The transition from locked to leaky modes in tropospheric radio propagation

1973 J. Phys. A: Math. Nucl. Gen. 6 177

(http://iopscience.iop.org/0301-0015/6/2/009)

J. Phys. A: Math., Nucl. Gen., Vol. 6, February 1973. Printed in Great Britain. © 1973

The transition from locked to leaky modes in tropospheric radio propagation

M G W Hayes

Cavendish Laboratory, University of Cambridge, Cambridge CB2 3RQ, UK

MS received 14 August 1972

Abstract. The propagation characteristics of a waveguide mode in the troposphere near the ground are studied by the simple phase integral method over a range of frequencies in which the mode changes from locked to leaky. The method is applied to a troposphere with a parabolic distribution for the square of the modified refractive index, and the results are compared with those from an exact analytic method. Particular attention is given to the contour of the phase integral in the complex height plane. This contour changes discontinuously at the transition from a locked to a leaky mode. An alternative phase integral formula is described in which there is no abrupt change at the transition. Some results from this formula are presented, and a derivation of it for locked modes is given.

1. Introduction

The phase integral method (Eckersley 1931, 1932a, b, c, Booker and Walkinshaw 1946, Heading 1962) can be used for solving many problems in radio propagation in a horizontally stratified system such as the ionosphere or troposphere. One of its simplest applications is for finding the reflection coefficient R for radio waves normally incident on an isotropic slowly varying ionosphere with a simple electron distribution $N(z)$ having just one maximum. Another application is to the study of waveguide modes in the troposphere near the ground when the atmosphere has a temperature inversion so that locked modes are possible. These two examples share certain features which are studied in this paper.

In the first example, the reflection coefficient R of the ionosphere is given by the approximate phase integral formula

$$R \simeq \mathrm{i} \exp\left(-2\mathrm{i}k \int_0^{z_{01}} \mu \,\mathrm{d}z \right) \tag{1}$$

(Budden 1961) where $\mu(z)$ is the refractive index, and z_{01} is the lower of the two real levels z_{01}, z_{02} where $\mu(z)$ is zero. For a frequency f which is less than the penetration frequency f_{p} of the ionosphere, $|R|$ is unity if the ionosphere is loss free. Above the reflection level there is a region $z_{01} < z < z_{02}$, similar to a 'potential barrier' in wave mechanics, where the wave is evanescent. At still greater heights $z > z_{02}$ a propagated wave, the transmitted wave, is again possible, but if the barrier is thick the transmission coefficient is negligible. For frequencies greater than f_{p} these levels z_{01}, z_{02} are complex conjugates and z_{01} is chosen such that $\mathrm{Im}\,(z_{01})$ is negative. There is no barrier and the reflection coefficient is very small. For a range of frequencies near f_{p} the formula (1) fails in a loss-free ionosphere. If the electron collision frequency ν is allowed for, however, this range of failure is reduced and if ν is sufficiently large there is no serious failure. There is then a continuous transition from strong reflection for $f < f_p$ to weak reflection and strong penetration for $f > f_p$. As f increases from small to large values the point z_{01} moves along a continuous curve in the complex z plane.

In the second example, for a tropospheric duct, a locked mode is possible if the frequency is great enough. For this mode the troposphere is then a perfect reflector. Here only the mode of lowest order, the first order mode, is studied but similar considerations apply to the other modes. The ground is always assumed to be a perfect conductor. Above the reflection level in the troposphere there is a barrier region similar to that in the first example. If the frequency is reduced, the reflection level gets higher and the barrier becomes thinner. If it is thin enough, there is some penetration and the mode becomes slightly leaky. The simplest form of the phase integral formula, (18) below, ignores this leakage. It predicts that the mode is 'well locked' provided that f is greater than a transition frequency f_T. The error is not serious when $f \gg f_T$ but may be intolerable if f is close to f_T. For frequencies less than f_T the mode is leaky, and if f is sufficiently small the phase integral method can again be used with good accuracy. The method uses a point z_0 in the complex z plane where the variable $q(z)$, (5) below, is zero. This z_0 fills the same role as the z_{01} in the first example.

Loss of energy from a mode through leakage is, in some respects, similar to loss of energy from a vertically incident wave through electron collisions in the ionosphere. It might be expected, therefore, that the transitions in these two examples, f decreasing through f_T and f increasing through f_p respectively, are similar. In fact they are different. The purpose of this paper is to study the transition from a locked to a leaky mode as f decreases through f_T. It is shown that the z_0 used in the phase integral method does not lie on a continuous curve as f is varied, but on two different curves, one for $f \gg f_T$ and the other $f \ll f_T$. The jump occurs in the transition range $f \simeq f_T$ where the simple phase integral formula must inevitably fail.

This paper is therefore aimed at understanding the basic wave propagation processes of the transition. It is not concerned with actual tropospheric models. A simple modified refractive index distribution, $\mu(z)$ in (2) below, has been chosen for which the exact solution of the wave equation is known, and this allows the validity of the phase integral formula to be tested. For actual distributions $\mu(z)$, where there is no known exact solution, either some approximation must be used, or a full numerical solution. The phase integral method is a quick and simple approximate method, and it is therefore important to know the extent of its validity.

2. Notation

A cartesian coordinate system, axes x, y, z is used, with the z axis directed vertically upwards, and the x–z plane is defined to be the plane containing the wave normal. The troposphere is assumed to be horizontally stratified so that the modified refractive index μ is a function only of z. The principal symbols used are as follows.

- θ angle between the wave normal and the vertical at a level where the refractive index is unity.
- μ modified refractive index.
- ω angular frequency.
- C $\cos\theta$.

c velocity of electromagnetic waves in free space.
f $\omega/2\pi$.
k ω/c.
n mode number.
S $\sin\theta$.

