{f}_4,
\end{aligned}\right\} \tag{16.15}$$

where

$$\left.\begin{aligned}
\mathrm{f}_1 &= \{2n_O(\rho_O^2 - 1)\}^{-\frac{1}{2}}(n_OE_x - n_O\rho_OE_y + \rho_O\mathscr{H}_x + \mathscr{H}_y),\\
\mathrm{f}_2 &= -\mathrm{i}\{2n_O(\rho_O^2 - 1)\}^{-\frac{1}{2}}(n_OE_x - n_O\rho_OE_y - \rho_O\mathscr{H}_x - \mathscr{H}_y),\\
\mathrm{f}_3 &= -\mathrm{i}\{2n_E(\rho_O^2 - 1)\}^{-\frac{1}{2}}(-\rho_On_EE_x + n_EE_y - \mathscr{H}_x - \rho_O\mathscr{H}_y),\\
\mathrm{f}_4 &= \{2n_E(\rho_O^2 - 1)\}^{-\frac{1}{2}}(-\rho_On_EE_x + n_EE_y + \mathscr{H}_x + \rho_O\mathscr{H}_y).
\end{aligned}\right\} \tag{16.16}$$

Here $\mathrm{f}_1$ to $\mathrm{f}_4$ are associated with the four wave types upgoing ordinary, downgoing ordinary, upgoing extraordinary, downgoing extraordinary, respectively. This order is a matter of choice and is determined here by (7.139) for the four $q$s. Any other choice would be permissible and in § 18.10 a different order is used with $\mathrm{f}_2, \mathrm{f}_3$ interchanged. The parameter $\psi$ in (16.14) is called the 'coupling parameter'. It was introduced by Försterling (1942) and plays an important part in all the theory of coupled equations. It may be regarded as a function of the state of the ionosphere at each level, and its properties are discussed in § 16.12. Since it is a derivative with respect to the height $z$, it is zero in a homogeneous medium.

The elements $\Gamma_{12}$, $\Gamma_{21}$ on the right of (16.15) are large near levels where $n_O = 0$. They are associated with coupling between the upgoing and downgoing ordinary waves, that is with reflection of the ordinary wave. Similarly $\Gamma_{34}$ and $\Gamma_{43}$ are associated with reflection of the extraordinary wave. The elements $\Gamma_{13}$ and $\Gamma_{31}$ are associated with coupling between the upgoing ordinary and extraordinary waves, and similarly $\Gamma_{24}$ and $\Gamma_{42}$ are associated with the two downgoing waves. It has been suggested that these two kinds of coupling may be responsible for the 'Z-trace' sometimes observed in ionograms at high latitudes; see § 16.8. These terms become large when $\psi$ is large, that is when $n_O \approx n_E$, for it is shown, (16.96), that $\psi$ has a factor $(n_O - n_E)^{-2}$; see § 16.12. Thus $\Gamma_{13}$, $\Gamma_{31}$, $\Gamma_{24}$ and $\Gamma_{42}$ have infinities of this order.

The elements $\Gamma_{14}$ and $\Gamma_{41}$ would be associated with coupling between the upgoing ordinary wave and the downgoing extraordinary wave. This could be strong only near points where $n_O + n_E = 0$, and (16.96) indeed shows that $\psi$ has a factor $(n_O + n_E)^{-2}$. But $n_O$ and $n_E$ are usually defined so that their real parts are not less than zero; see § 16.5 and (16.52). This condition, therefore, could never occur. A similar conclusion applies for $\Gamma_{23}$, $\Gamma_{32}$. It has been suggested that coupling of these types might be responsible for a reflection known as the 'coupling echo'. This subject, however, is more complicated, and is discussed in § 16.14.

When the earth's magnetic field is neglected so that $\psi = 0$, the equations (16.15) separate into two pairs, one pair containing only $\mathsf{f}_1$ and $\mathsf{f}_2$ and the other pair containing only $\mathsf{f}_3$ and $\mathsf{f}_4$. By the substitutions $\mathsf{f}_1 = n_O^{\frac{1}{2}} E_y^{(1)}$, $\mathsf{f}_2 = n_O^{\frac{1}{2}} E_y^{(2)}$ the first pair can be converted to the form (7.46), (7.47). Similarly the second pair can be converted to the same form.

### (b) *Other special cases*

There are two other special cases where the first order coupled wave equations have a fairly simple form. The first is when the earth's magnetic field is vertical, see problem 16.2, and the second is when the earth's magnetic field is horizontal and in the plane of incidence, that is for north–south or south–north propagation at the magnetic equator. The equations are similar in general form to (16.15). They are not needed in this book, and they have not been extensively used. The algebraic details were given by Budden and Clemmow (1957) and by Budden (1961a, § 18.15).

## 16.3. Coupled equations near a coupling point

The derivation of the first order coupled equations in § 7.15 depended on the use of the transforming matrix **S** in (6.53) and of its inverse. Thus it is necessary that **S** shall be non-singular. At a coupling point, where two of the $q$s are equal, the two corresponding columns (7.118) or (7.134) of **S** are the same so **S** is there singular and $\mathbf{S}^{-1}$ does not exist. To derive a set of coupled equations that can be used throughout a domain of the $z$ plane containing a coupling point, an alternative treatment must be used. The following treatment is based on the method of Heading (1961) who gave the general mathematical theory. It is here written in the language of radio propagation in stratified media (Budden, 1972). A treatment very similar to that of the present section was given by Ludwig (1970). Heading's method has also been used in the study of mode conversion, by Fuchs, Ko and Bers (1981).

Let $z_p$ be a coupling point where $q_1 = q_2$ and consider a domain of the complex $z$ plane containing $z_p$ in which there is no other coupling point and in which the coefficients in the Booker quartic $F(q) = 0$, (6.15) or (7.128), are analytic functions of $z$. Then, where the roots $q_i$ are distinct, each of them is analytic. A factor $(q - q_3)(q - q_4)$ can be removed from the quartic, leaving a quadratic for $q$ whose coefficients are analytic in the domain. Each of its two roots, however, contains a square root proportional to $q_1 - q_2$, that is zero when $z = z_p$. Thus $z_p$ is a branch point of $q_1$ and $q_2$. Let

$$\alpha = \tfrac{1}{2}(q_1 + q_2), \quad \beta = \tfrac{1}{2}(q_1 - q_2). \tag{16.17}$$

Then $\alpha$ is analytic but $\beta$ has a branch point at $z_p$ and $\beta^2$ is analytic there. These symbols must not be confused with the coefficients $\alpha, \beta$ in the Booker quartic (6.15), (6.16).

For the columns $\mathbf{s}_j$ of **S** we now choose the form (7.134) with the factor $(A_j F_j)^{-\frac{1}{2}}$ omitted. Thus in this section the method of normalising **S** is different from that used in §§7.14–7.16. Then on using (7.119) (16.17)

$$\tfrac{1}{2}(\mathbf{s}_1 - \mathbf{s}_2) = \beta \begin{pmatrix} a_3 \\ 2\alpha + a_1 \\ 3\alpha^2 + \beta^2 + 2\alpha a_1 + a_2 \\ a_5 \end{pmatrix} = \beta \mathbf{s}_\beta. \tag{16.18}$$

This is zero at $z = z_p$ solely because of the factor $\beta$. The other factor $\mathbf{s}_\beta$ is a column matrix whose elements are all analytic and in general not all zero in the domain. It is also convenient to define

$$\mathbf{s}_\alpha = \tfrac{1}{2}(\mathbf{s}_1 + \mathbf{s}_2) = \begin{pmatrix} a_3\alpha + a_4 \\ \alpha^2 + \beta^2 + a_1\alpha + a_2 \\ \alpha^3 + 3\alpha\beta^2 + a_1(\alpha^2 + \beta^2) + \alpha a_2 \\ a_5\alpha + a_6 \end{pmatrix} \tag{16.19}$$

whose elements are analytic and in general not all zero in the domain.

Now let **U** be a $4 \times 4$ matrix whose columns are $\mathbf{s}_\alpha, \mathbf{s}_\beta, \mathbf{s}_3, \mathbf{s}_4$. They are linearly independent in the domain so that **U** is analytic and non-singular, and $\mathbf{U}^{-1}$ exists. The matrix **U** will now be used instead of **S** to derive coupled equations. It can be shown from (16.17)–(16.19) and (7.85) that

$$\mathbf{TU} = \mathbf{UA} \quad \text{where} \quad \mathbf{A} = \begin{pmatrix} \alpha & 1 & 0 & 0 \\ \beta^2 & \alpha & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & q_4 \end{pmatrix}. \tag{16.20}$$

Instead of $\mathbf{f}$ in (6.53), define four new independent variables $\mathrm{c}_i$ forming a column $\mathbf{c}$ where

$$\mathbf{e} = \mathbf{Uc}, \quad \mathbf{c} = \mathbf{U}^{-1}\mathbf{e} \tag{16.21}$$

and substitute for $\mathbf{e}$ in (7.80). Then $\mathbf{c}$ must satisfy

$$\mathbf{c}' + \mathrm{i}\mathbf{Ac} = \mathbf{\Lambda c} \tag{16.22}$$

where

$$\mathbf{\Lambda} = -\mathbf{U}^{-1}\mathbf{U}'. \tag{16.23}$$

The four equations (16.22) are the required new form of the first order coupled equations.

Any transforming matrix must satisfy two requirements throughout the chosen domain of the $z$ plane: (a) it must be non-singular so that its inverse exists, (b) it must be analytic so that it can be differentiated. The new transforming matrix $\mathbf{U}$ satisfies both these requirements. Hence the new coupling matrix $\mathbf{\Lambda}$ is analytic and bounded throughout the domain.

In a sufficiently slowly varying medium the non-diagonal elements of $\mathbf{\Lambda}$ can be neglected to a first approximation. Then the waves associated with $\mathrm{c}_3, \mathrm{c}_4$, that is with $q_3, q_4$ are independently propagated and are the same as given by (7.117). Those associated with $\mathrm{c}_1, \mathrm{c}_2$, that is with $q_1, q_2$, however, remain strongly coupled because of the non-diagonal elements of $\mathbf{A}$. With this approximation (16.22) gives

$$\begin{pmatrix} \mathrm{c}_1 \\ \mathrm{c}_2 \end{pmatrix}' = -\mathrm{i}\begin{pmatrix} \alpha + \mathrm{i}\Lambda_{11} & 1 \\ \beta^2 & \alpha + \mathrm{i}\Lambda_{22} \end{pmatrix}\begin{pmatrix} \mathrm{c}_1 \\ \mathrm{c}_2 \end{pmatrix}. \tag{16.24}$$

Now $\Lambda_{11}, \Lambda_{22}$ contain derivatives with respect to $z$ and are therefore small in a slowly varying medium. Suppose that they are small enough to be neglected. Let

$$\begin{pmatrix} \mathrm{c}_1 \\ \mathrm{c}_2 \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \exp\left(-\mathrm{i}k\int^z \alpha\,\mathrm{d}z\right). \tag{16.25}$$

Then (16.24) gives

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}' = -\mathrm{i}\begin{pmatrix} 0 & 1 \\ \beta^2 & 0 \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}. \tag{16.26}$$

If $\Lambda_{11}$ and $\Lambda_{22}$ may not be neglected but are equal, the $\alpha$ in (16.25) is replaced by $\alpha + \mathrm{i}\Lambda_{11}$. If they are unequal it is still possible to reduce the equations to the form

(16.26) by replacing (16.25) by a matrix transformation

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \boldsymbol{L}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \tag{16.27}$$

and it can be shown that the $2 \times 2$ matrix $\boldsymbol{L}$ is analytic in the domain.

Now (16.26) gives

$$h_1'' + \beta^2 h_1 = 0, \quad h_2 = \mathrm{i}h_1' \tag{16.28}$$

and it is these equations that determine the behaviour of the solutions near an isolated coupling point, that is a zero of $\beta$. The first is of the same form as (7.6) with $\beta$ instead of $q$, and we may therefore use Langer's (1937) form (8.72) of the uniform approximation. Let

$$\zeta = \left(\frac{3}{2}\mathrm{i}k\int_{z_\mathrm{p}}^{z}\beta\mathrm{d}z\right)^{2/3} \tag{16.29}$$

(compare (8.68)). Then the solution analogous to (8.72) is

$$h_1 = \zeta^{\frac{1}{4}}\beta^{-\frac{1}{2}}\mathrm{Ai}(\zeta), \quad h_2 = -\zeta^{-\frac{1}{4}}\beta^{\frac{1}{2}}\mathrm{Ai}'(\zeta) + (\text{small terms})\mathrm{Ai}(\zeta) \tag{16.30}$$

where $\mathrm{Ai}'(\zeta)$ is $\mathrm{dAi}(\zeta)/\mathrm{d}\zeta$; this prime does *not* mean $k^{-1}\mathrm{d}/\mathrm{d}z$. The 'small terms' are zero if $\beta^2$ is proportional to $z - z_\mathrm{p}$. Since $\zeta$ contains a cube root, it can take three different values and there are three different solutions (16.30) but only two are independent because of the linear relation (8.63) satisfied by the Airy integral functions. The physical conditions of the problem are used to choose the correct solution. For some examples see § 16.7.

If the chosen domain of the complex $z$ plane extends out from $z_\mathrm{p}$ to where $\zeta$ is large enough, and if the uniform approximation (16.30) can still be used there, the Airy integral function may be set equal to a combination of its two asymptotic approximations (8.27), (8.28). These give for $h_1$ from (16.30) and (16.29) the two solutions

$$h_1 \propto \beta^{-\frac{1}{2}}\exp\left(\mp \mathrm{i}k\int_{z_\mathrm{p}}^{z}\beta\mathrm{d}z\right) \tag{16.31}$$

and where these are used in (16.25) they give, with (16.17):

$$c_1 \propto (q_1 - q_2)^{-\frac{1}{2}}\exp\left(-\mathrm{i}k\int^{z} q_1\,\mathrm{d}z\right), \quad c_2 \propto (q_2 - q_1)^{\frac{1}{2}}\exp\left(-\mathrm{i}k\int^{z} q_2\,\mathrm{d}z\right) \tag{16.32}$$

and hence the field variables **e** can be found from (16.21). These two solutions give the two W.K.B. solutions (7.136) associated with the two roots $q_1$, $q_2$ of the Booker quartic. Thus the form (16.22) of the coupled equations leads to a solution (16.30) which is uniformly valid near the coupling point $z_\mathrm{p}$ and at points sufficiently remote from it that the W.K.B. solutions take over.

Some examples of the use of this method are given in the following § 16.4. Equations (16.29)–(16.32) are used in the study of the phase integral method for coupling in §16.7.

