17

Coalescence of coupling points

17.1. Introduction

The first order coupled wave equations (16.22) were derived and studied in § 16.3 for the neighbourhood of a coupling point $z = z_p$ where two roots q_1, q_2 of the Booker quartic equation are equal. If this coupling point is sufficiently isolated, a uniform approximation solution can be used in its neighbourhood and this was given by (16.29), (16.30). It was used in § 16.7 to study the phase integral formula for coupling. The approximations may fail, however, if there is another coupling point near to z_p, and then a more elaborate treatment is needed. When two coupling points coincide this is called 'coalescence'. This chapter is concerned with coupled wave equations when conditions are at or near coalescence.

A coupling point that does not coincide with any other will be called a 'single coupling point'. It is isolated only if it is far enough away from other coupling points and singularities for the uniform approximation solution (16.30) to apply with small error for values of $|\zeta|$ up to about unity. Thus an isolated coupling point is single, but the reverse is not necessarily true.

Various types of coalescence are possible. The two coupling points that coalesce may be associated with different pairs q_1, q_2 and q_3, q_4 of roots of the Booker quartic. This is not a true coalescence because the two pairs of waves are propagated independently of each other. The two coupling processes can then be treated at the same time as was done in the example of § 16.4, and the two coupling points still behave as though they are single. Alternatively the two pairs may have one member in common, for example q_1, q_2 and q_1, q_3. This type does occur, as was shown for example by Smith, M.S. (1974a) but he does not call it coalescence. (See Smith's figs. 1 and 6 where $S = S_a, S_c, S_d$ or S_f). It occurs when the curve of q versus X has a vertical tangent at a point that is also a point of inflection. There is an example in fig. 6.6(e). The theory has not been fully worked out and is not given in this book. Finally the two coupling points may both be associated with the same pair of roots, say q_1, q_2. This is the type that is now to be studied. It is shown that, even for this type alone, there are two different kinds of coalescence, to be called a 'coalescence of the first kind' C1, and a 'coalescence of the second kind' C2.

For C1 at exact coalescence, one incident wave gives rise to two other waves with roughly equal amplitudes. For example in fig. 15.5 it gives both a reflected and a transmitted wave. When the coupling points move apart one of these waves gets weaker and the other stronger, and when the points are well separated the incident wave gives only one other wave.

For C2 at exact coalescence, one incident wave travels right through the region near the coalescence just as it would if the coupling points were not there. There is then only one emergent wave. The fields of both the parent incident wave, before the coalescence is reached, and the product wave afterwards, are given by a W.K.B. solution that applies for the whole wave path going right through the coalescence. When the coupling points move apart a second wave is produced and the original product wave gets weaker. When the coupling points are well separated the original product wave has negligible amplitude, and the second wave now takes over as a new product wave.

The C1 is important in the theory of partial penetration and reflection, § 15.10, figs. 15.5, 15.6. The C2 is important in the theory of radio windows §§ 17.6–17.9, and of limiting polarisation §§ 17.10–17.11. For a version of the theory of coalescence more detailed than that gives here, see Budden and Smith (1974).

17.2. Further matrix theory

In this chapter some further properties of matrices are needed in continuation of § 7.14. They can be found in standard text books, for example, Aitken (1956), Heading (1958), Frazer, Duncan and Collar (1938), Graham (1979). We use the 4×4 square matrices $\mathbf{T}$ (7.81), and $\mathbf{S}$ whose columns $\mathbf{s}_i$ are the four eigen columns of $\mathbf{T}$ with eigen values q_i. Let

$$\mathbf{C}(q) = \mathbf{T} - q\mathbf{1} \tag{17.1}$$

where $\mathbf{1}$ is the unit 4×4 matrix. Then $\mathbf{C}$ is called the characteristic matrix of $\mathbf{T}$. The q s are given by the Booker quartic equation which may be written

$$F(q) \equiv \det \mathbf{C} = 0; \tag{17.2}$$

compare (7.84). The transpose of the matrix of the cofactors of the elements of $\mathbf{C}$ in (17.2) is denoted by adj $\mathbf{C}$. (Note that Budden and Smith (1974) use adj $\mathbf{C}$ to mean the untransposed matrix of the cofactors.) The elements of the column $\mathbf{s}_i$ may be chosen to be any linear combination of the columns of adj $\mathbf{C}$ when $q = q_i$, and it is now assumed that the same combination is used for all four $\mathbf{s}_i$s. At a coupling point, where two q_is are equal, say $q_1 = q_2$, the two values of any element of adj $\mathbf{C}$ are also equal so that $\mathbf{s}_1 = \mathbf{s}_2$. It might happen that, at a coupling point all cofactors of $\mathbf{C}$ are zero. Then $\mathbf{s}_1, \mathbf{s}_2$ must be found in some other way, and it can be shown that they can be chosen so that they are independent. Then it can be further shown that the coupling point cannot be single. It must be a coalescence. The proof is given by Budden and Smith (1974, appendix §(b)).

The notation (16.17) for α, β is now used. It was shown in § 16.3 that β^2 is analytic near a single coupling point z_p so it can be expanded in a series

$$\beta^2 = \tfrac{1}{4}(q_1 - q_2)^2 = -a(z - z_\text{p}) + b(z - z_\text{p})^2 + \cdots. \qquad (17.3)$$

In the special case $q_1 = -q_2$ this is the same as the series (8.3) for q^2. There is another coupling point $z = z_\text{q}$ where $q_1 = q_2$ and

$$z_\text{q} - z_\text{p} \approx a/b. \qquad (17.4)$$

If $|z_\text{q} - z_\text{p}|$ is large enough, b/a must be small, and the two coupling points are then isolated. Coupled equations of the form (16.22) can then be used near each of them. But if z_p and z_q are close together, b/a is no longer small. Then one of two things can happen.

First it may happen that the matrix $\mathbf{U}$ in (16.20), (16.21) remains non-singular in the limit $z_\text{q} = z_\text{p}$ so that (16.21)–(16.30) can still be used. This would require that, in the limit, $\mathbf{s}_1$ and $\mathbf{s}_2$ are non-zero and equal. This always happens if at least one column of adj $\mathbf{C}$ is non-zero in the limit, because this column can be used for $\mathbf{s}_i$. Then we say that the coupling points approach a coalescence of the first kind, abbreviated by C1. Provided that only these two coupling points are near the domain considered, β^2 in (17.3) is with good approximation a quadratic function of z–z_p and equation (16.28) can be transformed to Weber's equation (15.70) or its equivalent (15.68) whose solutions are the parabolic cylinder functions or the function $\mathrm{F}(\zeta; A)$ (15.82). Further details are given in the following § 17.3.

Second it may happen that in the limit $z_\text{p} = z_\text{q}$ all elements of adj $\mathbf{C}$ are zero. Then the columns of adj $\mathbf{C}$ cannot be used to give the eigen columns $\mathbf{s}_i$, which must be constructed in some other way. It is then found that $\mathbf{s}_1$ and $\mathbf{s}_2$ are not parallel even though $q_1 = q_2$, and the coalescence is then said to be of the second kind, C2. For this case it can be shown that, when z_p and z_q are close together but not exactly coincident, there is a point close to them in the complex z plane where the first two columns $\mathbf{s}_\alpha, \mathbf{s}_\beta$ of $\mathbf{U}$ are parallel. Then $\mathbf{U}$ is singular. The transformation (16.21) cannot be used and a different one must be found. This is given in § 17.4. It introduces new variables w_1, w_2 that each satisfy a form of Weber's equation but it is different from the form used for C1.

For either C1 or C2, (17.4) shows that for exact coalescence $a = 0$, whence from (17.3) $\beta = \tfrac{1}{2}(q_1 - q_2)$ has a factor z–z_p, and is therefore analytic at and near the coalescent coupling points. Since $\alpha = \tfrac{1}{2}(q_1 + q_2)$ also is analytic, it follows that q_1 and q_2 are both analytic throughout a domain containing the coalescence.

17.3. Coalescence of the first kind, C1

At a coupling point two roots of the Booker quartic (17.2) are equal, and it is required that simultaneously

$$\det \mathbf{C} = 0 \quad \text{and} \quad \partial(\det \mathbf{C})/\partial q = 0. \tag{17.5}$$

Elimination of q from these gives

$$\Delta = 0 \tag{17.6}$$

where Δ is the discriminant (6.38)–(6.40) of the quartic and is a function of z. The solutions z of (17.6) are the coupling points in the complex z plane. To simplify the discussion it will now be assumed that electrons are the only species that appreciably affect the wave propagation. It will also be assumed that the collision frequency varies so slowly with z that $\partial Z/\partial z$ can be neglected. Then Δ depends on z only because of X for electrons, and (17.6) may be written

$$\Delta(X) = 0. \tag{17.7}$$

It is easy to extend the argument to deal with cases where these assumptions are not true.

At a coalescence, two solutions of (17.7) occur at the same z, which requires that simultaneously

$$\Delta(X) = 0 \quad \text{and} \quad \frac{\partial}{\partial z}\Delta(X) = \frac{\partial \Delta(X)}{\partial X}\frac{\partial X}{\partial z} = 0 \tag{17.8}$$

and the last of these can be satisfied either by $\partial\Delta/\partial X = 0$, or $\partial X/\partial z = 0$. The second occurs at a turning point of the function $X(z)$, such as the maximum of an ionospheric layer. The first is a property of $\Delta(X)$ and thence via (17.5) and (17.1) of the matrix $\mathbf{T}$, and does not depend on the ionospheric distribution functions.

Suppose now that (17.7) has a solution $X = X_c$, possibly complex, that is a coupling point where $q_1 = q_2$ and that is isolated in the complex X plane. Then the eigen columns $\mathbf{s}_1$, $\mathbf{s}_2$ can be chosen so that they are non-zero for all X near X_c and are identical when $X = X_c$. Suppose further that there is a value of X very close to X_c where $\partial X/\partial z = 0$, $z = z_m$. Then there are two values z_p, z_q of z where $X = X_c$, such that $z_p - z_m \approx z_m - z_q$. Thus conditions are near to coalescence of these two coupling points z_p, z_q in the z plane. Since the eigen columns $\mathbf{s}_1$, $\mathbf{s}_2$ are non-zero, the matrix $\mathbf{U}$ is non-singular in a domain containing z_p, z_q and the transformation (16.21) can be used. Conditions are near to C1.

Suppose that β^2 is exactly a quadratic function of z so that in (17.3) there are no terms after the term in $(z - z_p)^2$. Then

$$z_q = z_p + a/b, \quad \beta^2 = b(z - z_p)(z - z_q). \tag{17.9}$$

Let

$$\xi = (k^2 b)^{\frac{1}{4}}\{z - \tfrac{1}{2}(z_p + z_q)\}, \quad A = \tfrac{1}{2}(k^2 b)^{\frac{1}{4}}(z_q - z_p); \tag{17.10}$$

compare (15.67). Then the first equation (16.28) becomes

$$\mathrm{d}^2h_1/\mathrm{d}\xi^2 + (\xi^2 - A^2)h_1 = 0. \tag{17.11}$$

This is the same as (15.68) and can be used in a similar way. At exact coalescence C1, the coupling points z_p and z_q coincide and $A^2 = 0$.