Other symbols are explained in the text as they are used. The amplitude of the waves depends upon time t through a factor $\exp(i\omega t)$, which is omitted throughout.

3. Statement of the problem

On the average the refractive index of the atmosphere decreases as z increases. Sometimes, for example when there is a temperature inversion (Appleton 1946), the rate of decrease can be so great that there is a duct in which well locked waveguide modes can propagate if the frequency is great enough. In the theory of this type of propagation, the earth's curvature can be allowed for by using a modified expression for the refractive index function (Booker and Walkinshaw 1946, Pekeris 1946) which then permits the assumption that the earth is flat. It is required to study the simplest problem of this kind. It is therefore assumed that the earth's surface is flat and perfectly conducting, and that the modified refractive index $\mu(z)$ is given by

$$(\mu(z))^2 = 1+\beta\{(z-z_{\rm m})^2-z_{\rm m}^2\} \tag{2}$$

where β is a real constant and $z_{\rm m}$ is a real height. This has a minimum value at $z = z_{\rm m}$ and there can be a duct at heights below this. Solutions are sought in which the electric intensity of the wave is horizontal (parallel to the y axis) and given by

$$E(z)\exp(-ikSx). \tag{3}$$

Then E must satisfy

$$\frac{{\rm d}^2E}{{\rm d}z^2}+k^2q^2E = 0 \tag{4}$$

where

$$q^2 = \mu^2-S^2$$
$$= C^2+\beta\{(z-z_{\rm m})^2-z_{\rm m}^2\} \tag{5}$$

and

$$C^2 = 1-S^2. \tag{6}$$

Further, E must satisfy the boundary condition at the perfectly conducting ground, namely

$$E(0) = 0. \tag{7}$$

At great heights the solution must represent an outgoing wave, that is one in which the z component of the Poynting vector is positive. Solutions for E which satisfy these boundary conditions are only possible for discrete eigenvalues of S, each of which is associated with one waveguide mode. Here we study only the least attenuated mode, that is the one for which $|{\rm Im}(S)|$ is least.

Equation (4) can be solved exactly in terms of the parabolic cylinder functions (Whittaker and Watson 1935). It can be shown (Budden 1961) that the required solution is the Weber function

$$E = D_\nu(\xi) \tag{8}$$

where

$$\xi = (4k^2\beta)^{1/4}(z - z_m)\exp(\tfrac{1}{4}i\pi) \tag{9}$$

and

$$\nu + \tfrac{1}{2} = k^2(C^2 - \beta z_m^2)(4k^2\beta)^{-1/2}\exp(-\tfrac{1}{2}i\pi) \tag{10}$$

and where the real positive values of the fractional powers are used. When $z = 0$, $\xi = \xi_0$ such that

$$\xi_0 = -z_m(4k^2\beta)^{1/4}\exp(\tfrac{1}{4}i\pi). \tag{11}$$

The values of β, k and z_m used later in this paper are such that $|\xi_0|$ is sufficiently large for only the first two terms in the asymptotic series approximation (Whittaker and Watson 1935 § 16.52) for (8) to be used. The boundary condition (7) then becomes

$$\left(1 - \frac{\nu^2+\nu+1}{\xi_0^2}\right)\xi_0^{2\nu+1}(2\pi)^{-1/2}(-\nu-1)!\exp(\nu\pi i - \tfrac{1}{4}\xi_0^2) = 1. \tag{12}$$

The fractional complex power $\xi_0^{2\nu+1}$ is defined as follows. Equation (11) is rewritten as

$$\xi_0 = z_m(4k^2\beta)^{1/4}\exp(\tfrac{5}{4}i\pi) \tag{13}$$

so that

$$\ln\xi_0 = \ln z_m + \tfrac{1}{4}\ln(4k^2\beta) + \tfrac{5}{4}i\pi \tag{14}$$

and

$$\xi_0^{2\nu+1} = \exp\{(2\nu+1)\ln\xi_0\}. \tag{15}$$

Now (14) and (15) define $\xi_0^{2\nu+1}$ unambiguously.

Equation (12), with (10) and (11), is an equation for finding the eigenvalue C. It was checked in a few cases by comparison with a numerical stepwise integration of (4) as used, for example, by Hartree *et al* (1946), and its accuracy was found to be very good for the frequencies studied in this paper. When C is known, the value z_0 of z which makes $q = 0$ is given from (5) by

$$z_0 = z_m \pm \left(z_m^2 - \frac{C^2}{\beta}\right)^{1/2}. \tag{16}$$

The two values of z_0 obtained by this method will be denoted by $z_{01}^{(w)}$ and $z_{02}^{(w)}$ where w indicates that the Weber function solution (8) has been used, and

$$\mathrm{Re}(z_{01}^{(w)}) < \mathrm{Re}(z_{02}^{(w)}). \tag{17}$$

The same problem can be solved approximately by the phase integral method. By analogy with (1), the reflection coefficient of the troposphere, as observed at the ground, for obliquely incident waves is given by

$$R \simeq i\exp\left(-2ik\int_0^{z_0} q\,dz\right) \tag{18}$$

where $q(z)$ is given by (5) and is chosen so that $\mathrm{Re}(q) > 0$. The boundary condition (7) requires that $R = -1$. These formulae were used by Booker and Walkinshaw (1946) who showed (Booker and Walkinshaw 1946 equation (37)) that the resulting equation for finding the eigenvalue C is

$$\int_0^{z_0} q\,\mathrm{d}z = \frac{1}{2}\left(n-\frac{1}{4}\right)\frac{c}{f} \tag{19}$$

together with (16), where n is a positive integer. It is important to choose the correct alternative of the two values (16). These values will be denoted by $z_{01}^{(\mathrm{p})}$ and $z_{02}^{(\mathrm{p})}$ where p indicates that the approximate phase integral method has been used, and

$$\mathrm{Re}(z_{01}^{(\mathrm{p})}) < \mathrm{Re}(z_{02}^{(\mathrm{p})}). \tag{20}$$

The discussion of the solutions of (19) and of the criterion for choosing z_0 in (16) is the main purpose of the present paper.