The solution (16.30) can be used to supply a rough criterion to show when the W.K.B. solutions are good approximations. The method is the exact parallel of that used in §§ 7.10, 8.10. It was shown in § 8.10 that the asymptotic approximations for the function Ai($\zeta$) can be used with an accuracy better than about 8 to 9% if $|\zeta| \geqslant 1$. Suppose that, near $z_p$, $\beta^2$ is a linear function of $z$ so that $\beta^2 \approx K(z - z_p)$. Then (16.29) gives

$$|\zeta|^{\frac{3}{2}} \approx \left|\frac{k\beta^3}{K}\right| = \left|\frac{k\beta^3}{\mathrm{d}(\beta^2)/\mathrm{d}z}\right| = \tfrac{1}{2}k\left|\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{\beta}\right)\right|^{-1} \tag{16.33}$$

and from (16.17) the criterion $|\zeta| \geqslant 1$ gives

$$\frac{1}{k}\left|\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{q_1 - q_2}\right)\right| \leqslant \tfrac{1}{4}. \tag{16.34}$$

If this is used near a reflection point in an isotropic ionosphere, we have $-q_2 = q_1 = q$ and then (16.34) and (7.61) are identical. If (16.34) is used to test the W.K.B. solution $\mathbf{c}_1$ in (16.32), that is the one with $q_1$ in the exponent, it must be applied for all three pairs $q_1$, $q_2$ and $q_1$, $q_3$ and $q_1$, $q_4$.

## 16.4. Application to vertical incidence

The case of vertical incidence is important in radio probing of the ionosphere and is useful for illustrating the methods of §16.3. It is convenient to choose the $x$ and $y$ axes so that the earth's magnetic field is in the $x$–$z$ plane. Some of the algebra for this case has been given in §§ 7.17, 16.2(a). The four roots $q$ of the Booker quartic are now the refractive indices for the upgoing and downgoing ordinary and extraordinary waves. The polarisation $\rho_O$ is the same for both upgoing and downgoing ordinary waves, as explained near the end of § 4.4. Similarly, from (4.29) the polarisation of both upgoing and downgoing extraordinary waves is $\rho_E = 1/\rho_O$.

The theory of § 16.3 dealt with the neighbourhood of only one isolated coupling point and ignored the others. It is possible, however, to deal with two coupling points, A, B, at the same time, provided that the two $q$s that are equal at A, say $q_1, q_2$, are different from the two that are equal at B, say $q_3, q_4$. Then for A the operations (16.18), (16.19) are applied to the columns $\mathbf{s}_1$, $\mathbf{s}_2$ of $\mathbf{S}$, and for B they are at the same time applied to the columns $\mathbf{s}_3$, $\mathbf{s}_4$. Two cases will now be considered.

### (a) *Reflection points*

One form of coupling point is a reflection point A for the two ordinary waves, where $n_O = 0$. Another is a reflection point B for the two extraordinary waves, where $n_E = 0$.

We therefore take

$$q_1 = -q_2 = n_{\mathrm{O}}, \quad q_3 = -q_4 = n_{\mathrm{E}}. \tag{16.35}$$

The four columns of the matrix $\mathbf{S}$ may be chosen to be the four sets given in round brackets in (7.145)–(7.148). Thus

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & \rho_{\mathrm{O}} & \rho_{\mathrm{O}} \\ -\rho_{\mathrm{O}} & -\rho_{\mathrm{O}} & -1 & -1 \\ -\rho_{\mathrm{O}} n_{\mathrm{O}} & \rho_{\mathrm{O}} n_{\mathrm{O}} & -n_{\mathrm{E}} & n_{\mathrm{E}} \\ n_{\mathrm{O}} & -n_{\mathrm{O}} & \rho_{\mathrm{O}} n_{\mathrm{E}} & -\rho_{\mathrm{O}} n_{\mathrm{E}} \end{pmatrix}. \tag{16.36}$$

For the ordinary waves (16.17) gives $\alpha = 0$, $\beta = n_{\mathrm{O}}$ and the first two columns $\mathbf{s}_\alpha$, $\mathbf{s}_\beta$ of $\mathbf{U}$ are $\frac{1}{2}(\mathbf{s}_1 + \mathbf{s}_2)$, $\frac{1}{2}(\mathbf{s}_1 - \mathbf{s}_2)/n_{\mathrm{O}}$. Similarly for the extraordinary waves, $\alpha = 0$, $\beta = n_{\mathrm{E}}$ and the last two columns of $\mathbf{U}$ are $\frac{1}{2}(\mathbf{s}_3 + \mathbf{s}_4)$, $\frac{1}{2}(\mathbf{s}_2 - \mathbf{s}_4)/n_{\mathrm{E}}$. Hence

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & -\rho_{\mathrm{O}} & 0 \\ -\rho_{\mathrm{O}} & 0 & 1 & 0 \\ 0 & -\rho_{\mathrm{O}} & 0 & 1 \\ 0 & 1 & 0 & -\rho_{\mathrm{O}} \end{pmatrix}, \quad \mathbf{U}^{-1} = (1-\rho_{\mathrm{O}}^2)^{-1} \begin{pmatrix} 1 & \rho_{\mathrm{O}} & 0 & 0 \\ 0 & 0 & \rho_{\mathrm{O}} & 1 \\ \rho_{\mathrm{O}} & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho_{\mathrm{O}} \end{pmatrix} \tag{16.37}$$

and from (16.23)

$$\mathbf{\Lambda} = \frac{\rho_{\mathrm{O}} \rho_{\mathrm{O}}'}{\rho_{\mathrm{O}}^2 - 1} \mathbf{1} - \psi \mathbf{C} \tag{16.38}$$

where $\mathbf{1}$ is the unit $4 \times 4$ matrix, $\psi$ is Försterling's (1942) coupling parameter (16.14), and

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 & 0, \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{16.39}$$

Further, from (16.20)

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ n_{\mathrm{O}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & n_{\mathrm{E}}^2 & 0 \end{pmatrix} \tag{16.40}$$

and (16.22) gives

$$\mathbf{c}' + \mathrm{i}\mathbf{A}\mathbf{c} = -\psi \mathbf{C}\mathbf{c} + \frac{\rho_{\mathrm{O}} \rho_{\mathrm{O}}'}{\rho_{\mathrm{O}}^2 - 1} \mathbf{c} \tag{16.41}$$

where $\mathbf{c}$ is given by (16.21). The last term can be removed by setting $\mathbf{c} = \mathbf{b}(\rho_{\mathrm{O}}^2 - 1)^{-\frac{1}{2}}$ (compare (16.25)). Then the equations (16.41) can be partitioned thus

$$\begin{pmatrix} \mathrm{b}_1 \\ \mathrm{b}_3 \end{pmatrix}' = -\mathrm{i} \begin{pmatrix} \mathrm{b}_2 \\ \mathrm{b}_4 \end{pmatrix} - \psi \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{b}_1 \\ \mathrm{b}_3 \end{pmatrix},$$

$$\begin{pmatrix} \mathrm{b}_2 \\ \mathrm{b}_4 \end{pmatrix}' = -\mathrm{i} \begin{pmatrix} n_{\mathrm{O}}^2 & 0 \\ 0 & n_{\mathrm{E}}^2 \end{pmatrix} \begin{pmatrix} \mathrm{b}_1 \\ \mathrm{b}_3 \end{pmatrix} - \psi \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{b}_2 \\ \mathrm{b}_4 \end{pmatrix} \tag{16.42}$$

Differentiate the first of these once, and use the second to eliminate $\mathrm{b}_2$, $\mathrm{b}_4$. This gives

$$\begin{pmatrix}\mathrm{b}_1\\ \mathrm{b}_3\end{pmatrix}'' + \begin{pmatrix} n_\mathrm{O}^2+\psi^2 & 0\\ 0 & n_\mathrm{E}^2+\psi^2\end{pmatrix}\begin{pmatrix}\mathrm{b}_1\\ \mathrm{b}_3\end{pmatrix} = -2\psi\begin{pmatrix}\mathrm{b}_3\\ \mathrm{b}_1\end{pmatrix}' - \psi'\begin{pmatrix}\mathrm{b}_3\\ \mathrm{b}_1\end{pmatrix}. \tag{16.43}$$

These are the second order coupled equations that Försterling (1942) derived by a different method. No approximations are used in their derivation. They are discussed in more detail in §§ 16.10–16.12. They or the first order equations (16.42) can be used near any reflection point where $n_\mathrm{O}=0$ or $n_\mathrm{E}=0$ because the transforming matrix $\mathbf{U}$ (16.37) was chosen to achieve this. They cannot be used near a coupling point where $n_\mathrm{O}=n_\mathrm{E}$, $\rho_\mathrm{O}=-1$ because $\psi$ and $\psi'$ are infinite there.

### (b) *Coupling points*

Coupling occurs between the two upgoing waves at points where $n_\mathrm{O}=n_\mathrm{E}$, and these are also coupling points for the two downgoing waves. These two coupling processes can be treated together. We again use (16.35) and by analogy from (16.17) we take

$$\alpha=\tfrac{1}{2}(n_\mathrm{O}+n_\mathrm{E})=\tfrac{1}{2}(q_1+q_3)=-\tfrac{1}{2}(q_2+q_4),$$
$$\beta=\tfrac{1}{2}(n_\mathrm{O}-n_\mathrm{E})=\tfrac{1}{2}(q_1-q_3)=\tfrac{1}{2}(q_4-q_2). \tag{16.44}$$

For the matrix $\mathbf{S}$, instead of (16.36), we now use

$$\mathbf{S}=\begin{pmatrix}1 & 1 & 1 & 1\\ -\rho_\mathrm{O} & -\rho_\mathrm{O} & -\rho_\mathrm{E} & -\rho_\mathrm{E}\\ -\rho_\mathrm{O}n_\mathrm{O} & \rho_\mathrm{O}n_\mathrm{O} & -\rho_\mathrm{E}n_\mathrm{E} & \rho_\mathrm{E}n_\mathrm{E}\\ n_\mathrm{O} & -n_\mathrm{O} & n_\mathrm{E} & -n_\mathrm{E}\end{pmatrix}. \tag{16.45}$$

The columns of $\mathbf{U}$ are $\tfrac{1}{2}(\mathbf{s}_1+\mathbf{s}_3)$, $(\mathbf{s}_1-\mathbf{s}_3)/(n_\mathrm{O}-n_\mathrm{E})$, $\tfrac{1}{2}(\mathbf{s}_2+\mathbf{s}_4)$, $(\mathbf{s}_2+\mathbf{s}_4)/(n_\mathrm{O}-n_\mathrm{E})$ so that

$$\mathbf{U}=\begin{pmatrix}E & 0\\ 0 & F\end{pmatrix}\begin{pmatrix}\mathbf{1} & \mathbf{1}\\ \mathbf{1} & -\mathbf{1}\end{pmatrix} \tag{16.46}$$

where

$$E=\begin{pmatrix}1 & 0\\ -\tfrac{1}{2}(\rho_\mathrm{O}+\rho_\mathrm{E}) & (\rho_\mathrm{E}-\rho_\mathrm{O})/(n_\mathrm{O}-n_\mathrm{E})\end{pmatrix},$$
$$F=\begin{pmatrix}-\tfrac{1}{2}(\rho_\mathrm{O}n_\mathrm{O}+\rho_\mathrm{E}n_\mathrm{E}) & (\rho_\mathrm{E}n_\mathrm{E}-\rho_\mathrm{O}n_\mathrm{O})/(n_\mathrm{O}-n_\mathrm{E})\\ \tfrac{1}{2}(n_\mathrm{O}+n_\mathrm{E}) & 1\end{pmatrix}. \tag{16.47}$$

The matrix $\mathbf{A}$ from (16.20) is $\mathbf{U}^{-1}\mathbf{T}\mathbf{U}$ whence with (7.85) and (16.44)

$$\mathbf{A}=\begin{pmatrix}B & 0\\ 0 & B\end{pmatrix},\quad B=\begin{pmatrix}\alpha & 1\\ \beta^2 & \alpha\end{pmatrix}. \tag{16.48}$$

The coupling matrix (16.23) is

$$\Lambda=\begin{pmatrix}X & W\\ W & X\end{pmatrix} \tag{16.49}$$

where

$$X=-\tfrac{1}{2}(F^{-1}F'+E^{-1}E'),\ W=\tfrac{1}{2}(F^{-1}F'-E^{-1}E'). \tag{16.50}$$

Then the coupled equations (16.22) are

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}' = -\mathrm{i}\boldsymbol{B}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \boldsymbol{X}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \boldsymbol{W}\begin{pmatrix} c_3 \\ c_4 \end{pmatrix},$$
$$\begin{pmatrix} c_3 \\ c_4 \end{pmatrix}' = -\mathrm{i}\boldsymbol{B}\begin{pmatrix} c_3 \\ c_4 \end{pmatrix} + \boldsymbol{X}\begin{pmatrix} c_3 \\ c_4 \end{pmatrix} + \boldsymbol{W}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}. \tag{16.51}$$

where $\mathbf{c}$ is given by (16.21), (16.46). These exhibit the expected symmetry between the solution $c_1, c_2$ for upgoing waves and $c_3, c_4$ for downgoing waves. They can be used near a coupling point where $n_O = n_E$ because $\mathbf{U}$ in (16.46) was chosen to achieve this, and the coupling matrices $\boldsymbol{W}$, $\boldsymbol{X}$ (16.50) are small in a slowly varying medium. They cannot be used near a reflection point where $n_O = 0$ or where $n_E = 0$ because there the matrix $\boldsymbol{F}$ is singular and the coupling matrices $\boldsymbol{W}$, $\boldsymbol{X}$ are infinite.

## 16.5. Coupling and reflection points in the ionosphere

For the further study of coupled equations it is necessary to know the distribution of the coupling points in the complex $z$ (height) plane. Reflection points are a special case of coupling points and the term 'coupling point' is therefore used to include them. They are now to be discussed for a stratified ionosphere in which only electrons are effective. A coupling point is where two roots of the Booker quartic are equal. It is therefore where the discriminant $\Delta$, (6.39), (6.40), is zero. The equation $\Delta = 0$ was examined by Pitteway (1959) who showed that the solutions of interest are of eight types; see §6.6. A summary of their main properties is given here. A full discussion is given by Pitteway (1959), Jones and Foley (1972) and Smith, M.S. (1974a). It is here assumed that $Z$ is independent of height $z$, and the positions of the coupling points in the complex $X$ plane are studied. This is equivalent to the use of a model ionosphere in which $X$ increases linearly with height $z$.