Usually β^2 is not exactly a quadratic function of z. Suppose that it has zeros at z_p, z_q which are close together and that its other zeros are remote from them. Then near z_p, z_q the function β^2 must be approximately quadratic and it is still possible to reduce (16.28) to the form (17.11) by using a uniform approximation (see §§ 7.9, 8.20) first given by Miller and Good (1953). Instead of (17.10) let ξ and A be defined by

$$k\int_{z_\mathrm{p}}^{z} \beta\,\mathrm{d}z = \mathrm{i}\int_{-A}^{\xi} (A^2 - \xi^2)^{\frac{1}{2}}\,\mathrm{d}\xi, \tag{17.12}$$

$$k\int_{z_\mathrm{p}}^{z_\mathrm{q}} \beta\,\mathrm{d}z = \mathrm{i}\int_{-A}^{A} (A^2 - \xi^2)^{\frac{1}{2}}\,\mathrm{d}\xi = \mathrm{i}\pi A^2. \tag{17.13}$$

Here (17.12) maps the complex z plane into the ξ plane, and (17.13) ensures that it is a one-to-one mapping. It is easily checked that they reduce to (17.10) in the limit when β^2 is exactly quadratic, provided that the signs of the square roots are suitably chosen, and it is now assumed that this choice of sign is made. The zeros of β at z_p, z_q map into the points $\xi = \mp A$. With this transformation, (16.28) becomes

$$\mathrm{d}^2h_1/\mathrm{d}\xi^2 + (\xi^2 - A^2)h_1 = [\text{small term}]\,\mathrm{d}h_1/\mathrm{d}\xi. \tag{17.14}$$

The right-hand side is analogous to the second term in brackets in (7.52). It is zero if β^2 is exactly quadratic, and small enough to be neglected if β^2 is sufficiently near to quadratic. The main interest of (17.12)–(17.14) is that although it can be used for C1, it is not suitable for C2 and a different form (17.32), (17.33) must be used, for which A^2 is not zero at exact coalescence.

The most important example of a coalescence C1 is when the coupling point at $X = X_\mathrm{c}$ is a reflection point, for example when q_1 and q_2 apply to the upgoing and downgoing ordinary wave respectively. Since, in most cases of practical interest, X_c is isolated in the complex X plane, the other two waves associated with q_3 and q_4 are independently propagated and can be disregarded. The reflection process is described by the second order equation (16.28). This is of the same form as (7.6) used for an isotropic ionosphere. If β^2 in (16.28) is a quadratic function of z, the equation can be used in exactly the same way as (7.6) was used in §§ 15.9–15.11 to study propagation in an ionosphere with a parabolic electron height distribution $N(z)$. The theory of §§ 15.9–15.11 can therefore be used also for studying partial penetration and reflection of the ordinary wave, and of the extraordinary wave in an anisotropic ionosphere. The main difference is that in (15.68) the dependent variable is the horizontal electric field E of a wave obliquely incident on an isotropic ionosphere, whereas in (17.11) it is h_1 which is a linear combination of E_x, E_y, $\mathscr{H}_x$, $\mathscr{H}_y$ given by (16.21), (16.25).

Other kinds of C1 are possible. In the example just discussed, the C1 occurred because a reflection point X_c was near the maximum of an ionospheric layer where $\partial X/\partial z = 0$. Now X_c might be some other kind of coupling point, for example where the upgoing ordinary and extraordinary waves are critically coupled. When electrons only are allowed for, this usually occurs in the height range 90 to 98 km for vertical incidence and lower for oblique incidence, so that it is unlikely to be near a layer maximum, and therefore unlikely to lead to a C1. For frequencies small enough for ions to be important, however, the theory has not been so fully explored and it is possible that this might give C1s that are important.

In both the above examples (17.7) was satisfied because of the proximity of a zero of $\partial X/\partial z$ to a coupling point. It could alternatively be satisfied by a zero of $\partial\Delta/\partial X$, even in a medium where X varies monotonically with z. No cases of C1s of this type have yet been found but the possibility has not been excluded.

17.4. Coalescence of the second kind, C2

Consider now a domain of the complex z plane where all the ion and electron concentrations are monotonic functions of z. It is assumed that all concentrations and collision frequencies are slowly varying functions of z. Let z_p, z_q be two coupling points in the domain, with $q_1 = q_2$ at both. These points can be made to coalesce by changes in one or more of the parameters S ($= \sin\theta$), l_x, l_y, ω.

It often happens that at exact coalescence the columns $\mathbf{s}_1$ and $\mathbf{s}_2$ found from columns of adj $\mathbf{C}$ (see § 17.2) are identically zero. Then $\mathbf{s}_\alpha$ (16.19) is zero, the resulting form of $\mathbf{U}$ is singular, and the transformation (16.21) cannot be used. It is possible to find non-zero $\mathbf{s}_1$ and $\mathbf{s}_2$ by finding them at some point z near $z_p = z_q$ and then dividing them by some analytic factor that tends to zero at exact coalescence, for example $z - z_p$ or $q_1 - q_2$; see end of § 17.2. But then $\mathbf{s}_1$ and $\mathbf{s}_2$ tend to different limits when $z \to z_p$, and are not parallel. Consequently $\mathbf{S}$ is non-singular throughout a domain of the complex z plane containing $z_p = z_q$, and the transformation (6.53) can be used just as if the coupling points did not exist. But, at $z = z_p = z_q$, $\mathbf{s}_\beta$ (16.18) has some infinite elements because its denominator β is zero so that again the transformation (16.21) cannot be used.

Now let the two coupling points move apart. Then at each of them $\mathbf{S}$ is singular and the transformation (6.53) cannot be used. But $\mathbf{U}$ can now be defined as in § 16.3, and is non-singular at and near both z_p and z_q. Thus the transformation (16.21) can now be used throughout a domain containing both coupling points.

Neither $\mathbf{U}$ with (16.21) nor $\mathbf{S}$ with (6.53) can be used both at exact coalescence and at near coalescence. A new transforming matrix $\mathbf{V}$ is needed instead of $\mathbf{U}$ or $\mathbf{S}$. It must be uniformly analytic and non-singular throughout a domain containing z_p

and z_q, both at exact coalescence and at near coalescence. A method of finding it is now to be given in outline. Some further results from matrix theory are needed for this. Their derivation is too long to give here but they are given in full by Budden and Smith (1974, appendix).

The columns $\mathbf{s}_1, \mathbf{s}_2$ used in (16.19), (16.18) are any column of adj $\mathbf{C}$ or they may be a linear combination of several columns provided that the same combination is used for both. Then if $\mathbf{s}_1$ and $\mathbf{s}_2$ are non-zero at a coupling point where $q_1 = q_2$ they must be parallel there. If $\mathbf{s}_\alpha$ is zero at a coupling point, including a coalescence, both $\mathbf{s}_1$ and $\mathbf{s}_2$ must be zero. This suggests that a different combination of columns of adj $\mathbf{C}$ should be used, in the hope that it will give a non-zero $\mathbf{s}_\alpha$ at the coalescence. But this cannot be done if all the elements of adj $\mathbf{C}$, that is all the cofactors of $\mathbf{C}$, are zero. This would mean that, at the coalescence, $\mathbf{C}$ is of rank two or less. If $\mathbf{C}$ were of rank one it would follow from (17.2) that $\partial^2 F/\partial q^2 = 0$, as well as (17.5), and this would mean that three roots of the quartic (17.2) are equal. This case could occur but is excluded from the present discussion. A necessary condition for a C2, therefore, is that $\mathbf{C}$ shall be of rank two. It is proved by Budden and Smith (1974) that this condition is also sufficient.

Suppose, now, that a suitable linear combination of columns of adj $\mathbf{C}$ has been chosen for $\mathbf{s}_i$. Then all four elements are analytic functions of q and of z. At a single coupling point where $q_i = q_1 = q_2$ the function $q_i(z)$ is not analytic because it has a branch point there, but α and β^2 (16.17) are analytic. By the method used in § 16.3 it can then be shown that the only linear combinations of $\mathbf{s}_1, \mathbf{s}_2$ that are analytic at and near a coupling point are some combinations of $\mathbf{s}_\alpha$, $\mathbf{s}_\beta$. Thus any new transforming matrix $\mathbf{V}$, replacing $\mathbf{U}$ in (16.21) must have columns that are all linear combinations of $\mathbf{s}_\alpha$, $\mathbf{s}_\beta$, $\mathbf{s}_3$, $\mathbf{s}_4$.

Since the $\mathbf{s}_\alpha$ found from columns of adj $\mathbf{C}$ is analytic, and zero where $z = z_p$, it must contain a factor that is some integral power of $z - z_p$. The power could be two or more if the C2 is at the maximum of an ionospheric layer. But in the cases studied here the power is unity and only this case will be considered. Hence the column

$$\mathbf{s}_\gamma = \mathbf{s}_\alpha/(z - z_p) \tag{17.15}$$

is analytic and non-zero in the domain containing z_p, z_q. At coalescence $z_p = z_q$, but $\mathbf{s}_\gamma$ is still non-zero and bounded. Both at and near coalescence $\mathbf{s}_\beta$ also is non-zero in the domain.

Now define a new transforming matrix $\mathbf{V}$ whose columns are $\mathbf{s}_\gamma$, $\mathbf{s}_\beta$, $\mathbf{s}_3$, $\mathbf{s}_4$. It is non-singular throughout the domain for conditions both at and near to coalescence. With this matrix the steps analogous to (16.20)–(16.26) are now followed through. It can be shown that

$$\mathbf{TV} = \mathbf{V}\mathfrak{B} \quad \text{where} \quad \mathfrak{B} = \begin{pmatrix} \alpha & z - z_p & 0 & 0 \\ \beta^2/(z - z_p) & \alpha & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & q_4 \end{pmatrix} \tag{17.16}$$

(compare (16.20)). Let

$$\mathbf{e} = \mathbf{V}\mathbf{d}, \quad \mathbf{d} = \mathbf{V}^{-1}\mathbf{e}. \tag{17.17}$$

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

$$\mathbf{d}' + \mathrm{i}\mathfrak{B}\mathbf{d} = \boldsymbol{\Psi}\mathbf{d} \tag{17.18}$$

where

$$\boldsymbol{\Psi} = -\mathbf{V}^{-1}\mathbf{V}'. \tag{17.19}$$

In a sufficiently slowly varying medium $\mathbf{V}'$ is small and $\mathbf{V}^{-1}$ is bounded so that the coupling matrix $\boldsymbol{\Psi}$ can be neglected to a first approximation (compare $\boldsymbol{\Gamma}$ in (7.109) and $\boldsymbol{\Lambda}$ in (16.22)). Then the waves associated with q_3, q_4 are independently propagated, so the two elements $\mathrm{d}_1, \mathrm{d}_2$ of $\mathbf{d}$ may be separated from the others to give

$$\begin{pmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \end{pmatrix}' = -\mathrm{i}k\begin{pmatrix} \alpha & z - z_\mathrm{p} \\ \beta^2/(z - z_\mathrm{p}) & \alpha \end{pmatrix}\begin{pmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \end{pmatrix} \tag{17.20}$$

(compare (16.24)). The further transformation

$$\begin{pmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}\exp\left(-\mathrm{i}k\int^z \alpha\,\mathrm{d}z\right) \tag{17.21}$$

now leads to

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}' = -\mathrm{i}k\begin{pmatrix} 0 & z - z_\mathrm{p} \\ \beta^2/(z - z_\mathrm{p}) & 0 \end{pmatrix}\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}. \tag{17.22}$$