An alternative and more accurate form of (18) has been suggested for frequencies near f_T (S Rotheram private communication), namely

$$R \simeq \mathrm{i}\exp\left(-2ik\int_0^{z_{01}} q\,\mathrm{d}z\right)\left\{1+\exp\left(+2ik\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z\right)\right\}^{-1}. \tag{21}$$

A derivation of this formula is given for slightly leaky modes in § 7. If (21) is used with (7) and (16), it gives yet another pair of values of z_0, which will be written $z_{01}^{(\mathrm{A})}$ and $z_{02}^{(\mathrm{A})}$ where A indicates that the alternative formula (21) has been used.

If $f \gg f_\mathrm{T}$, that is for well locked modes, the last exponential term in (21) is very small. If it is neglected, (21) and (18) are the same, provided that z_{01} is used for the z_0 in (18).

Now (21) can also be written thus:

$$R \simeq \mathrm{i}\exp\left(-2ik\int_0^{z_{02}} q\,\mathrm{d}z\right)\left\{1+\exp\left(-2ik\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z\right)\right\}^{-1}. \tag{22}$$

For $f \simeq f_\mathrm{T}$ the modulus of the last exponential term in (22) is approximately unity. As f decreases, this term decreases. If it is neglected then (22) is the same as (18), provided that z_{02} is used for the z_0 in (18). The exponential term does not, however, tend to zero as f goes to zero. For the examples considered it has a minimum value of about 0·1 at $f \simeq 9$ MHz, and it then increases as f decreases further. Nevertheless, for frequencies in the transition region (22) tends to (18) with z_{02} used for z_0, as f is reduced. There is an appreciable range of frequencies $f \ll f_\mathrm{T}$ for which (22) is approximately the same as (18).

This change, from the use of z_{01} in the simple phase integral formula (19) for well locked modes, to the use of z_{02} for leaky modes, is shown to be necessary, in § 5, and is one of the main results of this paper.

4. Locked modes

Solutions will now be studied for modes of the smallest order $n = 1$, so that (19) becomes

$$\int_0^{z_0} q\,\mathrm{d}z = \frac{3}{8}\frac{c}{f}. \tag{23}$$

In any mode the electromagnetic field near the ground may be considered as composed of two crossing waves, one travelling obliquely upwards and the other obliquely downwards. One is converted into the other by reflection at the ground and again at the level $z = z_{01}^{(p)}$ where $q = 0$. Equation (23) is the condition that after two reflections the original wave is again produced, as required for a self-consistent mode. For locked modes, the reflection level $z_{01}^{(p)}$ is real so that, from (16) C^2 is real and less than βz_m^2. It is clear that the reflection level must be the smaller, $z_{01}^{(p)}$, of the two values (16). This is confirmed when the more accurate condition (12) is used. When the square root in (16) is small the two values of z_0 are close together. Between them is the barrier mentioned in § 1, and since this barrier is thin, some energy can leak through it. This possibility, however, is ignored when (23) is used and this equation predicts locked modes provided that

$$\frac{C^2}{\beta} \leqslant z_m^2. \tag{24}$$

For the equality sign in (24), we have from (16) and (23)

$$z_{01}^{(p)} = z_{02}^{(p)} = z_m, \qquad f = f_T = \tfrac{3}{4} c \beta^{-1/2} z_m^{-2} \tag{25}$$

where f_T is the transition frequency mentioned in § 1. The positions of the points (20) in the complex z plane, as given by solving (23) and (16) for various frequencies $f > f_T$, are shown in figure 1. Also shown are the more accurate values $z_{01}^{(A)}$, $z_{02}^{(A)}$ given by using

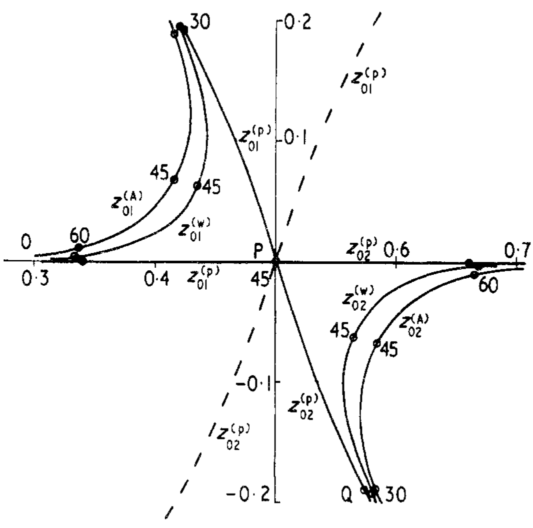

Figure 1. The complex z plane showing the loci of the points z_0 where $q = 0$. The curves $z_{01}^{(w)}$, $z_{02}^{(w)}$ were obtained from the Weber function solution (8), $z_{01}^{(p)}$, $z_{02}^{(p)}$ were obtained from the approximation (19) of the phase integral method, and $z_{01}^{(A)}$, $z_{02}^{(A)}$ were obtained from the alternative formula (21). The numbers by the points show the frequency in MHz. In this example $z_m = 0{\cdot}5$ km, $\beta = 4 \times 10^{-4}$ km^{-2}, the transition frequency f_T is 45·0 MHz and the troposphere is loss free. The dotted line shows 'negative order mode' solutions, obtained when $z_{01}^{(p)}$ is used for z_0 in (19) when $f < f_T$.