Table 16.1

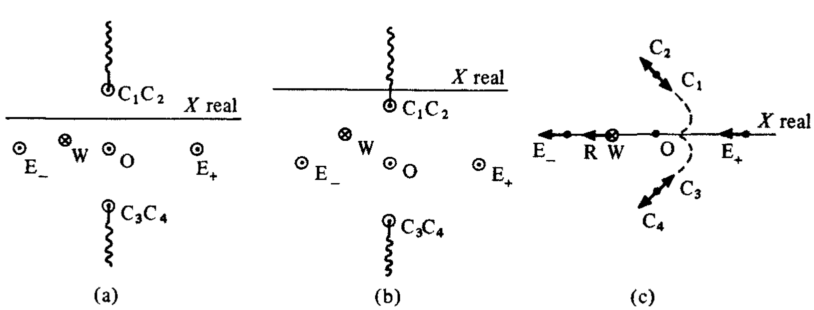
| Location in $X$ plane | Symbol | Description | |
|---|---|---|---|
| $U$ | O | Reflection. | Ordinary wave |
| $U - Y$ | $E_-$ | Reflection. | Extraordinary wave |
| $U + Y$ | $E_+$ | Reflection. | Extraordinary wave (Z-mode) |
| $U + \mathrm{i}Z_t$ | $C_1$ | Coupling. | Upgoing ordinary and extraordinary |
| $U - \mathrm{i}Z_t$ | $C_3$ | | |
| $U + \mathrm{i}Z_t$ | $C_2$ | Coupling. | Downgoing ordinary and extraordinary |
| $U - \mathrm{i}Z_t$ | $C_4$ | | |
| $X_\infty$ | R | Reflection, coinciding with point of resonance W | |

First consider vertical incidence. The eight types of coupling point are then as given in § 4.10 and are as in table 16.1. The positions of these points in the complex $X$ plane are shown in fig. 16.1. The first three are reflection points (4.74). The next four are where two values of $n^2$ are equal as given by (4.79). The last is where one value of $n^2$ is infinite, (4.76). The coupling point R and the resonance point W are really separate effects but they coincide for vertical incidence. For oblique incidence, when $l_x \neq 0$, they are at different positions.

The points $C_1$, $C_2$ coincide but they are associated with different coupling processes. Near $C_1$ the two upgoing waves are strongly coupled but they are propagated independently of the two downgoing waves, which are also strongly coupled near $C_2$. Similarly $C_3$ and $C_4$ coincide but are associated with independent coupling processes.

The various coupling points are branch points of the multiple valued functions $n(z)$ at vertical incidence, or $q(z)$ at oblique incidence. Here we continue to discuss vertical incidence. To aid the description it is useful to introduce branch cuts. The two values $\pm n_O(z)$ are equal (zero) at the reflection point O. This case is similar to the reflection point $z_0$ in an isotropic ionosphere, described in § 8.21. There a branch cut was used, shown as a wavy line in fig. 8.8. It runs from $z_0$ to infinity without crossing the real $z$ axis. For a study of the reflection of the ordinary wave a similar branch cut running from 0 to $\infty$ may be used. Similarly the two values $\pm n_E(z)$ are equal (zero) at the two reflection points $E_+$ and $E_-$ and branch cuts running from each of them to infinity, without crossing the real axis, may be used when studying the reflection of the extraordinary wave. The exact positions of these cuts are not important and their choice depends on the particular problem being solved. Since the refractive index of

Fig. 16.1. The complex $X$ plane showing the positions of the reflection and coupling points. W is the point of resonance where one refractive index is infinite. (a) and (b) are for vertical incidence. The wavy lines indicate branch cuts. In (a) $Z < Z_t$ and conditions are greater than critical. In (b) $Z > Z_t$ and conditions are less than critical. In (c), $Z = 0$. The arrows show how the coupling points move when the angle of incidence is increased from zero. The black dots are their positions at vertical incidence.



an upgoing wave is usually defined to be the value for which $\mathrm{Re}(n) > 0$, a convenient choice that is often used is to take as the branch cuts the lines whose equations are

$$\mathrm{Re}\{n_{\mathrm{O}}(z)\} = 0, \quad \mathrm{Re}\{n_{\mathrm{E}}(z)\} = 0. \tag{16.52}$$

The refractive indices $n_{\mathrm{O}}$, $n_{\mathrm{E}}$ for the two upgoing waves are equal at $\mathrm{C}_1$ and at $\mathrm{C}_3$, and branch cuts running from these points to infinity are shown as wavy lines in fig. 16.1. Again the exact positions of these cuts are not important and their choice is a matter of convenience. In § 4.15 the terms ordinary and extraordinary were defined as given by $\mathrm{Re}(S_{\mathrm{R}})$ positive and negative respectively where $S_{\mathrm{R}}$ is given by (4.93). Thus a convenient choice for the branch cuts is the lines whose equations are

$$\mathrm{Re}(S_{\mathrm{R}}) = 0 \tag{16.53}$$

so that $S_{\mathrm{R}}^2$ in (4.93) is real and negative. These lines run from $\mathrm{C}_1$ in the direction away from $\mathrm{C}_3$, and from $\mathrm{C}_3$ away from $\mathrm{C}_1$, as in fig. 16.1.

In fig. 16.1a, $Z < Z_{\mathrm{t}}$ and the points $\mathrm{C}_1$, $\mathrm{C}_2$ are on the positive imaginary side of the real $X$ axis. This is the usual case for high frequencies, for then $Z$ is small, and conditions are said to be 'greater than critical'. When $Z = Z_{\mathrm{t}}$ the points $\mathrm{C}_1$, $\mathrm{C}_2$ are on the real $X$ axis and this is called 'critical coupling'; see § 16.6. For $Z > Z_{\mathrm{t}}$ as in fig. 16.1(b), all the coupling points, including $\mathrm{C}_1$, $\mathrm{C}_2$, are on the negative imaginary side of the real $X$ axis and conditions are said to be 'less than critical'. Instead of the terms 'greater than...' and 'less than critical', some authors use the terms 'quasi-transverse' and 'quasi-longitudinal' respectively, and 'critical coupling' is called the 'QL–QT transition'. The terms 'quasi-longitudinal' and 'quasi-transverse' are also used to describe certain approximations in magnetoionic theory; see § 4.17. The two usages are related but not exactly the same. The concepts of greater than and less than critical do not involve approximations.

When $X$ is a linear function of $z$, fig. 16.1 shows how the coupling points are arranged in the complex $z$ (height) plane. When $X$ is a more complicated function of $z$, the distribution of coupling points is less simple. For example if $X \propto \mathrm{e}^{\alpha z}$, and if $z_0$ is the position of some coupling point, then the same kind of coupling point appears again at $z = z_0 + 2\pi i r/\alpha$ where $r$ is any positive or negative integer. Thus every type of coupling point occurs an infinite number of times. Usually it is only those that are nearest to the real $z$ axis that are important. Some examples of this case were given by Budden (1961a, figs. 20.6, 20.8). In practice $Z$ also is a function of $z$ and this affects the positions of the coupling points.

Next let the wave be obliquely incident. The behaviour of the coupling points now depends on many factors and is complicated. The description here is given for one case only, needed in later sections. For other cases see Pitteway (1959), Budden and Terry (1971), Jones and Foley (1972), Smith, M.S. (1974a). It is here assumed that the plane of incidence is the magnetic meridian plane so that $l_y = 0$, and that collisions are neglected, $Z = 0$, and that $Y < 1$ as in figs. 6.4–6.6. The angle of incidence $\theta =$

arcsin $S$ is assumed to be real and increases from zero so that the wave normal of the incident wave moves nearer to the direction of the earth's magnetic field, that is of $Y$. Some of the features of the process can be followed in figs. 10.4, 10.7. The reference line AB or MN is at first through the origin and moves to the right. See also figs. 6.4–6.7. For all values of $S$, the resonance point W remains fixed at $X = (1 - Y^2)/(1 - Y^2 l_z^2)$. When $S = 0$ the points $E_+$, $E_-$, O and R (same as W) are on the real $X$ axis as shown in fig. 16.1(c). When $S$ increases slightly, O does not move; the line A′B′, fig. 10.4, still intersects one of the refractive index surfaces for the ordinary wave where $X = 1$, and the wave is here reflected with a 'Spitze' in its ray, § 10.9. The point R moves to the left of W and can be seen in the insets of figs. 6.4, 6.5. The points $E_+$, $E_-$ move to smaller values of $X$. The points $C_1$, $C_2$ now separate, and $C_3$, $C_4$ also separate, as shown by arrows in fig. 16.1(c). When $S$ increases further, $C_1$ and $C_3$ move on to the real $X$ axis and coincide there. At this stage three roots of the Booker quartic are equal. The value of $S$ is between the values used in figs. 6.6(a, b), and is such that the upper continuous curve has a vertical tangent at its point of inflection. When $S$ increases still further, $C_1$, $C_3$ separate again but remain on the real $X$ axis, fig. 6.6(b). At this stage the line A′B′ in fig. 10.4 is still to the left of P and the reflection point O is still where $X = 1$. The ordinary wave is still reflected with a 'Spitze'. With $S$ increasing, $C_3$ moves to the left and eventually coincides with O, at $X = 1$; fig. 6.6(c). This happens when $S = S_A = l_x\{Y/(Y+1)\}^{\frac{1}{2}}$, (6.47), fig. 6.3. The reference lines A′B′ in fig. 10.4 and MN in fig. 10.7 now go through the window point P. The coupling points O and $C_3$ here both apply to the same two roots of the Booker quartic. This is an example of a coalescence C2, studied later in §§ 17.4, 17.6. It is associated with the phenomenon of radio windows, §§ 17.6, 17.7. When $S$ increases to a value greater than $S_A$, the point O moves to where $X < 1$ and there is no longer a Spitze in the ordinary ray. Now $C_3$ remains at $X = 1$. The point $C_1$ moves to smaller $X$ and eventually $C_1$ and $C_3$ coincide at $X = 1$, fig. 6.6(e), when $S = \sin\Theta$, where $\Theta$ is the angle between $Y$ and the vertical. Thereafter $C_1$ remains at $X = 1$, and $C_3$ moves down to where $X < 1$, fig. 6.6(f).

The sequence just described is modified when $l_y \neq 0$, when $Z \neq 0$ and when $Y > 1$. These cases are all discussed by Smith, M.S. (1974a).

Another example of coupling points has been mentioned in § 13.9. They are analogous to $C_1$, $C_3$ and denoted by $z = z_p, z_q$ respectively. They are there used in the theory of ion cyclotron whistlers, in which the effects of heavy positive ions and collisions are allowed for.

## 16.6. Critical coupling

The term 'critical coupling' is customarily used only for vertically incident waves. It refers to one of the two coupling points $C_1$, $C_2$ described in § 16.5. $C_1$ is associated with the two upgoing waves and $C_2$ with the two downgoing waves, and for vertical incidence they are at the same position in the complex $z$ plane. This position depends on frequency and on the functions $X(z)$, $Z(z)$. Critical coupling occurs when $C_1$ and $C_2$ are on the real $z$ axis. The points $C_1$ and $C_2$ are where $n_O = n_E$ and the condition for this was discussed in §§ 4.4, 4.10. $X(z)$ and $Z(z)$ are both real, and if only electrons are allowed for, the condition is (4.26)–(4.28). For frequencies of a few hundred hertz or less, where heavy ions are important, critical coupling can also occur. It is associated with the theory of ion cyclotron whistlers and has been discussed in § 13.9. In this case critical coupling is also known as 'exact crossover', and the condition for it is (13.60)–(13.62). See also fig. 13.12(c), where $z_p$ denotes the coupling points $C_1$ and $C_2$. In this section we deal only with the first case where only electrons are considered.

At critical coupling (4.28) shows that the collision frequency $\nu$ has the transition value $\omega_t = \frac{1}{2}\omega_H \sin^2 \Theta / |\cos \Theta|$. This is independent of frequency. The function $\nu(z)$ is believed to be a monotonically decreasing function so that it attains the value $\omega_t$ at one fixed height, to be called the 'transition height'. In temperate latitudes $\omega_t$ is about $9 \times 10^5\,\mathrm{s}^{-1}$ and the transition height is about 90 km. The second condition in (4.26) is that at the transition height $X = 1$, and this depends on frequency. Thus in general the relations (4.26)–(4.28) are not simultaneously true at any real height. There is, however, always one frequency, the plasma frequency at the transition height, that makes $X = 1$. It is called the 'critical coupling frequency', and denoted by $f_c$. It depends on the state of the ionosphere and therefore on time of day, season and geographical position. In temperate latitudes in the day time it is usually in the range 200 kHz to 1 MHz and is rarely less than 150 kHz. At night it is smaller, of the order of 20 to 50 kHz in temperate latitudes. In the tropics, however, the transition height is lower and $f_c$ is smaller.

For frequencies greater than $f_c$, conditions are greater than critical. The coupling points $C_1$, $C_2$ are on the positive imaginary side of the real $z$ axis, fig. 16.1(a). This is the usual condition when probing the ionosphere with frequencies greater than about 1 MHz. The branch cut in fig. 16.1(a) does not cross the real $z$ axis and each wave remains either ordinary or extraordinary at all heights, without a change of name. The ordinary wave is reflected near where $X = 1$ and the extraordinary near where $X = 1 - Y$, and these two reflections are easily observed, and have different equivalent heights.

For frequencies less than $f_c$, conditions are less than critical, and this is the regime in temperate latitudes for frequencies less than about 50 kHz at night, and less than about 500 kHz to 1 MHz in the day time. The coupling points $C_1$, $C_2$ are now on the negative imaginary side of the real $z$ axis, and the branch cut that ends at $C_1$, $C_2$ now crosses the real $z$ axis, fig. 16.1(b). Consider, first, a wave that travels vertically upwards from the ground, whose polarisation after it enters the ionosphere is that of an ordinary wave. This wave has a reflection point where $X = 1 - \mathrm{i}Z$. But, at these low frequencies, $Z = \nu/2\pi f$ is large so that the real and imaginary parts of $X$ at the reflection point are comparable. The effect of this is that the reflection coefficient is very small and this reflection is rarely observed. Because of the collisions, $\mu = \mathrm{Re}(n_O)$ does not go near to zero at the real height where $X = 1$. This effect can be seen in figs. 4.7–4.9, 16.5, 16.6 which are for conditions less than critical. For other curves showing this see, for example, Lepechinsky (1956, fig. 3a), Ratcliffe (1959, figs. 7.4, 10.3), Budden (1961a, fig. 6.14), Rawer and Suchy (1967, fig. 31). At the level where $X = 1$ the condition $Z > Z_t$ now holds and the real $z$ axis crosses the branch cut. The value of $n_O$ at a point on one side of the cut is the same as $n_E$ at the adjacent point on the other side. The field components are the same at these two adjacent points. Thus where $X > 1$ the wave is an extraordinary wave. This is simply a change of name, as explained in §4.16, and is not the result of any mode conversion process. The wave travels on up to near where $X = 1 + Y - \mathrm{i}Z$ and here it is reflected. At this greater height $Z$, though still large, is less than $1 + Y$. This means that $n_E(z)$ does decrease towards zero near this level, and the resulting reflection of the extraordinary wave is great enough to be observed; see, for example, fig. 16.5–16.7. The downgoing extraordinary wave again passes the level where $X = 1$ and its name changes back to 'ordinary wave' below this. Thus the reflected wave observed at the ground has the polarisation of an ordinary wave, but is reflected as an extraordinary wave. This wave is called an 'initial ordinary' wave (Cooper, 1961). For very low frequencies it is the most commonly observed reflection.