This will now be studied first for the case where β^2 is exactly a quadratic function of z given by (17.9), so that (17.22) gives

$$\mathrm{d}g_1/\mathrm{d}z = -\mathrm{i}k(z - z_\mathrm{p})g_2, \quad \mathrm{d}g_2/\mathrm{d}z = -\mathrm{i}kb(z - z_\mathrm{q})g_1. \tag{17.23}$$

Now change the dependent variables to

$$w_1 = g_1 b^{\frac{1}{2}} + g_2, \quad w_2 = g_1 b^{\frac{1}{2}} - g_2 \tag{17.24}$$

and let

$$\xi = (k^2 b)^{\frac{1}{4}}\{z - \tfrac{1}{2}(z_\mathrm{p} + z_\mathrm{q})\}, \quad P = \tfrac{1}{2}(k^2 b)^{\frac{1}{4}}(z_\mathrm{q} - z_\mathrm{p}); \tag{17.25}$$

compare (17.10). Then the equations are

$$\mathrm{d}w_1/\mathrm{d}\xi = -\mathrm{i}(\xi w_1 - P w_2), \tag{17.26}$$

$$\mathrm{d}w_2/\mathrm{d}\xi = \mathrm{i}(\xi w_2 - P w_1). \tag{17.27}$$

If w_1 or w_2 is eliminated we obtain respectively

$$\mathrm{d}^2 w_1/\mathrm{d}\xi^2 + (\xi^2 - P^2 + \mathrm{i})w_1 = 0, \tag{17.28}$$

$$\mathrm{d}^2 w_2/\mathrm{d}\xi^2 + (\xi^2 - P^2 - \mathrm{i})w_2 = 0. \tag{17.29}$$

These are both of the form (15.68), (17.11) with $A^2 = P^2 \pm \mathrm{i}$. By a further change of independent variable they could be transformed to Weber's equation (15.70). At exact coalescence C2, $P = 0$, $A^2 = \pm\mathrm{i}$, whereas for exact coalescence C1, from (17.10), $A^2 = 0$. This is a very important difference.

When $P \neq 0$, let one of the solutions of (17.28) be selected. Then w_2 is found from

(17.26) and the recurrence relations for the solutions of these equations (Whittaker and Watson, 1927, p. 350, last formula in § 16.61) show that it also satisfies (17.29). Thence via (17.24), (17.21), (17.17) the components **e** of the electromagnetic field can be found. If $|\xi|$ is large so that z is far enough away from z_p, z_q, the appropriate W.K.B. solutions of (17.28), (17.29) can be used for w_1, w_2 and these in turn give the W.K.B. solutions for the characteristic waves. They display the Stokes phenomenon which means that the circuit relations for these functions show that there is mode coupling near z_p, z_q, between these characteristic waves.

When $P=0$ the behaviour is different. The equations (17.23)–(17.27) now have two independent exact solutions. The first is

$$w_1 = \exp(-\tfrac{1}{2}\mathrm{i}\xi^2), \quad w_2 = 0. \tag{17.30}$$

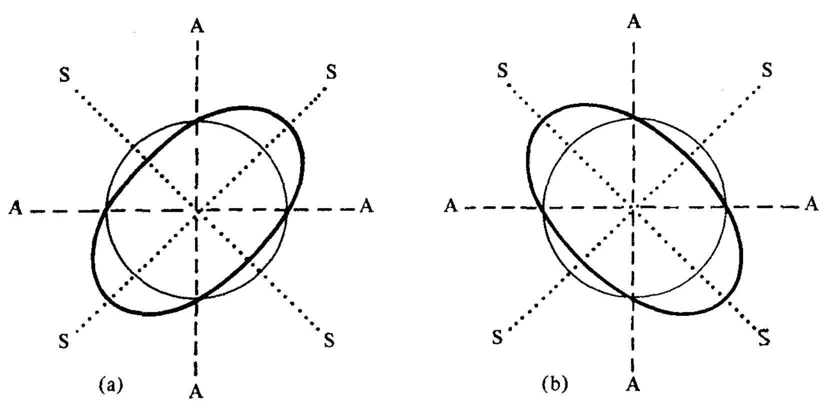
It is dominant on the Stokes lines where $\arg \xi = \frac{1}{4}\mathrm{i}\pi, \frac{5}{4}\mathrm{i}\pi$. Because it is exact there can be no subdominant term here and no Stokes phenomenon. The expressions given by Heading (1962a, p. 67) for the Stokes multipliers show that they are zero here. The Stokes multipliers for $\arg \xi = \frac{3}{4}\mathrm{i}\pi, -\frac{1}{4}\mathrm{i}\pi$ are not zero but here the solution (17.30) is subdominant. The Stokes diagram is shown in fig. 17.1(a). The thick curve has no breaks because the Stokes phenomenon does not occur at a C2. The second solution is

$$w_1 = 0, \quad w_2 = \exp(\tfrac{1}{2}\mathrm{i}\xi^2). \tag{17.31}$$

Its Stokes diagram fig. 17.2(b) has a continuous thick curve because, again, the Stokes phenomenon does not occur.

When β^2 is not exactly a quadratic function of z, a uniform approximation similar to (17.12), (17.13), but not quite the same, may be used. Instead of (17.25) let ξ and P

Fig. 17.1. Stokes diagrams for solutions of (17.26), (17.27), that is of (17.28), (17.29), when $P=0$. S denotes Stokes and A denotes anti-Stokes lines.



be defined by

$$k\int_{z_p}^{z}\beta\,dz = i\int_{-P}^{\xi}(P^2-\xi^2)^{\frac{1}{2}}\,d\xi, \tag{17.32}$$

$$k\int_{z_p}^{z_q}\beta\,dz = i\int_{-P}^{P}(P^2-\xi^2)^{\frac{1}{2}}\,d\xi = \tfrac{1}{2}i\pi P^2. \tag{17.33}$$

Then (17.24) and (17.26)–(17.29) must be modified by the inclusion of small terms. These are zero when β^2 is exactly quadratic. With a suitable choice of the signs of the square roots, (17.32), (17.33) are then the same as (17.25).

At exact coalescence C2, $P = 0$ and (17.32) shows that

$$\exp(\pm\tfrac{1}{2}i\xi^2) = \exp\left(\pm ik\int_{z_p}^{z}\beta\,dz\right). \tag{17.34}$$

Let this now be used in the two solutions (17.30), (17.31). For one of them the factor $\exp(-ik\int^z\beta\,dz)$ appears in g_1, g_2 from (17.24), and the transformation (17.21) shows that d_1, d_2 and thence **e**, from (17.17) contain the factor $\exp(-ik\int^z q_1\,dz)$, where (16.17) has been used. Similarly for the other solution the field components **e** contain the factor $\exp(-ik\int^z q_2\,dz)$. These are just the exponential factors that appear in the W.K.B. solutions for the two waves associated with q_1 and q_2 respectively.

At a C2, therefore, there is no coupling between these two waves. They are independently propagated and their fields are given by the W.K.B. solutions (7.112) for points z at and near $z_p = z_q$. This occurs because the Stokes phenomenon is absent. A physical explanation is that at a C2, if the eigen columns $\mathbf{s}_i$ are constructed by dividing some column of adj **C** by a suitable factor, then the columns $\mathbf{s}_1$, $\mathbf{s}_2$ are not parallel even though $q_1 = q_2$. This means that the two waves have different polarisations and wave impedances, just as they do at an ordinary point where the waves are distinct, and so they are independently propagated. It also means that at *exact* coalescence the matrix **S** in (7.85) is non-singular at all points in a domain containing $z_p = z_q$ so that the transformation (6.53) could be used. This behaviour should be contrasted with a C1 where the polarisations are the same at a coalescence, and the coupling is strong.

For conditions that are near but not exactly at C2, one pair of Stokes multipliers for each of the equations (17.28), (17.29) is still small but not exactly zero. Thus there is now a small amount of coupling. It gets stronger as the coupling points move further apart.

The best known example of coalescence C2 is the phenomenon of radio windows. This is described in §§ 17.6–17.9. Another example of a rather different kind is limiting polarisation, §§ 17.10, 17.11.

When conditions are near to coalescence C2, the physical description of what happens to the waves depends on $\arg(z_q - z_p)$. To illustrate this it is useful to discuss two extreme special cases. For the first $\arg(z_q - z_p) = \pm\frac{1}{2}\pi$, which arises in the study of ion cyclotron whistlers, § 17.5. For the second $\arg(z_q - z_p) = 0$ or π, which arises in the study of radio windows, § 17.6.

17.5. Ion cyclotron whistlers

The theory of ion cyclotron whistlers was given in § 13.9. It was studied for vertically incident waves and was associated with two coupling points z_p, z_q shown in fig. 13.12. Their position in the complex z plane is given by

$$\mathfrak{G}(z) = \pm \mathrm{i} \sin^2\Theta/\cos\Theta \tag{17.35}$$

where Θ is the angle between the earth's magnetic field and the vertical, and $\mathfrak{G}$ is given by (13.60). The two points coalesce only if $\Theta = 0$, that is at a magnetic pole of the earth. In practice when collisions are allowed for $\mathfrak{G}$ is complex when z is real, and the condition $\mathfrak{G} = 0$ requires a complex value of z. When $\Theta = 0$, therefore, the coalescence is not on the real z axis, and has only a small influence on the propagation. An upgoing electron whistler wave remains an electron whistler at all heights, and ion whistlers are not formed near the poles. When Θ differs from zero the two coupling points separate and for large enough Θ one of them, z_p in fig. 13.12(c), is on the real z axis. This is the condition of exact crossover, § 13.9, and is an example of critical coupling, § 16.6. For still larger Θ, $\mathrm{Im}(z_p)$ is positive, and an upgoing electron whistler now goes over continuously to an ion cyclotron whistler at points above $\mathrm{Re}(z_p)$. In the description given in § 13.9, the other coupling point z_q did not come into the discussion because it was assumed to be too far away from the real z axis.

If collisions were negligible the behaviour would be different, for then $\mathfrak{G}$ is real when z is real. The coalescence would be on the real z axis when $\Theta = 0$, and it would be necessary to consider both coupling points. For $\Theta \neq 0$, $\mathrm{Im}(z_p)$ is now always positive. Although the neglect of collisions is unrealistic, the description of what would happen to the waves is a useful illustration of a coalescence C2, and it is therefore given briefly here.

The dependence of the two values of n^2 on frequency was sketched in fig. 13.8(b). The continuous curves are for $\Theta = 0$, and the dotted curves for Θ non-zero but small. Similarly curves of n^2 versus height z can be plotted and are sketched in figs. 17.2. Both values of n^2 are real for all real z. For $\Theta = 0$, fig. 17.2(a), $n_O^2 = \varepsilon_2$ and $n_E^2 = \varepsilon_1$. The curves cross at a height $z = z_0$ and this is the point of coalescence. The ordinary wave is the electron whistler and has left-handed circular polarisation $\rho = \mathrm{i}$ at all heights. The other wave has right-handed circular polarisation $\rho = -\mathrm{i}$ at all heights. At exact coalescence there is no coupling between these waves and so no formation of an ion cyclotron whistler.