(21) instead of (18), and the values (17) found from (12) and (16). It is seen that the agreement of all three methods is good when f is large, but as f decreases the positive imaginary part of both $z_{01}^{(A)}$ and $z_{01}^{(w)}$ increases. The associated values of C and S are complex, so that, from (3), the mode is attenuated as it travels in the horizontal x direction. The simple phase integral method, however, gives a real value $z_{01}^{(p)}$, provided $f > f_T$, and there would then be no attenuation.

Figure 1 shows that the alternative form (21) of the phase integral formula gives results which are more accurate than for the simple form (18), but the accuracy is still poor when f is close to f_T.

Figure 2 shows the positions of the eigenvalues C, in the complex C plane, for various frequencies f, obtained from all three methods. It is seen that the agreement of all the methods is good when f is large but as f decreases the positive imaginary part of C, found from the alternative phase integral formula (21) and from (12), increases steadily. The simple phase integral method, however, gives a real value for C for all frequencies $f > f_T$.

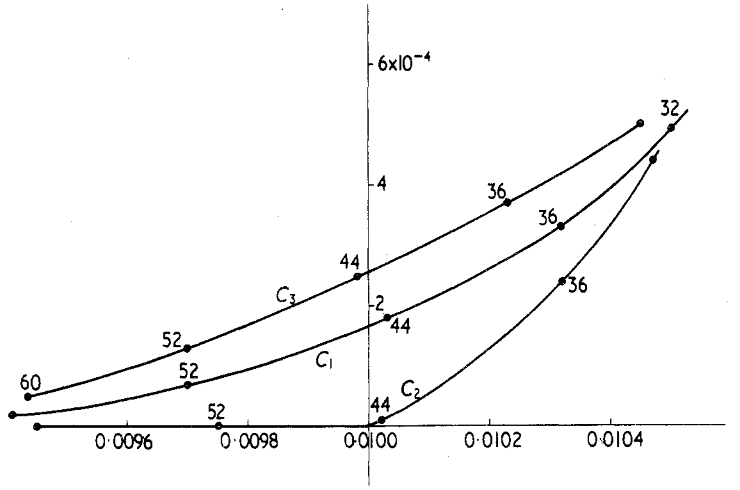

Figure 2. The complex C plane showing the loci of the eigenvalues of C satisfying the boundary conditions for the least attenuated mode. Curve C_1 was obtained from the Weber function solution (8), curve C_2 was from the approximation (19) of the phase integral method and curve C_3 was obtained from the alternative formula (21). The numbers by the points show the frequency in MHz. The values of z_m, β and f_T are the same as in figure 1 and the troposphere is again loss free.

5. Leaky modes

Suppose now that the frequency is decreased so that $f < f_T$. The resulting values $z_{01}^{(w)}$ and $z_{02}^{(w)}$ are shown in figure 1, and it is seen that they lie on the same continuous curves as for $f > f_T$. The same applies to the pair $z_{01}^{(A)}$, $z_{02}^{(A)}$ which are also shown. As f decreases, the imaginary part of $z_{01}^{(w)}$, and of $z_{01}^{(A)}$, continues to increase, and the associated value of S has an increasing negative imaginary part which shows that the attenuation of the mode through leakage increases as f decreases. The transition from locked to leaky modes is therefore gradual and continuous.

Now consider the solution when $f < f_T$ of equations (16) and (23) which use the simple phase integral formula (18). For z_0 in (23), either $z_{01}^{(p)}$ or $z_{02}^{(p)}$ may be chosen. It is found that if $z_{01}^{(p)}$, with the smaller real part, is consistently chosen, no value of S satisfying (23) can be found. If that z_0 with the positive imaginary part is consistently used, then an iterative solution of (16) and (23) gives the z_0 which also has the greater real part. Furthermore, if this z_0 is used, the resulting value of S has a positive imaginary part so that the amplitude of the field in the mode would increase as it travelled. Thus the mode would be of 'negative order' and could not be excited by any source (Wait 1957). It is therefore necessary to choose that z_0 with negative imaginary part. The resulting value of S then makes this z_0 also have the larger real part so that it must be labelled $z_{02}^{(p)}$. The corresponding values of $z_{01}^{(p)}$ and $z_{02}^{(p)}$ are shown in figure 1. They lie on the continuous lines through z_m. When f is small enough, the agreement between the points $z_{02}^{(p)}$, $z_{02}^{(A)}$ and $z_{02}^{(w)}$ is good (near Q in figure 1).

Figure 2 shows the eigenvalues C in the complex C plane for frequencies $f < f_T$ obtained from all three methods. The curves from the alternative phase integral formula (21) and from (12) continue smoothly from the corresponding curves for frequencies $f > f_T$. The curves from the simple phase integral formula, however, show a sharp discontinuity of gradient at $f = f_T$.

According to the simple phase integral method, therefore, the transition from locked to leaky modes is abrupt and occurs when f decreases and passes through f_T. The point z_0 used in (23) moves along a continuous line formed by the two lines OP and PQ in figure 1. There is no attenuation when $f > f_T$. For $f < f_T$ the attenuation increases as f decreases. For frequencies near f_T, however, the accuracy is very poor. For frequencies less than f_T, that is for leaky modes, the value of z_0 in (16) with the negative imaginary part must be used in (23). This is also the value with the greater real part.

For the alternative phase integral formula (21) or (22) the transition from locked to leaky modes is continuous. It is convenient, however, to use the form in which the exponential term in the correction factor in curly brackets has modulus less than unity. This means using (21) for locked modes, and (22) for leaky modes. If the correction factor is neglected there is again an abrupt change from the use of z_{01} for locked modes to z_{02} for leaky modes. The latter point has the greater real part. In the ionospheric problem of § 1, by contrast, the value of z_0 used in the phase integral method is never the one with the larger real part.