Consider next a wave that travels vertically upwards from the ground and whose polarisation after it enters the ionosphere is that of an extraordinary wave. One reflection point for this is where $X = 1 - Y - \mathrm{i}Z$ but since $Y > 1$ this point is not near the real $z$ axis. Cooper (1961, 1964) has shown that it can give some reflection but this is too weak to be important. Now follow this wave upwards along the real $z$ axis. When it goes above the level where $X = 1$ it is an ordinary wave because here $Z > Z_t$, and the branch cut is crossed. It is not therefore reflected near where $X = 1 + Y - \mathrm{i}Z$ but travels on upwards into the magnetosphere as an ordinary wave, that is a whistler mode wave.

The upgoing extraordinary wave can also be followed along a path in the complex $z$ plane that runs on the negative imaginary side of the coupling point $\mathrm{C}_1$. This is a branch point of $n_O(z)$ and $n_E(z)$, and by going on the other side of it we ensure that the branch cut is not crossed and the wave does not change its name but remains an extraordinary wave where $\mathrm{Re}(X) > 1$. Thus on the real axis where $X > 1$ there is an extraordinary wave, generated in the region near $\mathrm{C}_1$ by a mode conversion process. The original unconverted wave is extraordinary where $X < 1$ and ordinary where $X > 1$. Paradoxically this initially extraordinary wave where $X < 1$ gives rise by mode conversion to an extraordinary wave where $X > 1$. An estimate of its amplitude can be made by the phase integral method, § 16.7, with the path running on the negative imaginary side of $C_1$. This extraordinary wave is now reflected where $X = 1 + Y - iZ$, and the reflected extraordinary wave travels downwards. If we follow it along the real $z$ axis, its name changes to ordinary on going below where $X = 1$, and it reaches the ground as an ordinary wave. But it also gives rise, by mode conversion near the coupling point $C_2$, to a wave that is extraordinary where $X < 1$, and this too reaches the ground.

It is now interesting to enquire what would be observed if conditions changed continuously from greater than critical to less than critical. This might happen if the transmitter frequency was continuously decreased, or if the frequency was fixed and the electron concentration $N(z)$ was increasing. In describing the phenomenon we shall suppose that the ionosphere does not change and that the transmitter frequency is slowly decreased. It is assumed that $N(z)$ and therefore $X(z)$ are monotonically increasing functions of the height $z$. The critical coupling frequency $f_c$ has some fixed value less that the electron gyro-frequency $f_H$. This discussion applies only for vertical incidence. The transmitted wave is assumed to be linearly polarised so that on entering the ionosphere it splits into ordinary and extraordinary waves with roughly equal amplitude.

For frequencies less than $f_H$ the ordinary wave reflection from where $X = 1 - iZ$ is weak, as explained above. For all frequencies, therefore, the strongest reflection is of the extraordinary wave from where $X = 1 + Y - iZ$. For frequencies $f \gg f_c$ this returns to the ground as an extraordinary wave. On emerging from below the ionosphere its polarisation is the limiting polarisation $\rho_E$ of the extraordinary wave. This is discussed in § 17.11 where it is shown that $\rho_E$ is given by (4.21) for some limiting level in the base of the ionosphere where $X \approx 0$. Thus

$$\rho_E = i[Z_t + \{(1 - iZ)^2 + Z_t^2\}^{\frac{1}{2}}]/(1 - iZ) \tag{16.54}$$

where the real part of the square root is positive, and $Z$ has its value at the limiting level.

For frequencies $f \ll f_c$ the reflected extraordinary wave is an ordinary wave when it goes below where $X = 1$ and so the strongest reflection reaching the ground has the limiting polarisation $\rho_O$ of an ordinary wave, given by

$$\rho_O = i[Z_t - \{(1 - iZ)^2 + Z_t^2\}^{\frac{1}{2}}]/(1 - iZ). \tag{16.55}$$

Other reflected waves can be produced by mode conversion and can reach the ground but they are weak when $f$ is not near $f_c$.

Thus there should be a change of polarisation from (16.54) to (16.55) as the frequency is decreased through $f_c$. There is a transition range of frequency near $f_c$ where waves of both polarisations reach the ground and the resultant polarisation would have to be found by combining these waves. The extent of this transition range is not known with certainty. If it is small, then as the transmitter frequency is decreased the polarisation of the received wave would switch over quickly from

(16.54) to (16.55). These two polarisations are elliptical with opposite senses. The two polarisation ellipses are as shown in fig. 4.2(a).

The author has heard this effect described by colleagues but does not know of any published account of observations of it. It has been called 'flipover'. This name may have been used in describing experiments where the polarisation ellipse of the received wave was delineated on the screen of a cathode ray oscillograph. When conditions changed from greater to less than critical, the observed ellipse flipped over from (16.54) to (16.55).

For measurements with vertically incident radio pulses of frequency 150 kHz at night, changes of polarisation have been observed (Parkinson, 1955) but they are more complicated than could be explained by the 'flipover' described above. They are linked with the occurrence of a 'coupling echo', which is an extra reflection from the level where $X \approx 1$; see § 16.14.

The theory of the transition through critical coupling has been studied by Davids and Parkinson (1955), Altman, Cory and Fijalkow (1970), Altman and Fijalkow (1970), Altman and Postan (1971), Smith, M.S. (1973a). A related topic was discussed by Lepechinsky (1956), and by Landmark and Lied (1957). Although the subject is complicated, the results of all these authors seem to be compatible with the conclusion that a change of polarisation of the reflection from near $X \approx 1 + Y$ would be expected when conditions change from less than to greater than critical, but the change would not occur suddenly.

## 16.7. Phase integral method for coupling

The phase integral formula for finding ionospheric reflection coefficients was described in § 8.21. It was based on the use of an Airy integral function in the uniform approximation solution (8.72) of the differential equations. This solution was for an isotropic ionosphere and for horizontally polarised waves but it was explained near the end of § 8.21 that the method can also be applied for an anisotropic ionosphere, and for the reflection of the ordinary or the extraordinary wave. A similar solution for the neighbourhood of a coupling point has been given at (16.30) and this too contains an Airy integral function. It is very similar to (8.72) and can be used to derive a phase integral formula for coupling. To illustrate the method it is described here for vertical incidence and for upgoing waves. A similar treatment is possible for downgoing waves, and, more generally, for oblique incidence.

Consider first the case where $Z < Z_t$ so that conditions are 'greater than critical'; see § 16.5. The coupling point $C_1$ is then on the positive imaginary side of the real $z$ axis, and is assumed to be so close to it that the W.K.B. approximations are not usable on the real $z$ axis near $C_1$; see fig 16.2(a). This part of the real axis is called the 'coupling region'. It is assumed that at real heights below this the only wave present is an upgoing ordinary wave. It is required to study the waves at real heights above the coupling region and to find the amplitudes there of the ordinary and extraordinary waves. These upgoing waves are reflected at the appropriate levels in the ionosphere. The resulting downgoing waves may also be present near $C_1$ but they are assumed to be independently propagated and they can therefore be ignored.

Now take $q_1 = n_O, q_2 = n_E$ so that, from (16.17), $\beta = \frac{1}{2}(n_O - n_E)$ and from (16.29) the $\zeta$ used in Ai($\zeta$) (16.30) is

$$\zeta = \left\{\frac{3}{4}\mathrm{i}k\int_{z_p}^{z}(n_O - n_E)\mathrm{d}z\right\}^{2/3} \tag{16.56}$$

where $z_p$ is the value of $z$ at $C_1$. The Stokes lines are where $\arg\zeta = 0, \pm\frac{2}{3}\pi$ and the anti-Stokes lines are where $\arg\zeta = \pm\frac{1}{3}\pi, \pi$ so that their equations are

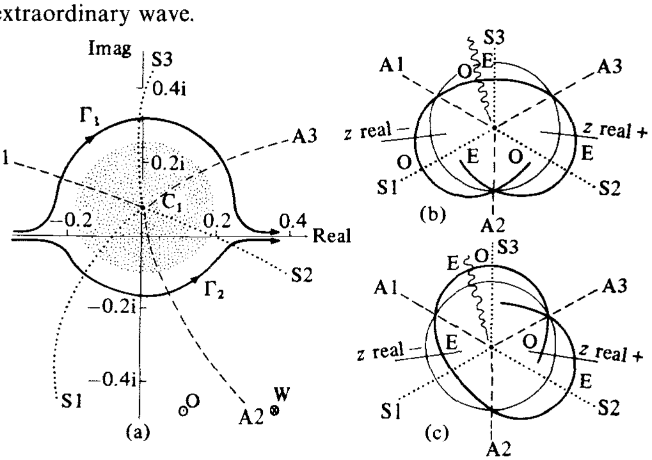
$$\text{Stokes:}\quad \mathrm{Re}\int_{z_p}^{z}(n_O - n_E)\mathrm{d}z = 0;\quad \text{anti-Stokes:}\quad \mathrm{Im}\int_{z_p}^{z}(n_O - n_E)\mathrm{d}z = 0. \tag{16.57}$$

These can be plotted in the complex $z$ plane, and examples are shown in fig. 16.2(a) (compare fig. 8.9(a)). For the upgoing ordinary wave the variables $\mathrm{c}_1, \mathrm{c}_2$ of (16.21) contain the exponential factor $\exp(-ik\int_0^z n_O\mathrm{d}z)$ and thence from (16.25) the variables $h_1, h_2$ contain the factor

$$\text{ordinary wave:}\quad \exp\left\{-\tfrac{1}{2}\mathrm{i}k\int^{z}(n_O - n_E)\,\mathrm{d}z\right\}. \tag{16.58}$$

Similarly for the extraordinary wave the variables $h_1$, $h_2$ contain an exponential factor (16.58) with $+\frac{1}{2}$ instead of $-\frac{1}{2}$. These are the exponential factors in the two terms used in the asymptotic approximation for Ai($\zeta$).

Fig. 16.2. (a) is the complex $z$ (height) plane and shows the Stokes lines S and anti-Stokes lines A radiating from the coupling point $C_1$. In this example $X = \mathrm{e}^{\alpha z}$ and the numbers by the axes are the values of $\alpha z$. $Y = 4$, $\Theta = 30°$, $Z = 0.5$ and $Z_t = 0.557$, so that conditions are greater than critical. (b) is the Stokes diagram when the incident wave from below is an ordinary wave, and (c) is the same when it is an extraordinary wave.



Suppose now that $Y > 1$, as in the example of fig. 16.2. On the real $z$ axis where $X \ll 1$, the incident ordinary wave, containing (16.58), is the dominant term. This can be seen as follows. For $X \ll 1$ the example of fig. 4.6 shows that $\mathrm{Re}(n_E) > \mathrm{Re}(n_O)$. Here let $z$ be given a positive imaginary increment so that the anti-Stokes line A1, fig. 16.2(a), is approached. The real part of the exponent in (16.58) then has a negative increment, so the modulus of the term decreases as A1 is approached. This shows that the term is dominant before A1 is crossed and subdominant after it is crossed. If $Y < 1$ a slightly different argument is needed but again it can be shown that the ordinary wave is the dominant term on the negative imaginary side of A1 (Budden 1961a, § 20.7). The function $\zeta$, (16.56), contains a cube root and has three values. The value to be chosen is that for which $\mathrm{Ai}(\zeta)$ has only a dominant term when $z$ is real and $X < 1$. The Stokes diagram for this $\mathrm{Ai}(\zeta)$ can now be drawn and is shown in fig. 16.2(b).

The W.K.B. solution below the coupling region may be written in the form (7.111), thus

$$\mathsf{f}_1 = \mathsf{K}_1 \exp\left\{ -ik \int_0^z n_O(z)\,\mathrm{d}z \right\} \tag{16.59}$$

where $\mathsf{K}_1$ is a constant. Let this solution be followed as we move along a path, $\Gamma_1$ in fig. 16.2(a), that goes on the positive imaginary side of $\mathrm{C}_1$ and is far enough from it for the W.K.B. solutions to be good approximations at all points on it. This solution is at first dominant. It becomes subdominant after A1 is crossed. It does not undergo the Stokes phenomenon when the Stokes line S3 is crossed because there is no dominant term present here. When the branch cut is crossed the name changes to extraordinary and $n_O$ changes to $n_E$ but (16.59) is continuous here. This solution, now an extraordinary wave, again becomes dominant after A3 is crossed and the real $z$ axis is again reached. Thus this W.K.B. solution is continuous for the whole path $\Gamma_1$. The ratio of the values of (16.59) at two points $z_a$, $z_b$ above and below the coupling region is

$$\mathsf{f}_1(z_a)/\mathsf{f}_1(z_b) = \exp\left\{ -ik \int_{z_b\,(\Gamma_1)}^{z_a} n(z)\,\mathrm{d}z \right\} \tag{16.60}$$

where $n$ means $n_O$ before the cut is crossed and $n_E$ after it is crossed. The values of $\mathsf{f}_1$ give the field components of the wave, from (6.53).

The formula (16.60) gives the amplitude of the extraordinary wave above the coupling region, produced by mode conversion from the incident ordinary wave below. It is an example of the phase integral formula for coupling, and was first given by Eckersley (1950).

The solution (16.59) can also be followed along a path, $\Gamma_2$ in fig. 16.2(a), that goes on the negative imaginary side of $\mathrm{C}_1$. This does not cross the branch cut so the wave remains an ordinary wave. The Stokes diagram, fig. 16.2(b), shows that this term now becomes subdominant, and disappears because of the Stokes phenomenon, after the Stokes line S2 is crossed and before the real $z$ axis is again reached where $X > 1$. This is above the level where $X = 1 - \mathrm{i}Z$ associated with reflection of the ordinary wave. Here the wave is evanescent and therefore strongly attenuated. At greater heights its amplitude would be extremely small so that even if present it would be of little interest.

When the incident wave is an extraordinary wave, let its W.K.B. solution below the coupling region be

$$\mathsf{f}_2 = \mathsf{K}_2 \exp\left\{-ik\int_0^z n_{\mathrm{E}}(z)\,\mathrm{d}z\right\}. \tag{16.61}$$

The phase integral method for coupling can be used in a similar way. The Stokes diagram is now as shown in fig. 16.2(c). If the path $\Gamma_2$ is used, the wave remains an extraordinary wave and does not undergo the Stokes phenomenon. Its amplitude $\mathsf{f}_2^{(\mathrm{E})}(z_{\mathrm{a}})$ above the coupling region is given by

$$\mathsf{f}_2^{(\mathrm{E})}(z_{\mathrm{a}})/\mathsf{f}_2(z_{\mathrm{b}}) = \exp\left\{-ik\int_{z_{\mathrm{b}}\,(\Gamma_2)}^{z_{\mathrm{a}}} n_{\mathrm{E}}(z)\mathrm{d}z\right\}. \tag{16.62}$$

If the path $\Gamma_1$ is used, again the solution does not undergo the Stokes phenomenon, but its name changes to ordinary after the branch cut is crossed. Its amplitude above the coupling region is given by

$$\mathsf{f}_2^{(\mathrm{O})}(z_{\mathrm{a}})/\mathsf{f}_2(z_{\mathrm{b}}) = \exp\left\{-ik\int_{z_{\mathrm{a}}\,(\Gamma_1)}^{z_{\mathrm{b}}} n(z)\,\mathrm{d}z\right\} \tag{16.63}$$

where $n$ means $n_{\mathrm{E}}$ before the cut is crossed and $n_{\mathrm{O}}$ afterwards. This formula gives the amplitude of the ordinary wave produced by mode conversion from the incident extraordinary wave. It is likely to be small because the ordinary wave here is evanescent.