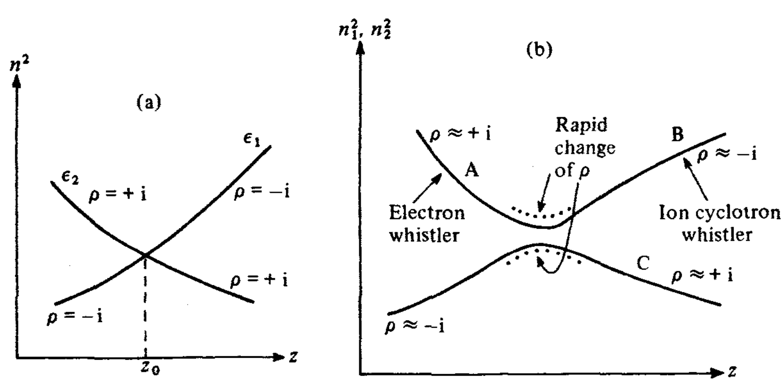
When Θ is small but non-zero the curves are as in fig. 17.2(b). The two coupling points have now separated slightly and are at complex conjugate values of z. There is no real value of z where the two values of n^2 are equal. This is to be contrasted, in §17.6, with fig. 17.3(b) where the slightly separated coupling points are both on the real z axis.

Suppose that an electron whistler wave travels up from below. Its squared refractive index is given by the part A of the upper curve in fig. 17.2(b), and its polarisation is very nearly left-handed circular $\rho \approx +\mathrm{i}$. It then encounters the coupling region near z_0. If the medium were extremely slowly varying, the squared refractive index would go continuously over to the part B of the curve and the polarisation would reverse its sense and go over to $\rho \approx -\mathrm{i}$. The electron whistler then changes its name to ion cyclotron whistler and there would be no mode conversion. But in practice this does not happen when Θ is small. In the coupling region the characteristic polarisation for the branch AB is changing very rapidly when z increases, even in a moderately slowly varying medium. Thus there is strong coupling with branch C, fig. 17.2(b), whose characteristic polarisation is nearly the same as for branch A. This topic is elucidated in more detail by Jones, D. (1970), and Wang (1971).

In this case therefore, mode conversion is the mechanism that makes an electron whistler remain an electron whistler. It is strong when Θ is small, and at coalescence when $\Theta = 0$ it dominates completely so that there is no ion cyclotron whistler. This will be called 'complete mode conversion'. For larger Θ the curves in fig. 17.2(b) are more widely separated, the polarisations change more slowly with changing z, and the coupling is much weaker. The electron whistler at great heights changes its name to ion cyclotron whistler and there is no mode conversion.

In this example, the two waves that are coupled at near coalescence are both vertically upgoing waves so that the phenomenon is coupling and not reflection.

Fig. 17.2. Dependence of n^2 on height z when collisions are neglected. In (a) the earth's magnetic field is vertical, $\Theta = 0$, and the frequency is the crossover frequency for the height z_0. At z_0 there is a C2. In (b) Θ is non-zero but small. Near z_0 conditions are near to, but not exactly at C2.



It should be contrasted with the example in the following section where the wave normals are oblique, and one wave is upgoing and the other downgoing.

It is paradoxical that complete mode conversion is invoked to explain why, at coalescence, an electron whistler remains an electron whistler at all heights, but this is a consequence of how 'mode conversion' is defined. For the two waves of fig. 17.2(b), the curves of n^2 versus z on the real z axis are continuous, and these waves are therefore taken to be the unconverted waves. This definition is retained in the limit of coalescence, fig. 17.2(a), where the two curves now touch and have a discontinuity of gradient. In the following section a case will be studied where the mode conversion at exact coalescence is a true and complete conversion from ordinary to extraordinary. When the coupling points are well separated the 'conversion' is a complete reflection of an upgoing ordinary wave to give a downgoing ordinary wave. For further details of the example in this section see Budden and Smith (1974) who give expressions for the columns $\mathbf{s}_\gamma$, $\mathbf{s}_\beta$ used in (17.16)–(17.19).

17.6. Radio windows 1. Coalescence

The most important practical application of the coalescence C2 is to the theory of radio windows and particularly the Ellis window; see end of § 16.8 and Ellis (1956). In this section a simple example is given to illustrate further the nature of a coalescence C2. The fuller theory is given in §§ 17.7–17.9.

Consider a stratified ionosphere and radio waves of frequency greater than f_H, so $Y < 1$. Thus only electrons need be considered and collisions will be neglected. The incident wave is an upgoing ordinary wave with its wave normal in the magnetic meridian plane, so $l_y = 0$. It is assumed that X is a monotonically increasing function of the height z. The angle of incidence is $\theta = \arcsin S$. We shall here study what happens when S is varied, with all other parameters held constant. Note that, in § 17.5, S was constant (zero) and it was the angle Θ between the earth's magnetic field and the vertical that was varied.

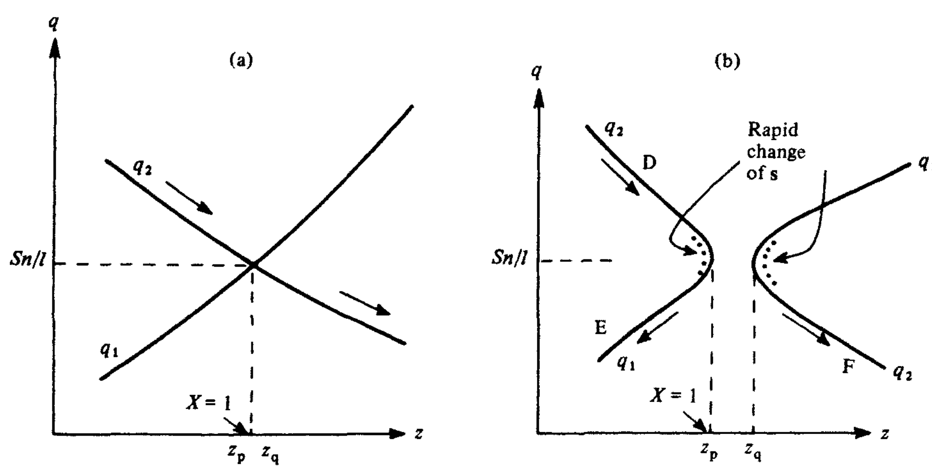
The coupling points in the complex X plane were studied for this case in § 16.5 and it was shown that when S is near to $S_A = l_x\{Y/(Y+1)\}^{\frac{1}{2}}$, (6.47), the two coupling points O and C_3 are on the real X axis. They are also on the real z axis at z_p, z_q respectively. They coalesce where $X = 1$ when $S \equiv S_A$. The formal proof that this is a coalescence C_2 was given by Budden and Smith (1974) who also discuss the expressions for the column matrices $\mathbf{s}_1$, $\mathbf{s}_2$ to be used in forming the transforming matrices $\mathbf{U}$, (16.20)–(16.23) and $\mathbf{V}$, (17.16)–(17.19).

Curves of q versus X for this case, with various values of S, were given in fig. 6.6, and fig. 6.6(c) applies for $S = S_A$. The relevant parts of these curves are sketched again in fig. 17.3 as curves of q versus z. Fig. 17.3(a) is for exact coalescence C2, that is $S = S_A$, and fig. 17.3(b) is for S slightly less than S_A. The coupling points z_p, z_q

are on the real z axis where the curves have vertical tangents in fig. 17.3(b). For $z < z_p$ the part of the curve marked q_2 is for the upgoing ordinary wave and the part q_1 is for the downgoing ordinary wave. Thus in fig. 17.3(b), $z = z_p$ is a reflection level for this wave. At each point on the curve the column matrix $\mathbf{s}$, meaning $\mathbf{s}_1$ or $\mathbf{s}_2$ (7.134) can be found, and it gives the polarisation and wave admittance of the wave field. Its elements are continuous as we proceed along the curve. Similarly for $z > z_q$ the part q_1 of the curve is for a downgoing extraordinary wave, that is a Z-mode wave, and the part q_2 is for an upgoing extraordinary wave. Again the column matrix $\mathbf{s}$ can be found for each point on this curve.

Between z_p and z_q is a region where q_1 and q_2 are complex conjugates. One of the waves decays and the other grows when z is increasing. This region, therefore, is like a potential barrier, and waves can tunnel through it. It is, however, different from the conventional simple barrier studied in § 15.9. When z_p and z_q are well separated, the barrier is wide and the tunnelling is weak. The two reflection processes at z_p and z_q then predominate. They are special cases of strong coupling between upgoing and downgoing waves. When $|z_q - z_p|$ is smaller the curves are as in fig. 17.3(b). There is some tunnelling because for the incident wave the column $\mathbf{s}$ changes rapidly as the part of the curve D near z_p is traversed. The value of $\mathbf{s}$ for curve D near z_p is closer to the value for curve F near z_q than it is to the value for curve E near z_p. Thus as the incident wave approaches z_p it is better matched in wave admittance and polarisation to the wave of branch F than it is to the reflected wave of branch E, and there is some mode conversion with tunnelling.

Fig. 17.3. Dependence of q on height z when collisions are neglected. In (a) $S = S_A$ and there is exact coalescence C2 for $z = z_p = z_q$. In (b) S is slightly less than S_A and conditions are near to, but not exactly at coalescence. Fig. (b) should be contrasted with fig. 17.2b.



The incident wave is ordinary but the wave that emerges above z_q is extraordinary. When $z_q - z_p$ is zero, the barrier has closed to zero width. The matching of the wave admittance and polarisation, as given by $\mathbf{s}$, for the upgoing ordinary wave and the upgoing extraordinary wave is now complete at $z = z_p = z_q$, and this $\mathbf{s}$ is very different from that for the downgoing waves. There is now no reflection of the ordinary wave. It continues on through the level $z = z_p = z_q$ and above this its name is changed to extraordinary. This wave is represented in fig. 17.3(a) by the continuous curve marked with arrows. Its fields are given by one of the W.K.B. solutions (7.112) at all heights including $z = z_p = z_q$ and its neighbourhood. It does not now undergo reflection, and therefore the downgoing wave is absent.

The amplitude of the transmitted wave above $z = z_q$ is therefore greatest at exact coalescence when $S = S_A$, $l_y = 0$, and this specifies a particular direction of the wave normal of the incident wave, to be called the 'central direction', or centre of the window. If now this wave normal direction is changed slightly so that it makes an angle η with the central direction, either S or l_y or both are changed. The coalescence is no longer exact and the transmitted wave is weaker. There is a range of angles η within a cone near the central direction such that the amplitude of the transmitted wave is greater than half its maximum value. This cone is called a 'radio window'. Note that it refers to a range of angles measured in free space. In practice for the ionosphere at frequencies greater than f_H, the cone is roughly circular with semi-vertical angle about 1.5°. The size of the cone depends on the gradient $\mathrm{d}N(z)/\mathrm{d}z$ of electron concentration at the level where $X = 1$, on the inclination arcsin S_A of the central direction to the vertical and on f and f_H. It is found to be almost independent of the collision frequency (neglected in this section). A method of calculating it and formulae (17.51), (17.52), are given in §17.7.

17.7. Radio windows 2. Formulae for the transparency

The meaning of the term 'radio window' was given in the preceding section, together with a description of the physical processes that the waves undergo near the level of the coupling points that cause it. The present section gives a more detailed theory and supplies formulae (17.51), (17.52) for calculating the transparency and angular width of a window. The following treatment is based on a paper by Budden (1980).

The idea of a 'window' for radio waves was first used by Ellis (1956, 1962) who applied it to explain the Z-trace; § 16.8. Several other applications have been suggested recently. Thus Jones, D. (1976a, 1977a, b) has suggested that they may play a part in one explanation of the terrestrial non-thermal continuum radiation (Gurnett and Shaw, 1973), and of the terrestrial kilometric radiation (Gurnett, 1975). These radiations are observed outside the earth and may possibly come from Cerenkov emission by energetic charged particles that enter the magnetosphere.