If losses due to electron collisions in the ionospheric plasma are allowed for, and if they are sufficiently great, the accuracy of the phase integral method is good for all frequencies including those near f_p. In the tropospheric duct problem, the medium is loss free. Leakage does not have the same effect as losses in the medium and there is always a failure of the phase integral method when f is near f_T. It is interesting to ask if losses in the tropospheric medium would reduce the range of f in which the phase integral method fails. This question is now examined.

6. Effect of losses in the medium

It is now supposed that the modified refractive index μ in the troposphere has a negative imaginary part, so that (1) is replaced by

$$(\mu(z))^2 = [1+\beta\{(z-z_m)^2 - z_m^2\}](1-i\eta) \tag{26}$$

where η is a real positive constant. Since βz_m^2 is very small, of the order of 10^{-4} in the example considered here, $\mathrm{Im}(\mu)$ is negative when z is real. This expression thus takes account of absorption of electromagnetic energy in the medium. Whilst this would rarely be important in the actual troposphere, it is possible for ducted propagation to occur in the ionosphere where there is a trough in the electron distribution $N(z)$ (Chang 1971a, b) and the losses in this medium could be important. Here, however, it is supposed that the lower reflector is still a perfectly conducting plane, so that the results can be compared with the loss free case studied in earlier sections. It is further assumed that the medium is nondispersive so that (26) is independent of frequency. This simplification would not apply to an ionospheric duct.

Figure 3 shows the loci of the points $z_{01}^{(w)}$, $z_{02}^{(w)}$ and $z_{01}^{(A)}$, $z_{02}^{(A)}$ for two values of the loss parameter η in (26). They do not change in form as η is increased but the two branches merely separate from each other. Also shown are the loci of $z_{01}^{(p)}$, $z_{02}^{(p)}$ from (23) for the same values of η. The point $z_{01}^{(p)}$ was used for z_0 in (23) when the exponential term in the correction factor in curly brackets in (21) had modulus less than unity, and $z_{02}^{(p)}$ was used for z_0 when this term in (21) had modulus greater than unity.

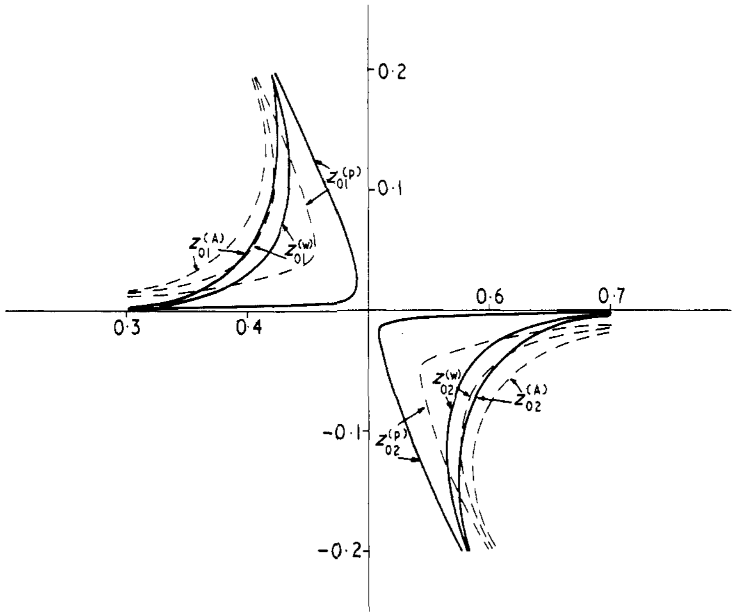

Figure 3. The complex z plane showing the loci of the points z_0 where $q = 0$, for two different values of η. The continuous curves are for $\eta = 0{\cdot}01$; the dotted curves are for $\eta = 0{\cdot}1$. The remaining labelling is the same as in figure 1 and the values of z_m, β and f_T are also the same as in figure 1, with which this figure should be compared.

The positions of the branch points are much less sensitive to changes in η than are the eigenvalues C. Furthermore the frequency range around f_T, where the simple phase integral approximation is poor, is reduced only very slightly. It is concluded that the effect of losses in the medium does not appreciably change the range or extent of failure of the phase integral method. In order to make the approximation good for all frequencies, the losses would have to be so large as to make the eigenvalues C wildly unrealistic. Again this contrasts with the ionospheric example, where moderate and realistic losses make the phase integral approximation good for all frequencies around f_p.

7. Discussion

This section attempts to explain why, when using the phase integral method, it is necessary to change from using $z_{01}^{(p)}$ (nearer to the ground) for well locked modes when $f \gg f_T$, to using $z_{02}^{(p)}$ (further from the ground) for leaky modes when $f \ll f_T$.

Consider first the case of a well locked mode travelling in the direction of x increasing. The electric field E of the least attenuated mode, in the plane $x = 0$ is now to be studied. The complex z plane for this case is shown in figure 4, and the points z_{01}, z_{02} are branch points of $q(z)$ from (5). The branch cuts are the lines $\mathrm{Re}(q) = 0$ and they do not cross the real z axis. In the following formulae the value of q with positive real part is used.