The method can similarly be used to deal with conditions that are less than critical. The coupling point $\mathrm{C}_1$ in fig. 16.2(a) would then be on the negative imaginary side of the real axis. The details are not given here.

The path $\Gamma_1$ is a 'good path' because the W.K.B. solutions are good approximations at all points on it. Provided that such a path exists, it is not necessary to use it in (16.60) or (16.63). The contour can be distorted in any convenient way provided that no singularities of $n(z)$ are crossed. In particular it can be moved as close as desired to the point $\mathrm{C}_1$. An example of this is given in the following section.

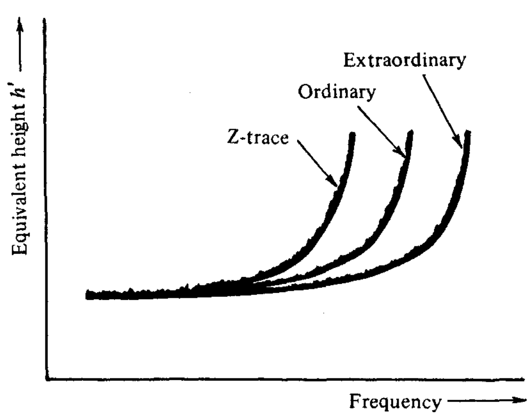
## 16.8. The Z-trace

When pulses of radio waves are reflected from the ionosphere at normal incidence and observed at the ground, using the ionosonde technique, §§ 1.7, 13.4, records of $h'(f)$ are obtained that normally consist of two curves, one for the ordinary and one for the extraordinary wave (figs. 13.2–13.4). Sometimes, especially at high latitudes, a third curve is observed in which $h'(f)$ for the F-layer has a greater value than for either of the other curves (Harang, 1936; Meek, 1948; Newstead, 1948; Toshniwal, 1935). The form of the $h'(f)$ record near the penetration frequency of the F-layer is then as sketched in fig. 16.3. The third echo is usually known as the Z-trace. Its polarisation shows that when it leaves the ionosphere it is an ordinary wave, but its equivalent height shows that it must have been reflected near the level where $X = 1 + Y$. In topside ionograms the Z-trace normally appears, fig. 13.6, and can be explained by simple ray theory, § 13.5, because the transmitter and receiver are immersed in the plasma. But this is not possible for ionograms observed at the ground since the transmitter and receiver are in free space.

If conditions were less than critical, § 16.5, the Z-trace would be observed at the ground with the polarisation of an ordinary wave, and the ordinary wave reflection from near where $X = 1$ would be either weak or absent. But this could only happen near a magnetic pole where $\sin\Theta$ and thence $Z_t$ are small so that the transition height is above the maximum of the F-layer. For a detailed study of the theory of the ordinary wave reflection from near $X = 1$ in these conditions see Budden and Smith (1973). For ionograms observed at the ground in latitudes not near a pole, the transition height is well below the maximum of the F-layer and the critical coupling frequency is much less than the penetration frequency of the F-layer, so that conditions must be greater than critical. Then it is not possible to explain the Z-trace by simple ray theory.

It has been suggested (e.g. Rydbeck, 1950, 1951) that the effect is produced by a twofold mode conversion process, for the upgoing and downgoing waves, associated with the coupling points $C_1$ and $C_2$ respectively.

Fig. 16.3. Sketch to show the appearance of the Z-trace in an ionogram observed from the ground.



It is instructive to estimate the order of magnitude of the Z-trace reflection assuming that this is the correct explanation. We use the phase integral formula (16.60) for the amplitude of the extraordinary wave produced by mode conversion from the incident ordinary wave. The frequency is near the penetration frequency of the F-layer so that $Z$ is small. For simplicity we take $Z = 0$ so that on the real $z$ axis $n_{\mathrm{O}}$ is real for $X < 1$, and $n_{\mathrm{E}}$ is real for $X > 1$. The contour $\Gamma_1$, fig. 16.2(a), can be distorted to coincide with the path $\mathrm{AQC_1QB}$ as sketched in fig. 16.4. The real part of the exponent in (16.60) is then determined entirely by the two sides of the line $\mathrm{QC_1}$. Assume that the electron concentration varies linearly with height $z$ so that $X = \alpha z$. Let $\mathscr{G}$ be the modulus of (16.60). Then

$$\mathscr{G} = \exp\left[\mathrm{Re}\left\{\frac{-\mathrm{i}k}{\alpha}\int_{(Q)}^{(C_1)}(n_{\mathrm{O}} - n_{\mathrm{E}})\,\mathrm{d}X\right\}\right]. \tag{16.64}$$

At Q, $X = 1$ and at $\mathrm{C_1}$, $X = 1 + \mathrm{i}Z_{\mathrm{t}}$. At intermediate points let $X = 1 + \mathrm{i}\zeta$. Then

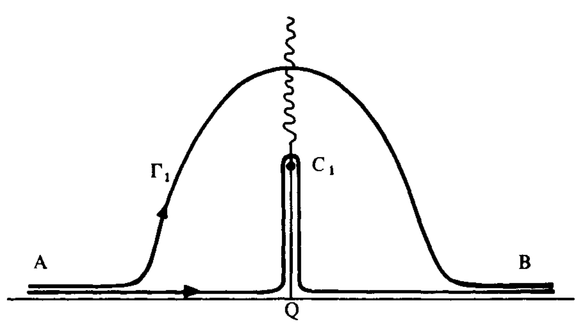
$$\mathscr{G} = \exp\left\{\frac{k}{\alpha}\int_0^{Z_{\mathrm{t}}}\mathrm{Re}(n_{\mathrm{O}} - n_{\mathrm{E}})\,\mathrm{d}\zeta\right\}. \tag{16.65}$$

When $\zeta = 0$, $n_{\mathrm{O}} - n_{\mathrm{E}} = -1$ and when $\zeta = Z_{\mathrm{t}}$, $n_{\mathrm{O}} - n_{\mathrm{E}} = 0$. Rydbeck (1950, 1951) and Pfister (1953) have given curves that show that $\mathrm{Re}(n_{\mathrm{O}} - n_{\mathrm{E}})$ varies smoothly and monotonically in the range $0 \leqslant \zeta \leqslant Z_{\mathrm{t}}$. Hence the integral in (16.65) is of order $-Z_{\mathrm{t}}$, and approximately

$$\mathscr{G} \approx \exp(-kZ_{\mathrm{t}}/\alpha) = \exp(-\omega_{\mathrm{t}}/\alpha c). \tag{16.66}$$

For temperate latitudes take $\omega_{\mathrm{t}} \approx 10^6\,\mathrm{s}^{-1}$, $\alpha = 0.1\,\mathrm{km}^{-1}$. Then $\mathscr{G} \approx \mathrm{e}^{-33} = 4.7 \times 10^{-15}$. To give the amplitude of the Z-trace this figure must be squared since the coupling process occurs twice. Hence the amplitude of the Z-reflection would be far too weak to be detected. But since it has often been observed even in low latitudes, for example in India (Toshniwal, 1935), there must be some other explanation. One possibility is that occasionally there can be a very steep gradient of electron

Fig. 16.4. The complex $z$ plane, showing possible contours for use with the phase integral formula (16.56).



concentration near the coupling level, so that $\alpha$ has a much larger value than was assumed above.

A more likely explanation (Ellis, 1956) is that the Z-trace arises from waves that travel obliquely so that in the coupling region the wave normal is nearly parallel to the earth's magnetic field. Then the two coupling points $C_1$, $C_3$ are very near together and approach coalescence, fig 16.1(c). There is a radio window, the 'Ellis window', at the level where $X = 1$ and the upgoing ordinary wave goes through it and is an extraordinary wave on the upper side. The theory of this effect is given in §§ 17.6, 17.7. This extraordinary wave would not return to a receiver near the transmitter after reflection. But if there are irregularities of electron concentration near the level where $X = 1 + Y$, some energy can be scattered and return along the same path. If the refractive index within an irregularity differs by $\Delta n_E$ from its value $n_E$ in the surrounding plasma, the scattered amplitude is proportional to $\Delta n_E/n_E$ (Booker, 1959). This is largest near the level $X = 1 + Y$ where $|n_E|$ is smallest.

## 16.9. Additional memory

In this section the summation convention, §§ 2.11, 14.8, for repeated suffixes is not used.

The W.K.B. solution for the general anisotropic stratified medium was given at (7.112). Apart from the constant $\mathsf{K}_j$ it contained two factors $\mathbf{s}_j$ and $\exp(-ik\int_0^z q_j\,dz)$. The first is the $j^{\text{th}}$ column of $\mathbf{S}$. It is given, for example, by (7.134) and it is then a function of the composition of the medium at the height $z$ only. It does not depend on how the medium varies with $z$. It will be called a 'local' property of the medium. The other factor has been called the 'phase memory' term; § 7.3. The exponent gives the change of complex phase of the wave in its passage to the point $z$. It depends on the whole of the medium that the wave has traversed, and is not a local property. In these solutions, therefore, it is important to distinguish between 'local' and 'memory' factors.

Because of the coupling process, one characteristic wave, with subscript $i$ say, gives rise as it travels to some of the other characteristic waves. The amplitude of wave $j$ thus generated in any element $\delta z_1$ at $z = z_1$ on the path of wave $i$ is proportional to $\Gamma_{ij}\delta z_1 \exp(-ik\int_0^{z_1} q_i\,dz)$. This infinitesimal wave $j$ then travels to another point $z_2$ with a phase memory factor $\exp\{-ik(\int_0^{z_1} q_i dz + \int_{z_1}^{z_2} q_j dz)\}$. The original wave $i$ reaches $z_2$ with a phase memory factor $\exp(-ik\int_0^{z_2} q_i dz)$. Thus the complex phase difference at $z_2$ between the parent wave $i$ and the infinitesimal wave $j$ is

$$\phi_{ij} = k\int_{z_1}^{z_2} (q_i - q_j)\,dz. \tag{16.67}$$

Now at $z_2$ many of these infinitesimal waves $j$ have to be added for all contributing

$z_1$. This could be done by an amplitude–phase diagram, which would give a spiral, and the resultant is small even when $\Gamma_{ij}$ is moderately large, because of the partial destructive interference of the contributions with the variable phase (16.67).

Consider now the wave generated in $\delta z_1$ through the action of the element $\Gamma_{jj}$ of the coupling matrix. It is of the same type as the parent wave $j$ and travels with the same phase memory factor. At $z_2$ the phase difference between it and the parent wave is $\phi_{jj} = 0$ from (16.67). When the contributions are added, therefore, they all have the same phase and reinforce each other so that the resultant may be large even when $\Gamma_{jj}$ is small. This shows that the diagonal elements $\Gamma_{jj}$ have to be treated differently from the non-diagonal elements. It was suggested in §7.15 that $k\Gamma_{jj}$ should be included with $-ikq_j$ in the phase memory term as in (7.117). This suggestion must now be examined.

The coupling matrix $\mathbf{\Gamma} = -\mathbf{S}^{-1}\mathbf{S}'$ was given at (16.5), (16.6) and this form used the version (7.134) for $\mathbf{S}$. But a different $\mathbf{S}$ can be used by normalising the columns differently, as explained at the end of § 7.15; for an example see § 16.3. This means that a new version $\check{\mathbf{S}}$ of $\mathbf{S}$ can be found by multiplying by an arbitrary diagonal matrix $\mathbf{D}$ on the right, and the variable $\mathbf{f} = \mathbf{S}^{-1}\mathbf{e}$ in (16.53) must be replaced by $\check{\mathbf{f}}$ thus

$$\check{\mathbf{S}} = \mathbf{SD}, \quad \check{\mathbf{S}}^{-1} = \mathbf{D}^{-1}\mathbf{S}^{-1}, \tag{16.68}$$

and

$$\check{\mathbf{f}} = \check{\mathbf{S}}^{-1}\mathbf{e} = \mathbf{D}^{-1}\mathbf{S}^{-1}\mathbf{e}, \quad \mathbf{e} = \mathbf{SD}\check{\mathbf{f}}. \tag{16.69}$$

Now $\mathbf{\Gamma}$ is replaced by

$$\check{\mathbf{\Gamma}} = \mathbf{D}^{-1}\mathbf{\Gamma}\mathbf{D} - \mathbf{D}^{-1}\mathbf{D}'. \tag{16.70}$$

The diagonal elements of $\mathbf{\Gamma}$ and $\mathbf{D}^{-1}\mathbf{\Gamma}\mathbf{D}$ are the same, so the effect of using $\check{\mathbf{\Gamma}}$ instead of $\mathbf{\Gamma}$ is to replace $\Gamma_{jj}$ by $\Gamma_{jj} - \mathsf{D}_{jj}^{-1}\mathsf{D}'_{jj}$. In cases where the $\Gamma_{jj}$ are not zero, therefore, it is always possible to change the normalisation of $\mathbf{S}$ by using a $\mathbf{D}$ in (16.68) with

$$\mathsf{D}_{jj} = \exp\left(k\int^{z}\Gamma_{jj}\,\mathrm{d}z\right), \quad j = 1, 2, 3, 4, \tag{16.71}$$

to give a new coupling matrix $\check{\mathbf{\Gamma}}$ with $\check{\Gamma}_{jj} = 0$. Then the factor (16.71) does not appear in $\mathsf{f}_j$. But it is not removed from the field components $\mathbf{e}$ because it has merely been transferred to the factor $\mathbf{D}$ in (16.69).

Every term of $\Gamma_{jj}$ has a factor that is a derivative $\mathrm{d}/\mathrm{d}z$. It can happen that $\Gamma_{jj} = \mathrm{d}\mathsf{L}_j/\mathrm{d}z$ where $\mathsf{L}_j$ is a local function only. Then (16.71) is also a local function and can be combined with the local factor $\mathbf{s}_j$ of the W.K.B. solution, (7.112), and it does not appear in the memory term. But many cases occur where $\Gamma_{jj}$ cannot be expressed in this way because $\Gamma_{jj}\mathrm{d}z$ is not a perfect differential. This was discussed by Smith, M.S. (1975) who gave examples for an electron plasma where $\Gamma_{jj}$ depends on the two functions $X(z)$ and $Z(z)$. Then $\int^z \Gamma_{jj}\mathrm{d}z$ depends on how the medium varies in the range of $z$ covered by the integral. In this respect it resembles $\int^z q\,\mathrm{d}z$. It has a memory content that is called 'additional memory'. Budden and Smith (1976) showed that there is additional memory for many types of wave propagation encountered in geophysics.