It has been suggested (Oya, 1974; Benson, 1975; Jones, D., 1976a, 1977a, b) that a similar mechanism might operate for the decametric radiation from Jupiter, and for the radio emission from pulsars. The properties of radio windows have been used in studies of radio emission from the solar corona and other astrophysical sources (see Melrose, 1974a, b, 1980, who gives further references). Radio windows are also of interest in the theory of thermo-nuclear plasmas (see, for example, Weitzner and Batchelor, 1979).

The theory is given here for a stratified cold plasma and for frequencies great enough to ensure that only electrons need be considered. Electron collisions are neglected, but their effect is mentioned at the end of this section. The transparency of a window depends on the direction of the wave normal of the incident wave. But any wave is refracted, so that the wave normal direction varies with level in the plasma. A reference level is needed that can be used to specify the wave normal direction unambiguously, and we choose for this a region of free space, as was done in § 17.6. This is a useful reference even if the waves never enter such a region. Various directions of the wave normal have to be considered and it is therefore best to assume that the earth's magnetic field is in the x–z plane, so that $l_y = 0$. Then S_1, S_2 are used for the x and y direction cosines of the wave normal of the incident wave in the reference free space; compare §§ 6.2–6.4. It is also assumed, at first, that the source is infinitely distant so that S_1, S_2 are necessarily real; see § 14.9, item (9). This applies for the astrophysical applications mentioned above. For sources on or near the earth it is necessary, in general, to use complex S_1, S_2 even when collisions are neglected. This is an application of complex rays (Budden and Terry, 1971, §§ 5, 6) and is mentioned in § 17.8.

Consider a wave that starts as an upgoing ordinary wave at some real level $z = z_a$ well below the coupling points z_p, z_q, and travels through the window to a real level $z = z_b$ well above the coupling points. Here it is an upgoing extraordinary wave. Suppose that there is a 'good path' Γ from z_a to z_b in the complex z plane, such that at all points on it the W.K.B. solution for this wave can be used. Then the ratio of the moduli of the variable f, (6.53) at z_b and z_a for this wave is, from (7.111)

$$\mathscr{W} = \exp\left\{\mathrm{Re}\left(-ik\int_{z_a\,(\Gamma)}^{z_b} q\,\mathrm{d}z\right)\right\}. \tag{17.36}$$

To show that there is such a good path, the Stokes diagram for the solutions must be studied. It can be shown that near z_a the W.K.B. term for the upgoing ordinary wave becomes more dominant when $\mathrm{Im}(z)$ increases; the reader is advised to check this. Hence for exact coalescence the Stokes diagram is as in fig. 17.1(b). If coalescence is not exact there is also a downgoing ordinary wave where $z \approx z_a$ and the Stokes diagram is as in fig. 17.4(b). In the region where $\mathrm{Im}(z)$ is positive it includes a continuous curve linking the incident ordinary wave and the resulting extraordinary wave. This shows that the W.K.B. solution is not affected by the Stokes phenomenon. There is a good path Γ as shown in fig. 17.4(a). This contour can now be distorted provided that no singularities of the integrand in (17.36) are crossed. Thus it can run along the real z axis, provided that it is indented near the branch points z_p and z_q of q, as in fig. 17.4(a). But now in the ranges z_a to z_p and z_q to z_b, q is purely real, since collisions are neglected. Hence (17.36) becomes

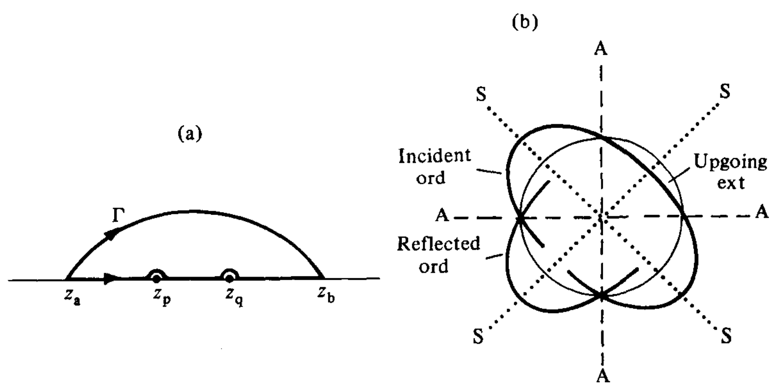
$$\mathscr{W} = \exp\left\{k\int_{z_p}^{z_q} \mathrm{Im}(q)\mathrm{d}z\right\}. \tag{17.37}$$

A simple formula for this, (17.51), (17.52) below, is now to be derived. At exact coalescence $z_p = z_q$ and $\mathscr{W} = 1$. For other conditions $\mathscr{W}$ is a measure of the transparency of the window.

The change of X in going from one edge of a window barrier to the other, that is from z_p to z_q in figs. 17.3(b) or 17.4(a), is of order 0.02 or less in practical cases; compare figs. 6.6(b, d). Thus the maximum thickness of a window that has appreciable transparency is about 1/50 of the scale height for the variation of the electron concentration $N(z)$. In the ionosphere this means that window barriers are unlikely to be thicker than about 1 km. In such small thicknesses it is useful to assume that $N(z)$ is linear and that the earth's magnetic field is constant. These assumptions are made here. In the magnetosphere the barriers are much thicker but there are many cases where the same assumption is useful. In other cases the barrier thickness can be several hundred km (Herring, 1980, 1984) and a separate theory is needed but not given here (see Weitzner and Batchelor, 1979).

Figs. 10.4, 10.7, 10.9 showed cross sections of the refractive index surfaces by the

Fig. 17.4. (a) is the complex z plane and shows possible contours for the integral in (17.36). (b) is the Stokes diagram when z_p and z_q are slightly separated, so that conditions are near but not exactly at coalescence C2. It should be compared with fig. 17.1.



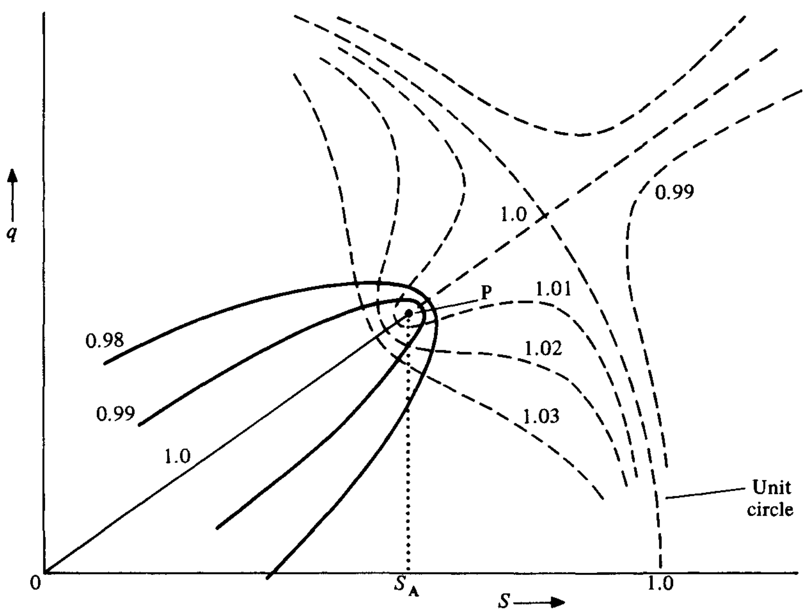
magnetic meridian plane when $Z = 0$, and they were drawn with the direction of the earth's magnetic field at the correct angle to the vertical axis. Thus the direction cosines of the common axis of the refractive index surfaces are

$$l_x, \quad 0, \quad l_z, \tag{17.38}$$

and l_x, l_z are both assumed to be positive.

For any point on one of the curves the radius from the origin is the refractive index n, the abscissa is $n \sin\psi = \sin\theta = S_1$ and the ordinate is $n \cos\psi = q$; see § 10.8. For these figures $S_2 = 0$ because the wave normal is in the magnetic meridian plane. It was shown in ch. 5 that the surfaces for the ordinary and extraordinary waves cannot have any points in common when $X \neq 1$ and $X \neq 0$. When $Y < 1$ and $X \to 1$ the surface for the ordinary wave shrinks down to the line PQ forming part of the symmetry axis, and the surface for the extraordinary wave shrinks to the remaining parts of that axis together with the unit sphere; see § 5.2. Thus the two surfaces touch only when $X = 1$ and then only at the two window points P, Q. The form of these curves near the window point P is sketched in fig. 17.5. When $Y > 1$ and $X \to 1$ the surface for the ordinary wave shrinks to the line PQ as before, together with the parts of the symmetry axis extending to infinity beyond the two other window points P_2, Q_2; figs. 10.9, 17.7.

Fig. 17.5. Cross section by the magnetic meridian plane of parts of typical refractive index surfaces (not to scale) for $Y < 1$. Continuous curves are for the ordinary wave, and broken curves are for the extraordinary wave in the Z-mode. The numbers by the curves are the values of X.



The surface for the extraordinary wave then shrinks to the segments PP_2 and QQ_2 of the symmetry axis, together with the unit sphere. Thus the surfaces touch only when $X = 1$ and then only at the four window points P, Q, P_2, Q_2. The form of the surfaces near P_2 is sketched in fig. 17.7. In this section the effect of the window near P is studied. Similar methods could be applied for Q. The points P_2, Q_2 are real only if $Y > 1$ and their effect is then sometimes called a 'second window'. The point P_2 is studied in § 17.9 and again similar methods could be applied for Q_2.

The refractive index surfaces are surfaces of revolution about the symmetry axis. Inspection of fig. 17.5 shows that near P they are approximately paraboloids. A quadratic approximation will now be derived that gives the equation of these paraboloids. An almost identical approximation has been used by Melrose (1980). Fig. 17.6 shows a cross section by any plane containing the symmetry axis of two typical refractive index surfaces near P. The line OA is the refractive index n and the angle $AOP = \Theta$ is the angle between the wave normal and the earth's magnetic field, the same as the Θ of chs. 4, 5. At P

$$n = n_1 = \{Y/(Y+1)\}^{\frac{1}{2}}, \quad \Theta = 0 \tag{17.39}$$

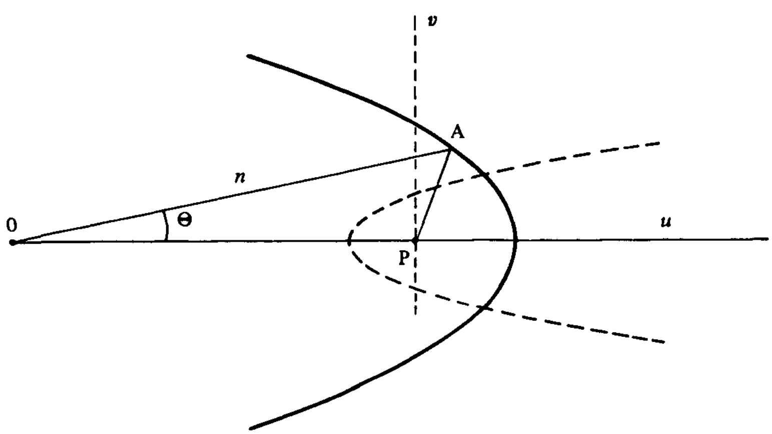
from § 5.2(7). Let

$$X = 1 + \sigma \tag{17.40}$$

where $|\sigma|$ is small. Choose new perpendicular axes u, v with the origin at P and with u along the symmetry axis. The form (4.65) of the dispersion relation is

$$An^4 - 2Bn^2 + C = 0 \tag{17.41}$$

Fig. 17.6. Cross section, by any plane containing the earth's magnetic field direction OP, of part of two refractive index surfaces near the window point P where they are, with good approximation, parabolae. The continuous curve is for $X < 1$ and the broken curve for $X > 1$.