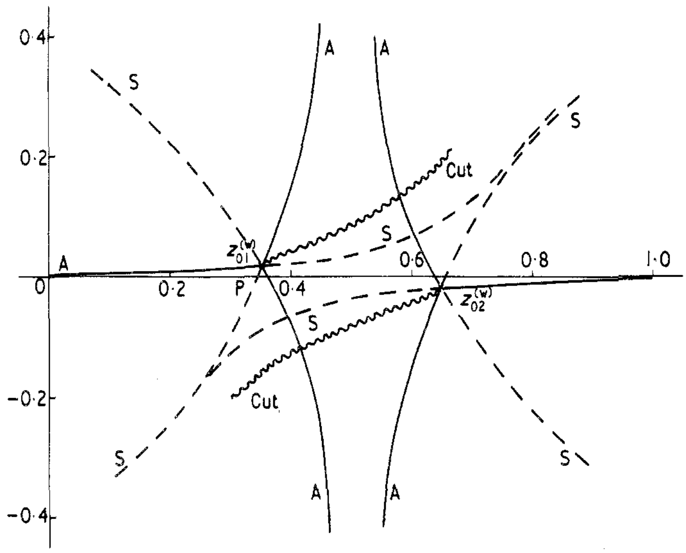

Figure 4. The complex z plane showing the branch points $z_{01}^{(w)}$, $z_{02}^{(w)}$ for a well locked mode obtained from the Weber function solution (8). The branch cuts are where $\mathrm{Re}(q) = 0$. The lines marked S and A are Stokes and anti-Stokes lines respectively.

In the duct there is an obliquely upgoing wave, labelled $D(0)$ in figure 5, whose electric field at the ground, $x = 0$, $z = 0$, is E_0. It is reflected near the level z_{01} to produce an obliquely downgoing wave $D(1)$ whose electric field at the ground is E_1. The reflection coefficient R_1 which would be measured by an observer at $z = 0$ is given by the phase

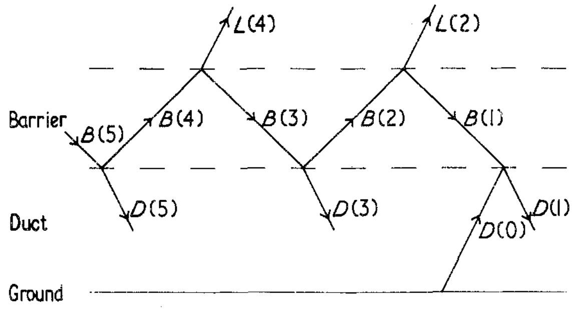

Figure 5. A schematic diagram for a mode which is only slightly leaky. It shows the upgoing and downgoing waves in the barrier region, and the resulting downgoing waves in the duct and leakage waves above the barrier.

integral formula (18) so that

$$\frac{E_1}{E_0} = R_1 \simeq \mathrm{i}\exp\left(-2\mathrm{i}k\int_0^{z_{01}} q\,\mathrm{d}z\right). \tag{27}$$

This result is reliable only if the reflection near z_{01} is not influenced by the other branch point z_{02}. This is achieved when the region between z_{01} and z_{02} is great enough. This region is the barrier which prevents appreciable leakage. According to (18) or (27) z_{01} is then purely real and the wave within the barrier contains a factor $\exp(+\mathrm{i}k\int^z q\,\mathrm{d}z)$ where q is positive imaginary. There is only one wave there, and this has its average Poynting vector horizontal and its amplitude decreases rapidly as z increases. It was shown in § 4, however, that a more accurate treatment, using either the alternative formula (21) or the analytic solution (8), shows that z_{01} has a positive imaginary part because the upper branch point for z_{02} always has a small influence. If we use either $z_{01}^{(\mathrm{A})}$ or $z_{01}^{(\mathrm{w})}$ but continue to suppose that there is only one wave in the barrier, with a factor $\exp(+\mathrm{i}k\int^z q\,\mathrm{d}z)$, then the q for this wave must still have a large positive imaginary part and it can be shown from (5) that it now has a small positive real part. This wave is called $B(1)$, and it still decreases in amplitude as z increases but the time average of the Poynting vector now has a small negative z component. This leads to the rather surprising result that energy is coming obliquely down in the barrier region towards the main reflecting region near z_{01}. As we proceed downwards on the real z axis the q for this downgoing wave $B(1)$ goes over continuously into the q for the reflected wave $D(1)$ in the duct, whose electric field at the ground is E_1 in (27). Thus for a point z in the barrier, the electric field of the wave $B(1)$ is:

$$\left(\frac{C}{q}\right)^{1/2} E_1 \exp\left(+\mathrm{i}k\int_0^z q\,\mathrm{d}z\right) \tag{28}$$

where the q with a positive real part is chosen and the path of integration is the real z axis. The factor $(C/q)^{1/2}$ arises from the factor $q^{-1/2}$ in the WKB approximate solution of (4). It is used later only where $z = 0$, and $q = +C$ so that it is then unity. The same formula (28) is also used later for other pairs of downgoing waves $D(2n+1)$ in the duct and $B(2n+1)$ in the barrier, where n is an integer. The approximate formula (27) therefore expresses the idea that there are two incident waves approaching the region near z_{01}, namely $D(0)$ in the duct and $B(1)$ in the barrier region. Their relative amplitudes and phases are such that there is no emergent upgoing wave in the barrier and that in the duct there is only the downgoing wave $D(1)$.

To explain the leakage there must clearly be at least one upgoing wave in the barrier region. We suppose that there is such an upgoing wave $B(2)$. Since this is not produced from $B(1)$ and $D(0)$ there must be another downgoing wave, called $B(3)$, in the barrier. This wave approaches the region near z_{01} and is partially reflected to give the upgoing wave $B(2)$ and partially transmitted to give a wave $D(3)$ travelling down in the duct. It is the upgoing wave $B(2)$ which is partially reflected near z_{02} to give $B(1)$, and partially transmitted to give an upgoing wave $L(2)$ above z_{02} which contributes to the leakage. Similarly, the downgoing wave $B(3)$ is produced by partial reflection near z_{02} of an upgoing wave $B(4)$ which is also partially transmitted to give a leaking wave $L(4)$. This process is now continued indefinitely. There is an infinite sequence of upgoing and downgoing waves in the barrier and they give rise to a sequence $D(1)$, $D(3)$, $D(5)$, ... of downgoing waves in the duct. This idea is illustrated schematically in figure 5.