In ray tracing, the local factor of the W.K.B. solution is ignored. It is assumed that the ray paths are given solely by the eikonal function, § 14.2, and for stratified media this is the phase memory $k\int^z q\,\mathrm{d}z$. The local factor is slowly varying whereas the phase memory has a factor $k$ that makes its effect large, especially at high frequencies. In some cases, at lower frequencies, the local factor $\mathbf{s}_j$ may contain a factor $\mathrm{g}_j$ that varies too rapidly to be ignored. It must then be allowed for in the eikonal by writing the phase memory as $k\int^z q\,\mathrm{d}z + \mathrm{i}\ln \mathrm{g}_j$. This is just the converse of the process for removing $\Gamma_{jj}$ described above.

Smith, M.S (1976) studied the properties of the form of $\mathbf{\Gamma}$ given at (16.5), (16.6) and examined the effect on ray paths of incorporating $\Gamma_{jj}$ into the eikonal. He showed that in general ray paths are not reversible. For example the path of an ordinary ray that goes from A to B via the ionosphere is not the same as the reverse of the path of an ordinary ray going from B to A. The effect is small but Smith gave examples at frequencies 100, 300 and 900 kHz where it could be important. The use of $\Gamma_{jj}$ affects the calculation of lateral deviation §§ 10.13, 10.14 and of the bearing errors that result from it. In another paper Smith, M.S. (1975) gave a possible physical interpretation of $\Gamma_{jj}$ and showed how its influence can be detected in the results of full wave calculations of the reflecting properties of the ionosphere.

The $\Gamma_{jj}$ become important in ray theory at low frequencies where the factor $k$, in the conventional phase memory $k\int q\mathrm{d}z$, gets small. But this is also where the W.K.B. approximations used in ray theory are expected to be inaccurate. It may be asked whether these two effects are really separate. The reason for thinking that the effect of $\Gamma_{jj}$ must be allowed for when the non-diagonal elements of $\mathbf{\Gamma}$ can still be neglected was explained at the beginning of this section. It is that the effect of $\Gamma_{jj}$ is cumulative whereas the effect of the non-diagonal elements is smaller because of destructive interference. But this explanation has not been fully tested and there is a need for further study.

## 16.10. Second order coupled equations

Some forms of coupled equations have been used in which the principal terms contain derivatives up to the second order, and the coupling terms have derivatives up to the first order. An example has already been given at (16.43). Second order coupled equations of this kind have been used mainly for waves that are vertically incident on the ionosphere. A few special cases for oblique incidence have been given (Heading, 1953; Budden and Clemmow, 1957) but not extensively studied. The most important type for vertical incidence is the pair of equations, known as the Försterling equations (16.43) or (16.90), (16.91) and these are the main subject of study in

§§ 16.11–16.13. They were formulated (Försterling, 1942) long before the first order coupled equations were known.

For vertical incidence $S=0$, $C=1$, the differential equations (7.77), (7.78) take the simpler form

$$\frac{\mathrm{d}E_y}{\mathrm{d}z}=\mathrm{i}k\mathscr{H}_x,\quad \frac{\mathrm{d}E_x}{\mathrm{d}z}=-\mathrm{i}k\mathscr{H}_y,\quad \mathscr{H}_z=0, \tag{16.72}$$

$$\frac{\mathrm{d}\mathscr{H}_y}{\mathrm{d}z}=-\frac{\mathrm{i}k}{\varepsilon_0}D_x,\quad \frac{\mathrm{d}\mathscr{H}_x}{\mathrm{d}z}=\frac{\mathrm{i}k}{\varepsilon_0}D_y,\quad D_z=0. \tag{16.73}$$

If $\mathscr{H}_x$, $\mathscr{H}_y$ are eliminated:

$$\frac{\mathrm{d}^2E_x}{\mathrm{d}z^2}+\frac{k^2}{\varepsilon_0}D_x=0,\quad \frac{\mathrm{d}^2E_y}{\mathrm{d}z^2}+\frac{k^2}{\varepsilon_0}D_y=0. \tag{16.74}$$

Now $D_x$, $D_y$, $D_z$ can be expressed in terms of $E_x$, $E_y$, $E_z$ from (7.79) and (3.35). The last of (16.73) gives

$$M_{zx}E_x+M_{zy}E_y+(1+M_{zz})E_z=0 \tag{16.75}$$

whence $E_z$ may be eliminated from the expressions for $D_x$, $D_y$. Then (16.74) become

$$\frac{1}{k^2}\frac{\mathrm{d}^2E_x}{\mathrm{d}z^2}+\left(1+M_{xx}-\frac{M_{xz}M_{zx}}{1+M_{zz}}\right)E_x+\left(M_{xy}-\frac{M_{xz}M_{zy}}{1+M_{zz}}\right)E_y=0 \tag{16.76}$$

$$\frac{1}{k^2}\frac{\mathrm{d}^2E_y}{\mathrm{d}z^2}+\left(1+M_{yy}-\frac{M_{yz}M_{zy}}{1+M_{zz}}\right)E_y+\left(M_{yx}-\frac{M_{yz}M_{zx}}{1+M_{zz}}\right)E_x=0. \tag{16.77}$$

Note that the four coefficients used here are the four elements $\mathsf{T}_{41}$, $-\mathsf{T}_{42}$. $\mathsf{T}_{32}$, $-\mathsf{T}_{31}$ of the matrix $\mathbf{T}$, (7.81). The $x$ and $y$ axes may now be chosen so that the vector $\mathbf{Y}$ is in the $x$–$z$ plane, that is $l_y=0$. Let $n_{\mathrm{O}}$, $n_{\mathrm{E}}$ be the two refractive indices for waves travelling upwards in a fictitious homogeneous medium with the properties of the actual ionosphere at each level, and let $\rho_{\mathrm{O}}$, $\rho_{\mathrm{E}}$ be the corresponding values of the wave polarisation $E_y/E_x$. Then the four quantities $n_{\mathrm{O}}$, $n_{\mathrm{E}}$, $\rho_{\mathrm{O}}$, $\rho_{\mathrm{E}}$ are functions of $z$ and are expressible in terms of the $M_{ij}$. They provide an alternative way of specifying the properties of the plasma at each level, and the coefficients in (16.76), (16.77) can be expressed in terms of them. It can be shown that the equations then become

$$\begin{aligned}\frac{1}{k^2}\frac{\mathrm{d}^2E_x}{\mathrm{d}z^2}+\frac{\rho_{\mathrm{O}}n_{\mathrm{E}}^2-\rho_{\mathrm{E}}n_{\mathrm{O}}^2}{\rho_{\mathrm{O}}-\rho_{\mathrm{E}}}E_x+\frac{n_{\mathrm{O}}^2-n_{\mathrm{E}}^2}{\rho_{\mathrm{O}}-\rho_{\mathrm{E}}}E_y=0,\\ \frac{1}{k^2}\frac{\mathrm{d}^2E_y}{\mathrm{d}z^2}+\frac{\rho_{\mathrm{O}}n_{\mathrm{O}}^2-\rho_{\mathrm{E}}n_{\mathrm{E}}^2}{\rho_{\mathrm{O}}-\rho_{\mathrm{E}}}E_y-\frac{n_{\mathrm{O}}^2-n_{\mathrm{E}}^2}{\rho_{\mathrm{O}}-\rho_{\mathrm{E}}}E_x=0.\end{aligned} \tag{16.78}$$

The proof is left as an exercise for the reader, who should also verify that the equations reduce to the correct form at the magnetic equator, and at the magnetic poles.

Equations (16.78) or equivalently (16.76), (16.77) are in a sense second order coupled equations. The first two terms of each equation are principal terms and the remaining term is the coupling term. But the coupling terms are not in general small and so this form is unsuitable for use with a method of successive approximation.

### 16.11. Försterling's coupled equations for vertical incidence

Försterling's equations (16.43) have already been derived in § 16.4 as a deduction from the first order coupled equations. They will now be derived by a shorter and more direct method that is similar to the original method (Försterling, 1942).

In the differential equations (16.74) the variables $E_x$, $E_y$, $D_x$, $D_y$ refer to the total fields. Each of them is now to be expressed as the sum of the fields for the ordinary and extraordinary waves, thus

$$E_x = E_x^{(\mathrm{O})} + E_x^{(\mathrm{E})}, \quad E_y = E_y^{(\mathrm{O})} + E_y^{(\mathrm{E})} \tag{16.79}$$

and similarly for $D_x$. $D_y$. The four new variables on the right in (16.79) may be made to satisfy two further relations. We now choose the $x$ and $y$ axes so that $l_y = 0$. Then they may be made to satisfy

$$E_y^{(\mathrm{O})}/E_x^{(\mathrm{O})} = \rho_{\mathrm{O}}, \quad E_y^{(\mathrm{E})}/E_x^{(\mathrm{E})} = \rho_{\mathrm{E}} \tag{16.80}$$

where $\rho_{\mathrm{O}}$, $\rho_{\mathrm{E}}$ are the two wave polarisations as given by magnetoionic theory, ch. 4. They apply for both upgoing and downgoing waves since the same system of axes is used for both; see end of § 4.4. Equations (16.80) are the required two further relations and they are used to *define* the ordinary and extraordinary waves in the variable medium. Now (4.29) gives $\rho_{\mathrm{O}}\rho_{\mathrm{E}} = 1$ so that $E_x^{(\mathrm{E})} = \rho_{\mathrm{O}} E_y^{(\mathrm{E})}$ and hence (16.79) becomes

$$E_x = E_x^{(\mathrm{O})} + \rho_{\mathrm{O}} E_y^{(\mathrm{E})}, \quad E_y = \rho_{\mathrm{O}} E_x^{(\mathrm{O})} + E_y^{(\mathrm{E})}. \tag{16.81}$$

The electric displacement $\boldsymbol{D}$ is derived from the electric field by the constitutive relations, which are the same for a homogeneous and a variable medium. For the ordinary and extraordinary component waves, (4.42), (4.43) show that

$$D_x^{(\mathrm{O})} = \varepsilon_0 n_{\mathrm{O}}^2 E_x^{(\mathrm{O})}, \quad D_y^{(\mathrm{O})} = \varepsilon_0 n_{\mathrm{O}}^2 E_y^{(\mathrm{O})}, \quad D_x^{(\mathrm{E})} = \varepsilon_0 n_{\mathrm{E}}^2 E_x^{(\mathrm{E})}, \quad D_y^{(\mathrm{E})} = \varepsilon_0 n_{\mathrm{E}}^2 E_y^{(\mathrm{E})}. \tag{16.82}$$

The total $\boldsymbol{D}$ must be the sum of the contributions from the two component waves. Hence

$$\varepsilon_0^{-1} D_x = n_{\mathrm{O}}^2 E_x^{(\mathrm{O})} + n_{\mathrm{E}}^2 \rho_{\mathrm{O}} E_y^{(\mathrm{E})}, \quad \varepsilon_0^{-1} D_y = n_{\mathrm{O}}^2 \rho_{\mathrm{O}} E_x^{(\mathrm{O})} + n_{\mathrm{E}}^2 E_y^{(\mathrm{E})}. \tag{16.83}$$

Equations (16.81), (16.83) are now substituted in (16.74). A prime $'$ is used to denote $k^{-1}\mathrm{d}/\mathrm{d}z$. Thus

$$E_x^{(\mathrm{O})\prime\prime} + \rho_{\mathrm{O}} E_y^{(\mathrm{E})\prime\prime} + 2\rho_{\mathrm{O}}' E_y^{(\mathrm{E})\prime} + \rho_{\mathrm{O}}'' E_y^{(\mathrm{E})} + n_{\mathrm{O}}^2 E_x^{(\mathrm{O})} + n_{\mathrm{E}}^2 \rho_{\mathrm{O}} E_y^{(\mathrm{E})} = 0 \tag{16.84}$$

$$\rho_{\mathrm{O}} E_x^{(\mathrm{O})\prime\prime} + E_y^{(\mathrm{E})\prime\prime} + 2\rho_{\mathrm{O}}' E_x^{(\mathrm{O})\prime} + \rho_{\mathrm{O}}'' E_x^{(\mathrm{O})} + n_{\mathrm{O}}^2 \rho_{\mathrm{O}} E_x^{(\mathrm{O})} + n_{\mathrm{E}}^2 E_y^{(\mathrm{E})} = 0. \tag{16.85}$$

Now (16.85) is multiplied by $\rho_{\mathrm{O}}$ and (16.84) is subtracted; similarly (16.84) is multiplied by $\rho_O$ and (16.85) is subtracted:

$$(\rho_O^2-1)(E_x^{(O)\prime\prime}+n_O^2E_x^{(O)})+\rho_O\rho_O''E_x^{(O)}+2\rho_O\rho_O'E_x^{(O)\prime}=2\rho_O'E_y^{(E)\prime}+\rho_O''E_y^{(E)} \quad (16.86)$$

$$(\rho_O^2-1)(E_y^{(E)\prime\prime}+n_E^2E_y^{(E)})+\rho_O\rho_O''E_y^{(E)}+2\rho_O\rho_O'E_y^{(E)\prime}=2\rho_O'E_x^{(O)\prime}+\rho_O''E_x^{(O)}. \quad (16.87)$$

The equations are now in the required coupled form, but some further simplifications are possible. New dependent variables are chosen which remove the first order derivative from the principal terms. Thus let

$$E_x^{(O)}=\mathscr{F}_O(\rho_O^2-1)^{-\frac{1}{2}},\quad E_y^{(E)}=\mathscr{F}_E(\rho_O^2-1)^{-\frac{1}{2}} \quad (16.88)$$

so that, from (16.81),

$$\mathscr{F}_O=(-E_x+\rho_OE_y)(\rho_O^2-1)^{-\frac{1}{2}},\quad \mathscr{F}_E=(\rho_OE_x-E_y)(\rho_O^2-1)^{-\frac{1}{2}} \quad (16.89)$$

and use $\psi$ from (16.14). Then

$$\mathscr{F}_O''+(n_O^2+\psi^2)\mathscr{F}_O=\psi'\mathscr{F}_E+2\psi\mathscr{F}_E', \quad (16.90)$$

$$\mathscr{F}_E''+(n_E^2+\psi^2)\mathscr{F}_E=\psi'\mathscr{F}_O+2\psi\mathscr{F}_O', \quad (16.91)$$

which are Försterling's coupled equations, arranged with the principal terms on the left and the coupling terms on the right. Note that in (16.43) $\mathsf{b}_1=\mathscr{F}_O, \mathsf{b}_3=-\mathscr{F}_E$.

For frequencies greater than about 1 MHz, the coupling parameter $\psi$ (6.14) is negligibly small at real heights, in nearly all cases of practical importance. If it is neglected, the equations (16.90), (16.91) become

$$\mathscr{F}_O''+n_O^2\mathscr{F}_O=0,\quad \mathscr{F}_E''+n_E^2\mathscr{F}_E=0 \quad (16.92)$$

which shows that the ordinary and extraordinary waves are then propagated and reflected independently of each other. The equations can then be treated in exactly the same way as (7.6) was treated when discussing propagation in an isotropic ionosphere in chs. 7, 8, 12, 15. This provides the justification for studying the ordinary and extraordinary waves independently, for vertical incidence, as was done in ch. 13. Instead of the variable $E_y$ used in (7.6), the field variables to be used are the Försterling variables (16.89). Then much of the theory of chs. 12, 15 can be applied also to an anisotropic ionosphere. In particular a phase integral formula for the reflection coefficient, analogous to (8.78), can be derived for each of the two waves.