where, for $U = 1$ and with (17.40)

$$A = -\sigma - Y^2 + (1+\sigma)Y^2\cos^2\Theta, \quad B = \sigma^2 + \sigma Y^2 - \tfrac{1}{2}(1+\sigma)Y^2\sin^2\Theta,$$
$$C = Y^2\sigma - \sigma^3 \tag{17.42}$$

from (4.66). Now

$$n\cos\Theta = n_1 + u, \quad n\sin\Theta = v. \tag{17.43}$$

Eliminate n and Θ from (17.41)–(17.43) and neglect all products of the small quantities u, v, σ of degree 3 or greater. This gives

$$v^2 - 4\sigma u/n_1 - 2\sigma^2/Y = 0 \tag{17.44}$$

which is the required approximation. Further

$$\boldsymbol{n}\cdot\boldsymbol{Y} = nY\cos\Theta = Y(l_x S_1 + l_z q) \tag{17.45}$$

where (6.13) has been used, whence from (17.43)

$$u = l_x S_1 + l_z q - n_1, \quad v^2 = q^2 + S_1^2 + S_2^2 - (l_x S_1 + l_z q)^2 \tag{17.46}$$

and substitution in (17.44) gives

$$q^2 l_x^2 - q\{4\sigma l_z/n_1 + 2S_1 l_x l_z\} + S_1^2 l_z^2 + S_2^2 - 4\sigma S_1 l_x/n_1 - 2\sigma^2/Y + 4\sigma = 0. \tag{17.47}$$

This quadratic for q is the approximation, near the window, for the Booker quartic equation. Its solutions are

$$q = \frac{l_z}{l_x}\left(S_1 + \frac{2\sigma}{S_A}\right) \pm \frac{1}{l_x}\left\{\sigma^2\left(\frac{2}{Y} + \frac{4l_z^2}{S_A^2}\right) - 4\sigma\left(1 - \frac{S_1}{S_A}\right) - S_2^2\right\}^{\frac{1}{2}} \tag{17.48}$$

where (6.47) for S_A has been used. The square root in (17.48) has two zeros where $\sigma = \sigma_p$, σ_q say. These correspond to the points in fig. 17.4(a) where $z = z_p$, z_q. For $\sigma_p < \sigma < \sigma_q$ the square root is imaginary and the waves are attenuated as they pass through the barrier. Outside this range q is real and the waves are unattenuated. The curves of q vs $\sigma = X - 1$ then have the shape shown in figs. 6.6(b, c, d) for $X \approx 1$.

The formulae (17.47), (17.48) are useful when arctan (l_x/l_z) exceeds about 10°. For smaller l_x they are not so reliable. Then both the earth's magnetic field and the wave normal are nearly vertical. This case has been treated by Budden and Smith (1973) but is not given in this book.

Suppose, now, that the electron concentration $N(z)$ varies linearly with z in the region of the window so that

$$\sigma = \frac{e^2 G z}{\varepsilon_0 m\omega^2} + \text{constant, where } G = \frac{\mathrm{d}N}{\mathrm{d}z}. \tag{17.49}$$

The attenuation of the waves by the barrier in nepers is, from (17.37)

$$-\ln\mathscr{W} = -\frac{\varepsilon_0 m\omega^2 k}{e^2|G|}\int_{\sigma_p}^{\sigma_q}\mathrm{Im}(q)\mathrm{d}\sigma. \tag{17.50}$$

Substitution of the square root in (17.48) for Im(q) gives finally

$$-\ln \mathscr{W} = 0.204\frac{f_H^{\frac{1}{2}} f^{\frac{5}{2}}}{|G|}\{J_1(S_1 - S_A)^2 + J_2 S_2^2\} \tag{17.51}$$

where f and f_H are in kHz, G is in m^{-4} and

$$J_1 = (Y+1)\{l_z^2(Y+\tfrac{1}{2}) + \tfrac{1}{2}\}^{-\frac{3}{2}}, \quad J_2 = \{l_z^2(Y+\tfrac{1}{2}) + \tfrac{1}{2}\}^{-\frac{1}{2}}. \tag{17.52}$$

to give the transparency of the window in decibels the numerical factor 0.204 in (17.51) should be replaced by 1.772.

These formulae show that the window is completely transparent when $S_1 = S_A$, $S_2 = 0$. The half angular width of the window may be defined as the values of $|S_1 - S_A|$ with $S_2 = 0$, or of S_2 with $S_1 = S_A$, that make the attenuation equal to 6dB, that is half power. These values are respectively, for $S_2 = 0$:

$$|S_1 - S_A|^2 \approx \frac{3.386|G|\{l_z^2(Y+\frac{1}{2}) + \frac{1}{2}\}^{\frac{3}{2}}}{f_H^{\frac{1}{2}} f^{\frac{5}{2}}(Y+1)} \tag{17.53}$$

and for $S_1 = S_A$:

$$S_2^2 \approx \frac{3.386|G|\{l_z^2(Y+\frac{1}{2}) + \frac{1}{2}\}^{\frac{1}{2}}}{f_H^{\frac{1}{2}} f^{\frac{5}{2}}}. \tag{17.54}$$

One of the solutions (17.48) is for waves travelling in the direction of X increasing and the other for the opposite direction. Since the imaginary parts are equal with opposite signs it follows that (17.50)–(17.54) apply for waves going through the window in either direction. Thus in the lower ionosphere they apply for a downgoing extraordinary wave in the Z-mode partially penetrating the window to give a downgoing ordinary wave below it.

In deriving (17.50)–(17.54) several approximations were used. The formulae were therefore tested (Budden, 1980) by comparing them with the results of calculations made by numerical integration of the governing differential equations for various values of S_1 and S_2. The agreement was found to be good. In Budden's formulae the factor $1/l_x$ in (17.48) was omitted in error. This is the same as cosec Θ in his notation – not the same Θ as used here. Thus in Budden's (1980) formulae (22), (23) a factor cosec Θ should be inserted, and in (24), (25) the sin Θ in the denominator should be deleted. In Budden's numerical example cosec Θ was 1.113. Insertion of this factor results in even better agreement with the numerically computed results.

When electron collisions were included in the numerical calculations it was found that the resulting additional attenuation was about the same for all values of S_1, S_2 near the window. Thus collisions did not alter the angular widths, and (17.53), (17.54) can still be used.

17.8. Radio windows 3. Complex rays

In §§17.6, 17.7 it was assumed that the source of the waves approaching a radio window is infinitely distant, so that the incident wave when in free space is a plane wave. The rays associated with it are the family of lines that are parallel in free space and whose x and y direction cosines S_1, S_2 are real and give the direction of the source. These various rays impinge on any plane $z = \text{constant}$ at various values of x and y, including complex values, as explained in § 14.9 items (9), (10). On entering the ionosphere the rays are refracted. If they enter an attenuating region, the refractive index n is complex and the rays are refracted at complex angles. Let a particular ray of the family pass through the point with coordinates x_1, y_1 before it enters the ionosphere. Suppose that this ray gets to a receiver with real coordinates x_2, y_2 in or beyond the attenuating region. Then x_1 or y_1 or both must be complex. A simple example of this was given in § 14.9 item (9).

The barrier of a radio window is like an attenuating medium in which the variable q, given by (17.48), is complex within the barrier. Thus for a ray that has real x, y before it enters the barrier, either x or y or both must be complex when z is real, after this ray leaves the barrier. For a ray that goes through the barrier and arrives at a receiver with real x, y, either x and y or both must be complex before this ray enters the barrier. Rays of this kind can be traced through the barrier and Budden (1980) gave some examples.

The use of complex rays for this problem is a form of the phase integral method; see § 17.7. It was shown that the results of this method are in agreement with the full numerical calculations, and therefore also with the approximate formulae (17.50)–(17.54) (Budden, 1980, § 8).

Suppose now that the source is not infinitely distant. For example it may be a point transmitter at the ground. Rays now leave it in all directions including complex directions. A ray that starts with real direction cosines and enters an attenuating part of the ionosphere is refracted to complex directions and subsequently has complex values of x or y or both when z is real. Such a ray cannot therefore reach a receiver with real x, y. The ray that gets to a receiver at a real point above the attenuating region must leave the source in a complex direction.

We now continue to assume that collisions are negligible so that the ionospheric plasma is loss-free. But in a window barrier q is complex and the waves are attenuated. The above conclusions for an attenuating region therefore apply also for a window barrier. The only exception is a ray that enters the window in its central direction. For this ray the barrier has zero thickness and there is no attenuation. A real ray with this direction can now go through real space from transmitter to receiver.

Any other ray that goes through the window must have a complex direction so that on it x or y or both are complex when z is real, except at the transmitter where it starts out and at the receiver where it arrives. This result was described by Budden and Terry (1971, § 6) who illustrated it with some numerical results.

For complex S_1, S_2 the formulae (17.36), and (17.39)–(17.49) can still be used, where Θ, n, u, v, are now complex. But (17.37) and (17.50)–(17.54) cannot be used and more complicated formulae are needed.

17.9. Radio windows 4. The second window

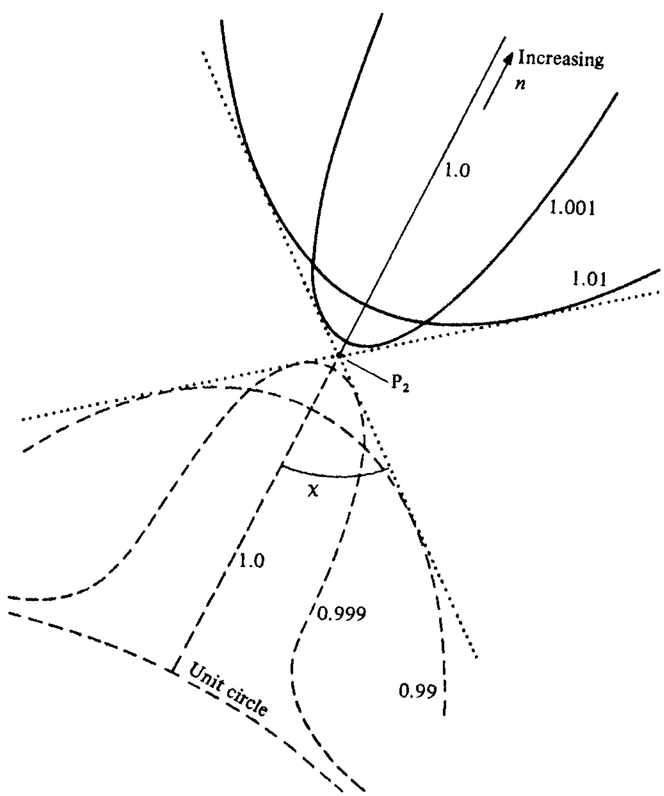
A second type of radio window associated with the window points P_2, Q_2, § 10.11, fig. 10.9, can occur if $Y > 1$ that is $f < f_H$. Fig. 17.7 shows part of the system of refractive index surfaces near a typical window point P_2. These can again be approximated by paraboloids of revolution as in § 17.7. It is only necessary to replace the value n_1 of n at P, (17.39) by its value at P_2 namely

$$n_2 = \{Y/(Y-1)\}^{\frac{1}{2}}, \quad \Theta = 0. \tag{17.55}$$

The final result is that (17.51) is still true, and (17.52) is replaced by

$$J_1 = (Y-1)\{l_z^2(Y-\tfrac{1}{2})-\tfrac{1}{2}\}^{-\frac{3}{2}}, \quad J_2 = \{l_z^2(Y-\tfrac{1}{2})-\tfrac{1}{2}\}^{-\frac{1}{2}}. \tag{17.56}$$

Fig. 17.7. The second window. Cross section by the magnetic meridian plane of parts of typical refractive index surfaces (not to scale) near the second window point P_2, for $Y > 1$. Continuous curves are for the ordinary wave in the whistler mode, and broken curves are for the extraordinary wave. The numbers by the curves are the values of X.