The amplitude of the upgoing wave $B(2)$ increases as it travels obliquely upwards in the barrier region. This is because the Poynting vector is almost horizontal so that the wave has come up obliquely from a point where x is large and negative. Since the mode is slightly attenuated its amplitude at this distant negative x is very large. When this wave reaches the region near z_{02} it is partially reflected to give the downgoing wave $B(1)$ and partially transmitted to give leakage of energy from the mode, wave $L(2)$ in figure 5.

The reflection coefficient for the conversion of wave $B(2n)$ into wave $B(2n-1)$ by downward reflection near z_{02} is given by the phase integral formula:

$$R_2 \simeq \mathrm{i} \exp\left(-2\mathrm{i}k \int_z^{z_{02}} q \, \mathrm{d}z \right). \tag{29}$$

The lower limit z is the coordinate of the point of observation, where $\mathrm{Re}(z) < \mathrm{Re}(z_{02})$. Both real and imaginary parts of q are positive so that the modulus of (29) is very large and much greater than unity. This is because the amplitudes of the incident and reflected waves are measured at the same point and the waves are inhomogeneous (Stratton 1941) so that this point is on different lines of energy flux for the two waves. The result (29) is used later to find the small amplitude of the incident wave $B(2n)$ that would have produced a reflected wave $B(2n-1)$ whose amplitude is large and known.

Similarly the downgoing wave $B(2n+1)$ produces some upgoing wave $B(2n)$ by reflection near z_{01}. The reflection coefficient is similar to (29), namely:

$$R_3 \simeq \mathrm{i} \exp\left(+2\mathrm{i}k \int_z^{z_{01}} q \, \mathrm{d}z \right) \tag{30}$$

This also has modulus greater than unity.

The resultant complex amplitude of the reflected wave in the duct is the sum of the amplitudes of the downgoing waves $D(1)$, $D(3)$, $D(5)$ ... in figure 5. By successive use of the formulae (27)–(30) these amplitudes can be expressed in terms of the amplitude of the incident wave $D(0)$. This process is shown in the following table.

Wave at ground $z = 0$	Wave in barrier	Formula used	Field
$D(1)$		(27)	$E_1 = E_0 \mathrm{i} \exp\left(-2\mathrm{i}k \int_0^{z_{01}} q \, \mathrm{d}z \right)$
	$B(1)$	(28)	$E_2 = \left(\frac{C}{q} \right)^{1/2} E_1 \exp\left(\mathrm{i}k \int_0^z q \, \mathrm{d}z \right)$
	$B(2)$	(29)	$\frac{E_2}{R_2} = -\mathrm{i}E_1 \left(\frac{C}{q} \right)^{1/2} \exp\left(\mathrm{i}k \int_0^z q \, \mathrm{d}z + 2\mathrm{i}k \int_z^{z_{02}} q \, \mathrm{d}z \right)$
	$B(3)$	(30)	$\frac{E_2}{R_2 R_3} = -E_1 \left(\frac{C}{q} \right)^{1/2} \exp\left(\mathrm{i}k \int_0^z q \, \mathrm{d}z + 2\mathrm{i}k \int_{z_{01}}^{z_{02}} q \, \mathrm{d}z \right)$
$D(3)$		(28)	$E_3 = E_1 \left\{ -\exp\left(2\mathrm{i}k \int_{z_{01}}^{z_{02}} q \, \mathrm{d}z \right) \right\}$
	$B(3)$	(28)	$\left(\frac{C}{q} \right)^{1/2} E_3 \exp\left(\mathrm{i}k \int_0^z q \, \mathrm{d}z \right)$
	$B(4)$, $B(5)$	(29), (30)	as for $B(2)$, $B(3)$ with E_1 replaced by E_3
$D(5)$		(28)	$E_5 = E_1 \left\{ \exp\left(2\mathrm{i}k \int_{z_{01}}^{z_{02}} q \, \mathrm{d}z \right) \right\}^2$

The sum of the amplitudes of $D(1)$, $D(3)$, $D(5)$ at the ground is thus a geometrical progression which can be summed provided that

$$\left|\exp\left(+2ik\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z\right)\right| < 1. \tag{31}$$

This is satisfied for well locked modes since $\mathrm{Im}(q)$ is positive when $\mathrm{Re}(q)$ is positive, and the path of integration is close to the real z axis. Thus the electric field E_r of the resultant reflected wave at the ground is

$$E_r \simeq E_0 \mathrm{i} \exp\left(-2ik\int_0^{z_{01}} q\,\mathrm{d}z\right)\left\{1+\exp\left(+2ik\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z\right)\right\}^{-1}. \tag{32}$$

This is the same as the alternative formula (21) whose properties have been studied in §§ 4–6.

For locked modes the upgoing wave $D(0)$ in the duct can never reach the level z_{02} and be reflected there. If it is followed continuously upwards on the real z axis, its field, as given by the WKB expression,

$$E \simeq q^{-1/2}\exp\left(-ik\int_0^z q\,\mathrm{d}z\right) \qquad (\mathrm{Re}(q) > 0) \tag{33}$$

is subdominant, and drops out discontinuously because of the Stokes phenomenon (Stokes 1858), where the Stokes line is crossed at the point P in figure 4. If it is followed on a path which passes on the positive imaginary side of z_{01} it goes over continuously into the downgoing wave $B(1)$.