The neglect of $\psi$ in (16.90), (16.91) is not permissible, even at high frequencies, in the lowest part of the ionosphere where the waves enter it, or emerge into free space. It is here that the phenomenon of limiting polarisation occurs; § 17.10, 17.11. But this level is remote from the reflection levels, so that (16.92) can still be used as a basis for finding reflection coefficients and equivalent heights.

The properties of the coupling parameter $\psi$ are discussed in the following section. Some applications of Försterling's equations are given in §§16.13–16.14.

### 16.12. Properties of the coupling parameter $\psi$

Equation (6.14) gives $\psi$ and in view of the relation $\rho_O\rho_E=1$ it can be written

$$\psi = \frac{\rho_O'}{\rho_O^2 - 1} = \frac{\rho_E'}{\rho_E^2 - 1} = \frac{1}{2k}\frac{d}{dz}\ln\left(\frac{\rho_O - 1}{\rho_O + 1}\right) = \frac{1}{4k}\frac{d}{dz}\ln\left(\frac{\rho_O + \rho_E - 2}{\rho_O + \rho_E + 2}\right). \quad (16.93)$$

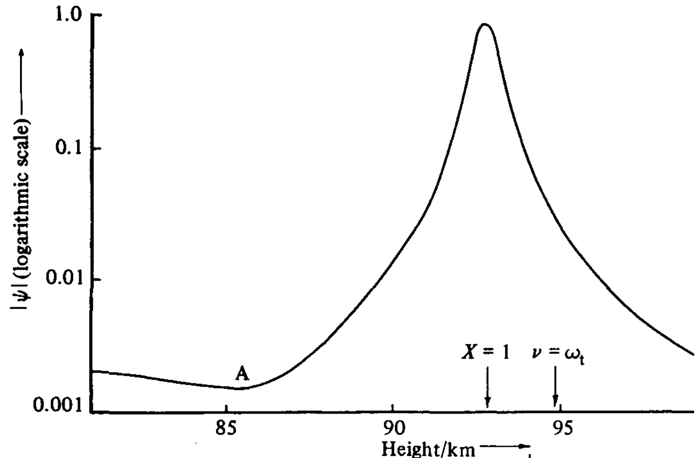
Now (4.20) and (4.27) show that

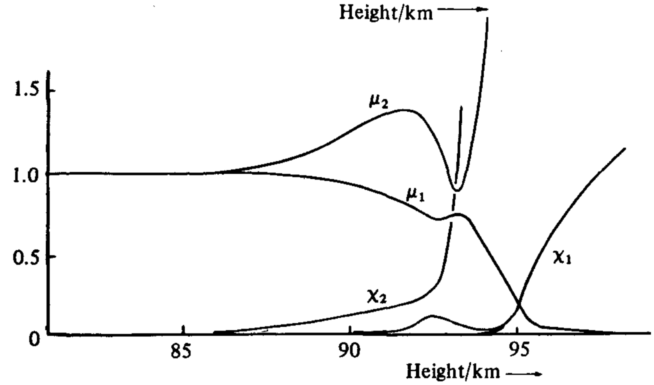
$$\rho_O + \rho_E = \frac{\mathrm{i}Y\sin^2\Theta}{(U - X)\cos\Theta} = \frac{2\mathrm{i}Z_t\,\mathrm{sgn}(\cos\Theta)}{U - X}. \quad (16.94)$$

For brevity it will be assumed that $\cos\Theta$ is positive as in the northern hemisphere so that $\mathrm{sgn}(\cos\Theta) = +1$. Then

$$\psi = \frac{-\frac{1}{2}\mathrm{i}Z_t}{Z_t^2 + (U - X)^2}(X' + \mathrm{i}Z') \quad (16.95)$$

Fig. 16.5. The function $|\psi|$, and the real parts $\mu$ and minus the imaginary parts $\chi$, of the two refractive indices, plotted against height $z$ for a frequency of 1 MHz in a typical ionospheric layer, chosen to simulate the E-layer in the daytime. In this example the ionosphere is a Chapman layer with its maximum at 115 km and having a penetration frequency of 4.4 MHz. The scale height $H$ is 10 km. The collision frequency is given by $\nu = \nu_0\exp(-z/H)$ and is equal to $10^6\,\mathrm{s}^{-1}$ at 90 km. $f_H = 1.12$ MHz, $\Theta = 23.27°$, $\omega_t = 5.98 \times 10^5\,\mathrm{s}^{-1}$.



where the prime $'$ means $k^{-1}\mathrm{d}/\mathrm{d}z$. Thus $\psi$ depends on the rates of variation with height $z$ of $X$, proportional to electron concentration, and $Z$, proportional to electron collision frequency. In a homogeneous medium $\psi$ is zero. It can also be shown from (4.67) that

$$\psi = \frac{-\frac{1}{2}\mathrm{i}X^2 Y^3 \sin^2\Theta\cos\Theta}{A^2(n_\mathrm{O}^2 - n_\mathrm{E}^2)^2}(X' + \mathrm{i}Z') \tag{16.96}$$

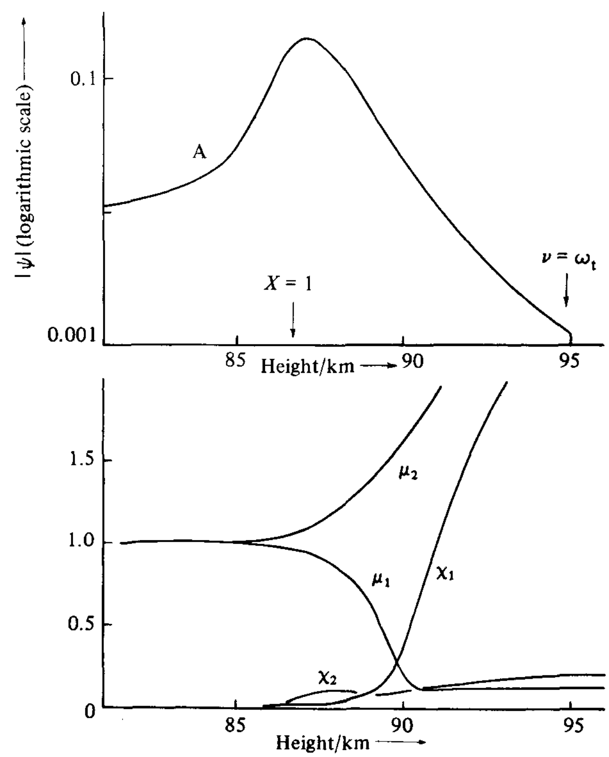
where $A$ is given by (4.66). The point where $A = 0$ is not a singularity of $\psi$ because there one of $n_\mathrm{O}^2$, $n_\mathrm{E}^2$ is infinite and the denominator is bounded and non-zero.

It is often convenient to treat $X$ and $Z$ as analytic complex functions of $z$, but real on the real $z$ axis. The functions $X(z)$, $Z(z)$ must have singularities but it is assumed that these are not near the real $z$ axis, and they are not considered further here. Then in the regions of the $z$ plane that are of interest, $X'$ and $Z'$ are bounded. The only singularities of $\psi$ are where the denominator is zero, that is where

$$X = 1 - \mathrm{i}(Z \pm Z_\mathrm{t}) \tag{16.97}$$

and these are simple poles of $\psi$. They are the coincident coupling points $\mathrm{C}_1$, $\mathrm{C}_2$ for the upper sign, and $\mathrm{C}_3$, $\mathrm{C}_4$ for the lower sign; § 16.5.

Fig. 16.6. The same as fig. 16.5 but for waves of frequency 160 kHz.



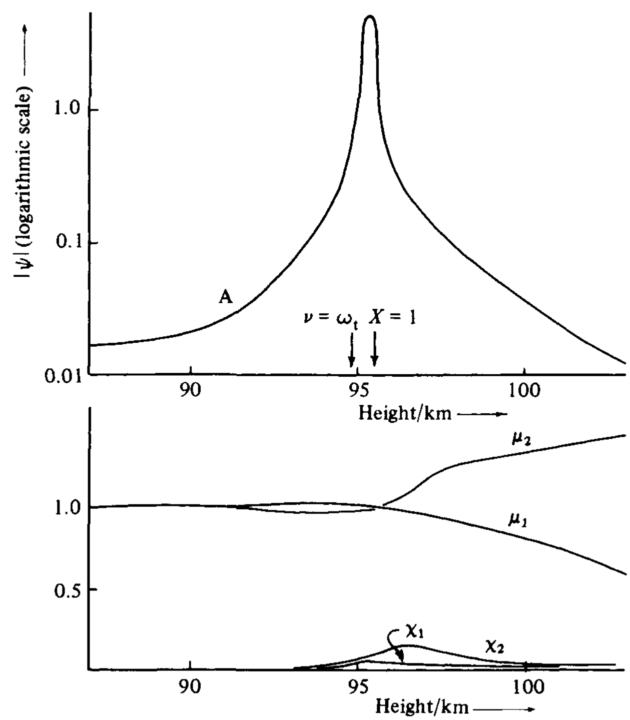
Below the ionosphere $X$ is zero, but $Z$ and $Z'$ are not zero, and

$$\psi = \tfrac{1}{2}Z'Z_t/\{(1 - iZ)^2 + Z_t^2\}. \tag{16.98}$$

In this case there are no electrons and the medium behaves like free space. Yet (16.98) shows that the coupling parameter is not zero. This has important consequences in the theory of limiting polarisation and is discussed in § 17.10, 17.11.

Some typical curves showing how $|\psi|$ depends on height $z$ are given in fig. 16.5–16.7. Here a model has been assumed for the ionosphere in which both $X$ and $Z$ vary with height $z$. Curves of the real parts and minus the imaginary parts, $\mu$, $\chi$ respectively, of the two refractive indices are also shown for comparison. The transition height where $\nu = \omega_t$ and the level where $X = 1$ are indicated. All curves of $|\psi|$ show a bend marked A in the figures. At levels below this the electron concentration is negligible and the last factor of (16.95) is determined by $Z'$. Thus at these low levels it is the variation of collision frequency with height that ensures that $\psi$ is not zero. At higher levels the electron concentration is appreciable and varies much more rapidly than the collision frequency. At levels above A, therefore, $Z'$ is negligible compared with $X'$. The curves show a maximum of $|\psi|$ near the level

Fig. 16.7. Curves for a frequency of 160 kHz in an ionospheric layer chosen to simulate the lower part of the E-layer at night. The ionosphere is the same as in fig. 16.5 except that the penetration frequency is now 0.5 MHz.



where $X = 1$. This is $\mathrm{Re}(X)$ for the poles (16.97) of $\psi$. The maximum is very marked when the level of $X = 1$ is near the transition height $\nu = \omega_t$, and less marked when these levels are well separated, fig. 16.6.

In figs. 16.5, 16.6 the height where $X = 1$ is below the transition height and conditions are greater than critical; see § 16.5. Thus at all heights $\mu_2$, $\chi_2$ refer to the ordinary wave and $\mu_1, \chi_1$ to the extraordinary wave. In fig. 16.7 the height where $X = 1$ is slightly above the transition height and conditions are less than critical. Hence $\mu_1$, $\chi_1$ refer to the ordinary wave below the level where $X = 1$, and to the extraordinary wave above this. Similarly $\mu_2$, $\chi_2$ refer to the extraordinary wave below, and to the ordinary wave, that is the whistler mode, above this level.

For frequencies greater than about 2 MHz, the level where $X = 1$ is nearly always well above the transition height and it is then permissible to neglect collisions, as was done in most of chs. 10, 12, 13. Then $\psi$ is given by

$$\psi \approx \frac{-\frac{1}{2}\mathrm{i}Z_t X'}{Z_t^2 + (1 - X)^2}. \tag{16.99}$$

If typical numerical values are inserted, it can be shown that at high frequencies $|\psi|$ remains small so that coupling effects are negligible. If $\psi$ is neglected, Försterling's equations give the two separate equations (16.92) already discussed in § 16.11.

In this and the previous section the coupling, for vertical incidence, between the ordinary and extraordinary waves is expressed by the coupling parameter $\psi$ and occurs because of the vertical gradients of $X$ and $Z$ in (16.95). It was assumed that the vector $\mathbf{Y}$ is in the $x$–$z$ plane and that its magnitude $Y$ and direction $\Theta$ are independent of the coordinates. Important instances can occur where these assumptions cannot be made. There can then be $z$ dependence of $Y$, $\Theta$ and the azimuth angle $\Phi$ of the plane containing $\mathbf{Y}$ and the $z$ axis, and these gradients can also give some coupling; see the comments at the end of § 18.8 on spatial variations of $\mathbf{Y}$. An extension of the theory to deal with this was given by Cohen (1960), for vertical incidence. He obtained coupled equations similar to (16.90), (16.91) but they contain two coupling parameters $\psi_O$, $\psi_E$ instead of one, whose definitions are more complicated than (16.93). The $\psi$s on the right of (16.90) must be replaced by $\psi_E$ and those on the right of (16.91) by $\psi_O$. The $\psi^2$ on the left of both (16.90), (16.91) is replaced by $\psi_O\psi_E$. There are corresponding modifications in the definition (16.89) of the dependent variables. Cohen has discussed important applications of his equations to radio propagation in the sun's atmosphere and in the interplanetary plasma, as well as ionospheric problems. Both Cohen (1960) and Titheridge (1971c) have studied the effect of a varying $\Theta$ on Faraday rotation

## 16.13. The method of 'variation of parameters'

In a number of important papers (Gibbons and Nertney, 1951, 1952; Kelso, Nearhoof, Nertney and Waynick, 1951; Rydbeck, 1950; Davids and Parkinson, 1955) a method of successive approximations has been applied to Försterling's coupled equations (16.90), (16.91). The method used is called 'variation of parameters'. It is a well known mathematical technique; see for example, Heading (1975c), Jeffreys and Jeffreys (1972). It is summarised here in the form needed for the Försterling equations.