The refractive index surfaces near P_2 and Q_2 have a real envelope that is a cone with its apex at P_2 or Q_2. It can be shown (see problem 5.9) that at the apex the generators make with the axis an angle χ given by

$$\tan\chi = \{2(Y-1)\}^{\frac{1}{2}}. \tag{17.57}$$

In fig. 17.7 it is assumed that the medium is horizontally stratified. A vertical line may be drawn in this diagram, as for Poeverlein's construction § 10.8, whose abscissa is $S = \sin\theta$ where θ is the angle of incidence. Then, at any level, the point where it cuts a refractive index surface has ordinate q. When this line goes through P_2, the wave normal of the incident wave has the central direction of the window. If this line is to cut real refractive index surfaces when $X \neq 1$, it is clearly necessary that χ shall exceed the angle arctan (l_x/l_z) between the earth's magnetic field and the vertical, so that

$$l_x/l_z < \{2(Y-1)\}^{\frac{1}{2}}, \quad \text{that is} \quad Y > \tfrac{1}{2}(1 + l_z^{-2}). \tag{17.58}$$

This is also the condition that the expressions (17.56) shall be real. If this is not satisfied, no waves coming from where $X \neq 1$ can reach the level of the window. In the ionosphere this means that the second window can only be effective at high enough latitudes.

Jones, D. (1976b) has suggested that the second window might be a possible mechanism for explaining the propagation of auroral hiss.

17.10. Limiting polarisation 1. Statement of the problem

In the earlier chapters it has been assumed that when a characteristic wave, ordinary or extraordinary, travels through the ionosphere or magnetosphere its polarisation and wave admittance are given at most levels z by the W.K.B. solutions (7.112). There are regions of failure near coupling points in the complex z plane, and some of these have been studied in ch. 16 and the present chapter. There is another region where failure of a different kind occurs, that is where the wave emerges from the plasma into free space or enters the plasma from free space. This effect must now be examined.

The equation for finding coupling points is $\Delta = 0$, (6.38), (6.39), where Δ is the discriminant of the Booker quartic equation. It was mentioned, § 6.6 (Pitteway, 1959), that Δ has a factor X^4 so that there are four coupling points where $X = 0$. From the symmetry of the equations it is clear that two of these must apply for upgoing waves whose q values both tend to C when $X = 0$, and the other two for downgoing waves whose q values tend to $-C$. Thus in both cases there is a coalescence where $X \to 0$. For this reason the subject of limiting polarisation comes within the present chapter. The criteria in §§ 17.2, 17.4 show that it is a coalescence C2.

For many purposes it is useful to assume that in the lowest part of the ionosphere X is some exponential type of function such as $e^{-\alpha z}$, or the Chapman function

(1.9), (1.10). Then $X \to 0$ when $z \to -\infty$. The actual coalescence is therefore at infinity. There is a region where X is extremely small, that extends through a large range of height z. This idea of a large region where $q_1 - q_2$ is small, with the actual coalescence at infinity is very different from the coalescence at non-infinite $z_p = z_q$ discussed in earlier sections. The subject of limiting polarisation therefore needs a completely separate treatment.

The two coalescent coupling points cannot be made to separate by changing any of the parameters in the equations. We are therefore concerned only with exact coalescence. Then the matrix **S**, § 6.10, is non-singular and the transformation (6.53) can be used, as explained in § 17.4. There is no need for the matrix **V** nor for the transformation (17.17).

The subject is discussed here first for vertically incident waves. The extension for oblique incidence is mentioned at the end of the following section.

The polarisations ρ of the two characteristic waves are given by (4.21). When $X = 0$ this becomes

$$\rho = \frac{\frac{1}{2}\mathrm{i}Y\sin^2\Theta \pm \mathrm{i}\{\frac{1}{4}Y^2\sin^4\Theta + (1-\mathrm{i}Z)^2\cos^2\Theta\}^{\frac{1}{2}}}{(1-\mathrm{i}Z)\cos\Theta} \tag{17.59}$$

so that even when there are no electrons ρ has a value that depends on Z. In the free space below the ionosphere it is possible, even when electrons are absent, to speak of the electron collision frequency ν and therefore of Z. These would be roughly proportional to the pressure and would therefore change markedly within the height range of about 60 or 70 km below the ionosphere. Thus if a pure ordinary wave could travel downwards it would show a marked change of polarisation (17.59) in this range. But clearly the actual polarisation cannot change because the medium behaves just like free space. Here the assumption that a characteristic wave retains its correct polarisation must fail. It is shown in the following section that some of the other characteristic wave is generated by mode conversion, and the amount is just enough to keep the polarisation constant. This mode conversion is only effective where the electron concentration is small. Within the ionosphere where the electrons are numerous, the ordinary and extraordinary waves retain their correct polarisation (4.21), and are independently propagated, except in regions of strong coupling near coupling points, as discussed in ch. 16. This independence of the two characteristic waves was mentioned in § 16.11 and shown to be expected where the coupling parameter ψ in Försterling's equations (16.90), (16.91) can be neglected.

The following questions must therefore be examined:

(a) What conditions are necessary for a characteristic wave to retain its polarisation as given by (4.21)?

(b) When one characteristic wave travels down through the ionosphere, what is its limiting polarisation when it reaches the ground?

These problems have been discussed by Booker (1936, § 6), Eckersley and Millington (1939), Budden (1952), Hayes (1971).

This discussion is given in terms of an ordinary wave. An exactly similar argument would apply for an extraordinary wave. Consider therefore an ordinary wave travelling down through the lower ionosphere into the free space below. Here the coupling parameter ψ is small, and in a small height range the amount of extraordinary wave generated by mode conversion is also small. But the two refractive indices n_{O}, n_{E} are both unity, so that the ordinary wave and the small extraordinary wave travel with the same phase velocity. Thus as more of the extraordinary wave is generated, it adds to the amplitude of that already present, because the small contributions arriving from higher levels all have the same phase. In other words, although the mode conversion is small, it is cumulative, and in the following section it is shown that this keeps the polarisation constant. At higher levels where the electrons make $n_{\mathrm{O}} \neq n_{\mathrm{E}}$, the ordinary and extraordinary waves have different phase velocities, so that the small contributions to the extraordinary wave are not phase coherent, and interfere destructively. This process was mentioned in § 16.9. There is then no cumulative generation of the extraordinary wave and the polarisation of the resultant wave remains very close to that of an ordinary wave. It will be shown that the transition occurs near the level where $X \approx 0.0001 - 0.01$ in practical cases and it is here that the limiting polarisation is determined. This is called the 'limiting region'. It is quite different from the 'coupling region' discussed earlier. In the coupling region $X \approx 1$ and $|\psi|$ is large. In the limiting region X and $|\psi|$ are both very small.

The problem is discussed here for waves emerging into free space. Let $\rho_{\mathrm{O}}(\text{out})$ be the limiting polarisation for a vertically emerging ordinary wave. A similar problem arises for waves entering the plasma. It is then required to find what polarisation $\rho_{\mathrm{O}}(\text{in})$ a wave must have such that, when well within the plasma, it consists of an ordinary wave only. It was shown at the end of § 4.4 that if the same axis system is used the polarisations ρ_{O} are the same for upgoing and downgoing ordinary waves, and similarly the two polarisations ρ_{E} for the extraordinary waves are the same. Thence it follows that $\rho_{\mathrm{O}}(\text{in}) = \rho_{\mathrm{O}}(\text{out})$.

When the waves are oblique a similar conclusion holds. Let the wave normal of the emerging wave have direction cosines proportional to S_1, S_2, q_{a} where, in free space, $q_{\mathrm{a}} = -C$. Consider now a wave that enters the plasma and let its wave normal have direction cosines proportional to $-S_1$, $-S_2$, q_{b} so that, in free space, $q_{\mathrm{b}} = C$, and the wave normals of the two waves are antiparallel. The quartic (6.15), (6.16) shows that there is a solution with $q_{\mathrm{b}} = -q_{\mathrm{a}}$ at all levels, and a slight extension of

(7.136), together with the last equations in (7.77), (7.78) shows that, for this solution, the ratios of the components of $\boldsymbol{E}, \mathscr{H}$ for the emergent wave are the same as those of $\boldsymbol{E}, -\mathscr{H}$ for the entering wave. It then follows that the entering and emerging waves have the same polarisations in free space. Note that this conclusion applies only when the wave normals of the emerging and entering waves are antiparallel. It would not apply if the wave normal of the entering wave had direction cosines $+S_1$, $+S_2$, C.

17.11. Limiting polarisation 2. Theory

For vertically downgoing waves in the lower ionosphere the first order coupled equations (16.15) are used, where f_2, f_4 refer to the downgoing ordinary and extraordinary waves respectively. It might, alternatively, be possible to use the second pair of the equations (16.51), but the form (16.15) is simpler. These equations are first to be studied in free space, so that $n_\mathrm{O} = n_\mathrm{E} = 1$. Then the coefficients of $\mathsf{f}_1, \mathsf{f}_3$ in the second and fourth equations (16.15) are zero, which shows that the upgoing and downgoing waves are propagated independently. A prime $'$ denotes $k^{-1}\mathrm{d}/\mathrm{d}z = \mathrm{d}/\mathrm{d}s$ where $s = kz$. Then the second and fourth equations are

$$\mathsf{f}_2' - \mathrm{i}\mathsf{f}_2 = -\mathrm{i}\psi\mathsf{f}_4, \quad \mathsf{f}_4' - \mathrm{i}\mathsf{f}_4 = \mathrm{i}\psi\mathsf{f}_2. \tag{17.60}$$

Two independent exact solutions of these are

$$\mathsf{f}_2 = \mathsf{f}_{2\mathrm{a}} = \mathrm{e}^{\mathrm{i}s}\exp\left(\int_p^s \psi\,\mathrm{d}s\right), \quad \mathsf{f}_4 = \mathrm{i}\mathsf{f}_2, \tag{17.61}$$

and

$$\mathsf{f}_2 = \mathsf{f}_{2\mathrm{b}} = \mathrm{e}^{\mathrm{i}s}\exp\left(-\int_p^s \psi\,\mathrm{d}s\right), \quad \mathsf{f}_4 = -\mathrm{i}\mathsf{f}_2 \tag{17.62}$$

where the lower limit p is some fixed point in the complex s plane. From (16.93)

$$\exp\int_p^s \psi\,\mathrm{d}s = K\left(\frac{\rho_\mathrm{O} - 1}{\rho_\mathrm{O} + 1}\right)^{\frac{1}{2}} \tag{17.63}$$

where K is a constant. If now (17.61) with (17.63) is used with $\mathsf{f}_1 = \mathsf{f}_3 = 0$, in (16.16) to find the field components, it gives

$$E_x = E_y = -\mathrm{i}K2^{-\frac{1}{2}}\mathrm{e}^{\mathrm{i}kz}. \tag{17.64}$$

Similarly (17.62) gives

$$E_x = -E_y = \mathrm{i}K^{-1}2^{-\frac{1}{2}}\mathrm{e}^{\mathrm{i}kz}. \tag{17.65}$$

The factor $\mathrm{e}^{\mathrm{i}kz}$ shows that these waves are both travelling downwards with phase velocity c, and that their polarisations are linear and independent of the height z. Thus they behave just as though they are in free space. Any linear combination of (17.61), (17.62), that is of (17.64), (17.65), is a solution of (17.60) so the resulting wave may have any chosen polarisation $\rho = E_y/E_x$. Hence let

$$\mathsf{f}_2 = a\mathsf{f}_{2a} + b\mathsf{f}_{2b}, \quad \mathsf{f}_4 = \mathrm{i}(a\mathsf{f}_{2a} - b\mathsf{f}_{2b}). \tag{17.66}$$

Then

$$\rho = (a+b)/(a-b), \quad a/b = (\rho+1)/(\rho-1). \tag{17.67}$$

This completes the proof that, in the free space below the ionosphere, the mode conversion is just enough to keep the polarisation constant.