Suppose now that the frequency is decreased towards f_T. Then the left hand side of (31) increases and eventually a frequency f_0 near f_T is reached for which it is unity. The condition for this is

$$\mathrm{Im}\left(\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z\right) = 0 \qquad \text{for} \qquad f = f_0 \tag{34}$$

which is the condition that an anti-Stokes line shall run from z_{01} to z_{02}. For frequencies less than f_0 the geometric progression used to derive (32), and thence (21), is divergent.

Equation (22) was obtained from (21) by multiplying numerator and denominator by $\exp(-2ik\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z)$. The same equation would have been obtained if the upgoing wave $D(0)$ had first been partially 'reflected' and 'transmitted' near z_{02}, and the sequence of multiple reflections had then taken place between z_{01} and z_{02}. This would give a convergent geometrical progression whose sum is the curly bracket term in (22). The resulting formula has good accuracy when $f \ll f_T$, that is for very leaky modes, as was shown in § 5. It is then very nearly the same as the simple phase integral formula (18) provided that z_0 now means z_{02}. A physical interpretation of this result, similar to that used for locked modes as illustrated in figure 5, is possible but less useful, because for leaky modes the points z_{01}, z_{02} have large imaginary parts. The multiply reflected waves have to be considered in complex space and their physical behaviour on the real z axis does not play any significant part.

It has been mentioned that the modulus of $\exp(-2ik\int_{z_{01}}^{z_{02}} q\,\mathrm{d}z)$ reaches a minimum value at some frequency $f_c \ll f_T$. If $f < f_c$ and f decreases, this term increases and tends to -1 as f goes to zero (S Rotherham private communication). When $f = f_c$ the point P, in figure 4, is at $z = 0$. When $f < f_c$ it is on the negative real z axis. A wave in the duct is completely reflected at $z = 0$ by the earth's surface and therefore, for $f < f_c$, it never undergoes the Stokes phenomenon associated with z_{01}. Hence (22) which modifies (18) by including the effect of z_{01} should not be used for $f < f_c$.

Also (12) is invalid at very low frequencies, because $|\xi_0|$ is very small. Therefore the asymptotic series approximation for (8) cannot be used. It should be emphasized that (12) agrees well with the numerical integration results for frequencies near the transition, and it is, in that sense, 'exact' in the transition region.

It is important to note that the amended form (21) or (22) of the phase integral formula is the same for both well locked and very leaky modes. There is, apparently, no correspondingly simple amendment of the formula (1) for ionospheric reflection. Amendments of (1) have been suggested by Rydbeck (1948) and Heading (1953) (see also Budden 1961 p 446), but they are more complicated than (21), (22) and the amendments are different for the two cases $f < f_p$ and $f > f_p$.

8. Conclusions

The simple phase integral formula can be used for finding the propagation constant of a mode in the troposphere in conditions where ducted propagation is possible provided that the frequency f is not too close to the frequency f_T where the transition from locked to leaky modes occurs. For locked modes $f \gg f_T$ it is necessary to use the branch point z_0 of q, whose real part is the smaller, whereas for leaky modes the z_0 with the greater real part must be chosen. There is thus an abrupt transition in the methods for treating locked and leaky modes. If the tropospheric medium is lossy, the transition is still abrupt. An alternative more accurate form of the phase integral formula can be used which does not show any abrupt transition in going from locked to leaky modes but its accuracy is poor for frequencies near the transition frequency f_T. It can be used, however, for appreciable frequency ranges $f_c < f \ll f_T$ and $f \gg f_T$. For locked modes, this formula can be interpreted in terms of a sequence of multiply reflected waves. For leaky modes a similar interpretation can be found but it does not seem to have any clear physical significance.

Acknowledgments

The author wishes to thank Dr J Blythe and especially S Rotheram of the Marconi Company for their helpful suggestions and criticisms, and Professor M V Wilkes FRS, Director of the University Mathematical Laboratory, for permission to use the computer TITAN. He is also grateful to the Science Research Council for a grant under the Cooperative Awards in Pure Science scheme, and to the Marconi Company who sponsored him. Above all, he would like to thank Dr K G Budden FRS, who suggested this work and gave much time in very helpful discussion and guidance throughout its course.

References

Appleton Sir E V 1946 *Meteorological Factors in Radio Wave Propagation* (London: Physical Society) pp 1–17

Booker H G and Walkinshaw W 1946 *Meteorological Factors in Radio Wave Propagation* (London: Physical Society) pp 80–127

Budden K G 1961 *Radio Waves in the Ionosphere* (London: Cambridge University Press)
Chang H 1971a *Radio Sci.* **6** 465–73
—— 1971b *Radio Sci.* **6** 475–82
Eckersley T L 1931 *Proc. R. Soc.* A **132** 83–98
—— 1932a *J. Inst. Elect. Engrs.* **71** 405–54
—— 1932b *Proc. R. Soc.* A **136** 499–527
—— 1932c *Proc. R. Soc.* A **137** 158–73
Hartree D R, Michel J G L and Nicholson P 1946 *Meteorological Factors in Radio Wave Propagation* (London: Physical Society) pp 127–68
Heading J 1953 *PhD Thesis* University of Cambridge
—— 1962 *Phase Integral Methods* (London: Methuen)
Pekeris C L 1946 *Phys. Rev.* **70** 518–22
Rydbeck O E H 1948 *Chalmers Tek. Högsk. Handl.* No 74
Stokes Sir G G 1858 *Trans. Camb. Phil. Soc.* **10** 105–28
Stratton J A 1941 *Electromagnetic Theory* (New York: McGraw-Hill)
Wait J R 1957 *Proc. Inst. Radio Engr.* **45** 760–7
Whittaker E T and Watson G N 1935 *Modern Analysis* 4th edn (London: Cambridge University Press)