Suppose that we seek a solution of the linear inhomogeneous equation

$$d^2u/ds^2 + uf(s) = g(s). \tag{16.100}$$

Consider the homogeneous equation

$$d^2u/ds^2 + uf(s) = 0 \tag{16.101}$$

and let $u_1(s)$, $u_2(s)$ be two independent solutions of it, whose Wronskian is

$$W(u_1, u_2) = u_1 du_2/ds - u_2 du_1/ds. \tag{16.102}$$

We now seek a solution of (16.100) of the form

$$u = Au_1 + Bu_2 \tag{16.103}$$

where $A$ and $B$ are functions of $s$, and since there are two of them we can impose one relation between them. Now

$$du/ds = u_1 dA/ds + u_2 dB/ds + A du_1/ds + B du_2/ds. \tag{16.104}$$

For the imposed relation take

$$u_1 dA/ds + u_2 dB/ds = 0. \tag{16.105}$$

Then

$$\frac{d^2u}{ds^2} = \frac{du_1}{ds}\frac{dA}{ds} + \frac{du_2}{ds}\frac{dB}{ds} + A\frac{d^2u_1}{ds^2} + B\frac{d^2u_2}{ds^2}. \tag{16.106}$$

Substitute this in (16.100) and use (16.101) with $u = u_1$, $u_2$. Then

$$\frac{du_1}{ds}\frac{dA}{ds} + \frac{du_2}{ds}\frac{dB}{ds} = g(s). \tag{16.107}$$

Now (16.105), (16.107) may be solved to give

$$\frac{dA}{ds} = -\frac{u_2 g}{W}, \quad \frac{dB}{ds} = \frac{u_1 g}{W} \tag{16.108}$$

and the solution (16.103) is

$$u = -u_1 \int_a^s \frac{u_2 g}{W} ds - u_2 \int_s^b \frac{u_1 g}{W} ds \tag{16.109}$$

where $a$ and $b$ are arbitrary constants to be determined from the physical conditions. In the cases of interest here the differential equations are in the normal form, that is, like (16.100) they have no first derivative term, so that the Wronskian $W$ is independent of $s$ and may be taken outside the integrals (16.109).

The application of this result to Försterling's equations (16.90), (16.91) will now be illustrated by considering a specific problem in which an ordinary wave is incident from below on the region around where $X = 1$ and where the coupling may be strong. The waves that go through this region may be reflected at a higher level, but the reflected waves will at first be ignored since the effect of coupling on waves coming down from above can be treated separately.

The coupling parameter $\psi$ is assumed to be negligible except in the coupling region, and is there assumed to be so small that the right-hand sides of (16.90), (16.91) may be neglected in the zero order approximation. The factors $\psi^2$ on the left will be neglected completely; they could be allowed for by replacing $n_O, n_E$ by $(n_O^2+\psi^2)^{\frac{1}{2}}$, $(n_E^2+\psi^2)^{\frac{1}{2}}$ respectively in the following argument. It is also assumed that conditions are greater than critical so that $n_O, n_E$ can be used unambiguously. Finally it is assumed that $n_O, n_E$ are not near zero in the coupling region (see figs. 16.5, 16.6) so that the reflection levels are not close to the coupling region and the W.K.B. solutions (16.111) below, can be used there. Take $s = kz$ so that in the zero order approximation the equations are

$$d^2\mathscr{F}_O/ds^2 + n_O^2\mathscr{F}_O = 0, \quad d^2\mathscr{F}_E/ds^2 + n_E^2\mathscr{F}_E = 0 \tag{16.110}$$

and their W.K.B. solutions are

$$\begin{aligned} \mathscr{F}_O^{(1)} &= n_O^{-\frac{1}{2}}\exp\left(-i\int_0^s n_O\,ds\right), \quad \mathscr{F}_O^{(2)} = n_O^{-\frac{1}{2}}\exp\left(i\int_0^s n_O\,ds\right), \\ \mathscr{F}_E^{(1)} &= n_E^{-\frac{1}{2}}\exp\left(-i\int_0^s n_E\,ds\right), \quad \mathscr{F}_E^{(2)} = n_E^{-\frac{1}{2}}\exp\left(i\int_0^s n_E\,ds\right). \end{aligned} \tag{16.111}$$

There is no upgoing extraordinary wave below the coupling region, and any downgoing waves above it are being ignored, so that at all levels the zero order approximation is

$$\mathscr{F}_O = \mathscr{F}_O^{(1)}, \quad \mathscr{F}_E = 0. \tag{16.112}$$

To obtain a more accurate solution, this result is substituted in the right-hand sides of (16.90), (16.91). This leaves (16.90) unaffected so that to the first order $\mathscr{F}_O$ is unchanged. Equation (16.91), however, becomes, if $\psi^2$ on the left is neglected:

$$\frac{d^2\mathscr{F}_E}{ds^2} + n_E^2\mathscr{F}_E = g(s) \text{ where } g(s) = \mathscr{F}_O^{(1)}\frac{d\psi}{ds} + 2\psi\frac{d\mathscr{F}_O^{(1)}}{ds}. \tag{16.113}$$

The formula (16.109) is now applied with $u_1 = \mathscr{F}_E^{(1)}$, $u_2 = \mathscr{F}_E^{(2)}$ so that $W = 2i$. Thus to the first order

$$\mathscr{F}_E = \tfrac{1}{2}i\mathscr{F}_E^{(1)}\int_a^s \mathscr{F}_E^{(2)}g(s)\,ds + \tfrac{1}{2}i\mathscr{F}_E^{(2)}\int_s^b \mathscr{F}_E^{(1)}g(s)\,ds. \tag{16.114}$$

The limits $a$, $b$ must now be found. When $s$ is outside the coupling region the integrands in (16.114) are zero and the integrals are constant. When $s$ is well below the coupling region, the first integral must be zero since $\mathscr{F}_E^{(1)}$ is an upgoing extraordinary wave which is absent there. Similarly the second integral must be zero when $s$ is above the coupling region since $\mathscr{F}_E^{(2)}$ is a downgoing extraordinary wave. Thus $a$ is any value of $s$ well below the coupling region and $b$ is any value well above it. Equation (16.114) shows that above the coupling region there is an upgoing extraordinary wave, the first term, and below it there is a downgoing extraordinary wave, the second term. These are generated by mode conversion, in the coupling region, from the incident upgoing ordinary wave. In fact (16.114) shows that the coupling region may be thought of as two distributed sources of these waves. The amplitudes of these sources at any level are the integrands.

For frequencies less than the electron gyro-frequency, the above method can also be used when the incident wave is an upgoing extraordinary wave, since this wave travels up through the coupling region to its reflection level near where $X = 1 + Y$. By mode conversion in the coupling region it gives rise to a downgoing ordinary wave below the coupling region, and an upgoing ordinary wave above it. Similarly the method can be used when the incident wave is a downgoing wave, ordinary or extraordinary, from above the coupling region.

The above theory is not only of great historical importance but also of physical interest, particularly in the interpretation of the integrands in (16.114) as distributed sources of waves produced by mode conversion. It was valuable as a method of computation in the days when the storage capacity of computers was less than it is today. But for the original basic differential equations (7.80), (7.81) the coupling and reflection points are ordinary points. Near them the field variables and the coefficients are bounded and continuous. They give no trouble when these basic equations are integrated numerically. Transformation of the equations to coupled form introduces, at the coupling points, singularities that are not present in the original equations. This and the need for repeatedly calculating $n_O$, $n_E$, $\psi$, or similar quantities, introduce complications when the Försterling equations or any other form of coupled equations are used in computing. The speed and large storage capacity of modern computers are now such that the original equations can be numerically integrated quickly and without approximations and this means that, for computing, the use of coupled equations and the method of variation of parameters are largely superseded.

## 16.14. The coupling echo

In measurements with radio pulses at vertical incidence with a frequency of 150 kHz, reflections have sometimes been observed that come from the level near where $X \approx 1$, as shown from the equivalent height of reflection. These are additional to the reflection that comes from the higher level where $X \approx 1 + Y$. They occur usually at night (Kelso *et al.*, 1951; Nertney, 1951, 1953; Lindquist, 1953; Parkinson, 1955).

At this frequency the values of $n_O$ and $n_E$ are close to unity when $X$ is near to 1; for examples see figs. 16.6, 16.7 which are for 160 kHz. This is because of the large value of $Z$. The reflection point where $X = 1 - iZ$ is a long way from the real $z$ axis so that the zero of $n_O$ is not apparent. But since there is a region of strong coupling near where $X \approx 1$, it was suggested that these reflections arise because of a coupling mechanism, and they are called 'coupling echoes'. Calculations using a method of successive approximation with variation of parameters, as in § 16.13 (Davids, 1952; Davids and Parkinson, 1955), have confirmed that reflections of this kind might be expected.

There is no coupling point associated with an upgoing ordinary and a downgoing extraordinary wave, nor with an upgoing extraordinary and downgoing ordinary wave. The coupling echo cannot be explained by a combination of separate simple coupling and reflection processes each associated with one coupling point. The regions where the W.K.B. solutions fail, surrounding the coupling points $C_1$ to $C_4$ and O, must overlap so that it is not possible to find a contour for use with the phase integral method. Smith, M.S. (1973a) has shown that for ionospheric models of a particular type, a coupling echo with detectable amplitude would not be expected when the function $N(z)$ is smoothly varying. Altman and Postan (1971) have made calculations for a range of models and frequencies. They show that coupling echoes can sometimes occur. The coupling processes that have to be invoked to explain them are more complicated than is generally assumed, and may be different for small and large values of the angle $\Theta$ between the earth's magnetic field and the vertical.

The possibility cannot be excluded that some 'coupling echoes' are in fact caused by scattering from irregularities of $N$ in the D-region, giving partial reflections, as described in § 11.13, and indeed some authors have used the term 'backscatter' for the processes that generate these echoes.

Further study of the theory of the coupling echo and further experimental results would be of the greatest interest. In particular it would be interesting to know how the polarisation of the reflected wave changes when the polarisation of the incident wave is changed.

## PROBLEMS 16

**16.1.** The coupling parameter $\psi$ (16.14) has simple poles in the complex $z$ (height) plane where $X = 1 - \mathrm{i}(Z \pm \mathrm{i}Z_t)$. Show that the residues are $\pm 1/(4k)$. Note that they are independent of $\mathrm{d}X/\mathrm{d}z$, $\mathrm{d}Z/\mathrm{d}z$.

**16.2.** For oblique incidence on the ionosphere when the earth's magnetic field is vertical, the Booker quartic is a quadratic equation for $q^2$, and its four solutions are $\pm q_{\mathrm{O}}, \pm q_{\mathrm{E}}$ for the ordinary and extraordinary waves respectively. The plane of incidence is the $x$–$z$ plane with $z$ vertical, and $\theta$ is the angle of incidence. For the ordinary wave the ratios $E_y/E_x$ and $-\mathscr{H}_x/\mathscr{H}_y$ are not the same but their geometric mean is denoted by $\rho_{\mathrm{O}}$, thus $\rho_{\mathrm{O}}^2 = -E_y\mathscr{H}_x/E_x\mathscr{H}_y$. In the same way $\rho_{\mathrm{E}}$ is defined for the extraordinary wave. Show that $\rho_{\mathrm{O}}\rho_{\mathrm{E}} = 1$. Let $\psi = \rho_{\mathrm{O}}'/(\rho_{\mathrm{O}}^2 - 1)$ where a prime ′ means $k^{-1}\mathrm{d}/\mathrm{d}z$. Let $\eta^2 = (U - X)/(UC^2 - X)$ where $C = \cos\theta$. Show that the second order coupled equations may be written

$$h_{\mathrm{O}}'' + \left\{ q_{\mathrm{O}}^2 + \psi^2 + \frac{\eta''}{\eta(\rho_{\mathrm{O}}^2 - 1)} \right\} h_{\mathrm{O}} = \left\{ -\psi' + \frac{\eta'' \rho_{\mathrm{O}}}{\eta(\rho_{\mathrm{O}}^2 - 1)} \right\} h_{\mathrm{E}} - 2\psi h_{\mathrm{E}}',$$

$$h_{\mathrm{E}}'' + \left\{ q_{\mathrm{E}}^2 + \psi^2 + \frac{\eta''}{\eta(\rho_{\mathrm{E}} - 1)} \right\} h_{\mathrm{E}} = \left\{ -\psi' + \frac{\eta'' \rho_{\mathrm{E}}}{\eta(\rho_{\mathrm{E}}^2 - 1)} \right\} h_{\mathrm{O}} - 2\psi h_{\mathrm{O}}',$$

where

$$h_{\mathrm{O}} = (\rho_{\mathrm{O}}^2 - 1)^{-\frac{1}{2}}(\eta E_x - \rho_{\mathrm{O}} E_y), \quad h_{\mathrm{E}} = (1 - \rho_{\mathrm{E}}^2)^{-\frac{1}{2}}(\eta E_x - \rho_{\mathrm{E}} E_y).$$

(See Budden and Clemmow, 1957).

**16.3.** Sketch the form of the $h'(f)$ curves you would expect if the critical coupling frequency is greater than the electron gyro-frequency. How would the polarisation of the reflected components depend on frequency?

**16.4.** Show that if the square matrix $\mathbf{S}$ of (6.53), (7.85) is normalised so that the coupling matrix $\mathbf{\Gamma} = -\mathbf{S}^{-1}\,\mathrm{d}\mathbf{S}/\mathrm{d}s$ (7.108) has zero trace, then det ($\mathbf{S}$) is independent of height $z = s/k$.

Note. The trace of a square matrix such as $\mathbf{\Gamma}$ is the sum of its diagonal elements and is written $\mathrm{Tr}(\mathbf{\Gamma})$. The adjugate $\mathbf{A} = \mathbf{S}^{-1}\det(\mathbf{S})$ of $\mathbf{S}$ is the transpose of the matrix of its cofactors; see § 7.14(4). To solve the problem the result $(\mathrm{d}/\mathrm{d}z)\det(\mathbf{S}) = \mathrm{Tr}(\mathbf{A}\,\mathrm{d}\mathbf{S}/\mathrm{d}z)$ is needed, and the reader should prove this. The proof is not difficult but the author has not found it in any textbook. The solution of the problem is given by Inoue and Horowitz (1966a) and by Rawer and Suchy (1967, p. 154), for the special case where all the diagonal elements of $\mathbf{\Gamma}$ are zero.

**16.5.** For a radio wave vertically incident on a horizontally stratified ionosphere, the horizontal Cartesian components of the electric intensity are $E_x$, $E_y$. The earth's magnetic field is in the $x - z$ plane. Let the polarisation of an ordinary wave be $\rho_{\mathrm{O}} = \mathrm{i}\tan\phi$. Show that the coupling parameter $\psi$ (16.93) is $\phi' = k^{-1}\mathrm{d}\phi/\mathrm{d}z$. Show that if $V$ is a two-dimensional vector with components $(E_x, -\mathrm{i}E_y)$, then if the $x$ and $y$ axes are rotated through a complex angle $\phi$, the components of $V$ in the new axes are $\pm(\mathrm{i}\mathscr{F}_{\mathrm{O}}, \mathscr{F}_{\mathrm{E}})$ where $\mathscr{F}_{\mathrm{O}}$, $\mathscr{F}_{\mathrm{E}}$ are the Försterling variables (16.89). What is the significance of this transformation when $V^2$ is zero?

[This version of Försterling's theory was given by Saha, Banerjea and Guha (1951).]

**16.6.** The coupling matrix $\mathbf{\Gamma}$ is $-\mathbf{S}^{-1}\mathbf{S}'$ from (7.108). Prove that the non-diagonal elements of $\mathbf{\Gamma}$ are given by

$$\Gamma_{ij} = \frac{1}{(q_i - q_j)}(\mathbf{S}^{-1}\mathbf{T}'\mathbf{S})_{ij} \quad \text{(not summed; } i \neq j).$$

[See Arantes and Scarabucci (1975). This result is useful for computing because $\mathbf{T}'$ is easier to compute than $\mathbf{S}'$.]