Now let the range of z be extended up to a level just above the limiting region where X is small but not zero. Here n_{O}, n_{E} differ from unity by small quantities of order X; compare (4.112). In the second equation (16.15) the coefficient of f_1 has a factor n'_{O} that is small because the medium is slowly varying. The coefficient of f_3 has factors $n_{\mathrm{O}} - n_{\mathrm{E}}$ and ψ that are both small. If any upgoing wave, represented by f_1 or f_3, is present, it can now generate a small amount of the downgoing wave f_2 by mode conversion, but the contributions to this from various levels are not phase coherent and therefore they interfere destructively. The same applies to the fourth equation (16.15). For this reason it is assumed that we can continue to ignore the effects of the upgoing waves and take $\mathsf{f}_1 = \mathsf{f}_3 = 0$. The coefficients of $\mathsf{f}_4, \mathsf{f}_2$ on the right of the second and fourth equations (16.15) respectively contain a factor $\frac{1}{2}(n_{\mathrm{O}} + n_{\mathrm{E}})(n_{\mathrm{O}}n_{\mathrm{E}})^{-\frac{1}{2}} = 1 + \mathrm{O}(X^2)$. Since this is multiplied by the small quantity ψ it is assumed that the small terms of order X^2 can be neglected. Then the equations become

$$\mathsf{f}'_2 - \mathrm{i}n_{\mathrm{O}}\mathsf{f}_2 = -\mathrm{i}\psi\mathsf{f}_4, \quad \mathsf{f}'_4 - \mathrm{i}n_{\mathrm{E}}\mathsf{f}_4 = \mathrm{i}\psi\mathsf{f}_2. \tag{17.68}$$

Now let

$$u = \mathsf{f}_4/\mathsf{f}_2 \quad \text{so that} \quad u' = \frac{\mathsf{f}'_4}{\mathsf{f}_2} - u^2\frac{\mathsf{f}'_2}{\mathsf{f}_4}. \tag{17.69}$$

Then u satisfies the differential equation

$$u' + \mathrm{i}u(n_{\mathrm{O}} - n_{\mathrm{E}}) - \mathrm{i}\psi(1+u^2) = 0 \tag{17.70}$$

which is a non-linear equation of Riccati type. If u can be found at some level $z = z_1$ below the ionosphere, the limiting polarisation ρ_1 can be found from (16.16) with $\mathsf{f}_1 = \mathsf{f}_3 = 0$, thus

$$\rho_1 = (\mathrm{i}\rho_{\mathrm{O}} + u)/(\mathrm{i} + \rho_{\mathrm{O}}u) \tag{17.71}$$

where ρ_{O} is the value of (17.59) for the ordinary wave where $z = z_1$. The wave above the limiting region is a pure ordinary wave so that there $\mathsf{f}_4 = 0$, $u = 0$. Equation (17.70) is suitable for numerical integration on a computer and has been used in this way by Barron (1960). The integration proceeds downwards starting just above the limiting region with $u = 0$. The same equation, with slightly different notation, has been used by Titheridge (1971b) in a study of the effect of a coupling region on Faraday rotation.

Now differentiate the first equation (17.68) once and use the result with (17.68) to eliminate f_4, f_4'. This gives

$$\mathrm{f}_2'' - \{\mathrm{i}(n_\mathrm{O} + n_\mathrm{E}) + \psi'/\psi\}\mathrm{f}_2' + (\mathrm{i}n_\mathrm{O}\psi'/\psi - \mathrm{i}n_\mathrm{O}' - n_\mathrm{O}n_\mathrm{E} - \psi^2)\mathrm{f}_2 = 0. \quad (17.72)$$

To remove the first derivative term take

$$v = \mathrm{f}_2\psi^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2}\mathrm{i}\int(n_\mathrm{O} + n_\mathrm{E})\mathrm{d}s\right\}. \quad (17.73)$$

Then

$$v'' + v\{\tfrac{1}{4}(n_\mathrm{O} - n_\mathrm{E})^2 - \tfrac{1}{2}\mathrm{i}(n_\mathrm{O}' - n_\mathrm{E}') + \tfrac{1}{2}\mathrm{i}(n_\mathrm{O} - n_\mathrm{E})\psi'/\psi - \psi^2 + \tfrac{1}{2}\psi''/2\psi - \tfrac{3}{4}(\psi'/\psi)^2\} = 0. \quad (17.74)$$

If the terms containing ψ are ignored, it can be confirmed that one solution is

$$v = \exp\left\{\tfrac{1}{2}\mathrm{i}\int(n_\mathrm{O} - n_\mathrm{E})\mathrm{d}s\right\}, \quad \mathrm{f}_2 \propto \exp\left(\mathrm{i}\int n_\mathrm{O}\mathrm{d}s\right), \quad \mathrm{f}_4 = 0 \quad (17.75)$$

so that the wave is a pure ordinary wave. Similarly if $n_\mathrm{O} - n_\mathrm{E}$ is ignored the solutions are (17.61), (17.62) and the wave travels as in free space. The limiting region is where the terms containing $n_\mathrm{O} - n_\mathrm{E}$ and those containing ψ are comparable, and it is here that the limiting polarisation is determined. There seems to be no simple way of solving this problem to give a concise expression for the limiting polarisation. To proceed further we shall therefore use a specific numerical example, typical of the lower ionosphere, to indicate the orders of magnitude of the various terms in the coefficient of v in (17.74).

Consider therefore a frequency $f \approx 2$ MHz so that $Y \approx \frac{1}{2}$ and take $\nu = 5 \times 10^5\,\mathrm{s}^{-1}$, so that $Z \approx 0.04$. Let $\Theta \approx 30°$ as for temperate latitudes so that $Z_\mathrm{t} \approx 0.07$ from (4.27). The value of ψ in the lower ionosphere is given by (16.98), where it is assumed that X is negligibly small and $X' \ll Z'$; compare figs. 16.5–16.7 at low levels. Then (16.98) gives $\psi \approx 0.04\,Z'$. The coefficient here is practically independent of Z and therefore of height z. Suppose that Z is proportional to $\exp(-z/H)$ where $H \approx 10$ km; compare § 1.6. Then $\psi \approx -0.04\,Z/kH \approx -4 \times 10^{-6}$, and ψ^2 is negligible compared with $\frac{1}{2}\psi''/\psi - \frac{3}{4}(\psi'/\psi)^2 = -\frac{1}{4}(kH)^{-2} \approx 1.4 \times 10^{-6}$. This would still be true for any other H because if H is changed, ψ^2, ψ''/ψ and $(\psi'/\psi)^2$ are all changed by the same factor.

When X is small the refractive indices are given to order X by (4.112), whence

$$n_\mathrm{O} - n_\mathrm{E} \approx \frac{XY(\frac{1}{4}Y^2\sin^4\Theta + U^2\cos^2\Theta)^{\frac{1}{2}}}{U(U^2 - Y^2)} \approx 0.6X. \quad (17.76)$$

Thus the orders of magnitude of the various terms in the coefficient of v in (17.74) are

$\frac{1}{4}(n_\mathrm{O} - n_\mathrm{E})^2$	$\frac{1}{2}\mathrm{i}(n_\mathrm{O}' - n_\mathrm{E}')$	$\frac{1}{2}\mathrm{i}(n_\mathrm{O} - n_\mathrm{E})\psi'/\psi$	$\frac{1}{2}\frac{\psi''}{\psi} - \frac{3}{4}\left(\frac{\psi'}{\psi}\right)^2$	ψ^2
$0.1X^2$	$0.3X'$	$7 \times 10^{-4}X$	1.4×10^{-6}	1.6×10^{-11}

If X in the lower ionosphere is proportional to $\exp(z/H_X)$, the second term is $0.3X/kH_X$ and is less than the third term if $H_X \gtrsim 10$ km. Thus when X varies very slowly with height, the second term may be ignored. When the third term is comparable with the fourth, it exceeds the first term, and here $X \approx 0.002$. This gives an estimate of the level z where the limiting polarisation would be determined. If X follows the Chapman law (1.9), (1.10) with reasonable values of N_m and H, it can be shown that in the lowest part of the layer the second term is larger than the third. The second and fourth terms are now approximately equal where $X \approx 0.0002$.

It is thus clear that the location of the limiting region depends very much on how X varies with height in the lowest part of the ionosphere. But the above figures suggest that it is likely to be where X is between about 0.0001 and 0.01. This determines a range for the height z_l, whence the range of the value of Z can be found.

To find the limiting polarisation ρ_l the approximate values of X and Z must be inserted in (4.21). But X is so small that it can be neglected, and (17.59) can then be used. In this formula ρ is also insensitive to the exact value of Z since in most cases of practical interest $Z \ll 1$. Thus it is not necessary to know z_l and Z with great accuracy. In practice it is good enough if z_l can be estimated to within about $\pm$ 10 km. For the same reason it is not possible to use measurements of the limiting polarisation to give information about Z where $z = z_l$. The observed polarisations often vary with time (Landmark, 1955; Morgan and Johnson, 1955) probably because the ionosphere moves and is not exactly horizontally stratified.

For the study of limiting polarisation when the wave normal is oblique, the theory is more involved and the coupled equations are too complicated to be worth using. It is better to use numerical integration of the basic equations (7.80). This was done by Hayes (1971), who studied how the polarisation varied with height z for a wave that is initially a pure ordinary wave and comes obliquely down through the lower ionosphere. The results depend on the angle of incidence, the azimuth of the plane of incidence and on the function $N(z)$ in the lowest part of the ionosphere. A case was included where this $N(z)$ has a maximum and a minimum, as proposed by Deeks (1966a, b). The general conclusion was the same as above for vertical incidence, namely that the limiting polarisation is found from (17.59) for some z_l in a region of the lowest part of the ionosphere where X is of order 0.0001 to 0.01. Only a rough estimate is needed for the value of Z in this region.