12

Ray theory results for isotropic ionosphere

12.1. Introduction

Although the earth's magnetic field has a very important influence on the propagation of radio waves in the ionospheric plasma, it is nevertheless of interest to study propagation when its effect is neglected. This was done in the early days of research in the probing of the ionosphere by radio waves, and it led to an understanding of some of the underlying physical principles; see, for example, Appleton (1928, 1930). In this chapter, therefore, the earth's magnetic field is ignored. For most of the chapter the effect of electron collisions is also ignored, but they are discussed in §§ 12.3, 12.11. The effect both of collisions and of the earth's magnetic field is small at sufficiently high frequencies, so that some of the results are then useful, for example with frequencies of order 40 MHz or more, as used in radio astronomy. For long distance radio communication the frequencies used are often comparable with the maximum usable frequency, §§ 12.8–12.10, which may be three to six times the penetration frequency of the F-layer. They are therefore in the range 10 to 40 MHz. This is large compared with the electron gyro-frequency which is of order 1 MHz, so that here again results for an isotropic ionosphere are useful, although effects of the earth's magnetic field have to be considered for some purposes.

This chapter is largely concerned with the use of pulses of radio waves and the propagation of wave packets, and uses results from ch. 10, especially §§ 10.2–10.6. But the ray direction and the wave normal direction are the same. There is no lateral deviation, §§ 10.12–10.14, and no phenomenon of the Spitze, §§ 10.9–10.11. The first half is largely concerned with probing the ionosphere at vertical incidence, and the second half with oblique incidence and ray paths.

The present chapter is not concerned with the wave polarisation of the waves, and the field of a wave will be denoted simply by E which may be thought of as one component of the electric field, although the arguments could equally well be applied

to a component of the magnetic field. The analysis in this chapter will usually be expressed in terms of the wave frequency f rather than the angular frequency $\omega = 2\pi f$.

12.2. Vertically incident pulses

Consider first a horizontally stratified ionosphere in which the electron concentration increases monotonically with height z. A vertically incident pulse of radio waves travels upwards with the group velocity $\mathscr{U} = c/n'$, from (5.66), until it is reflected at some height z_0 where $n = 0$, and returns to the ground $z = 0$. The height $z_0(f)$ is called the 'true height' of reflection. It depends on the carrier frequency f of the pulse. The time of travel of the wave packet is P'/c where P' is the equivalent path or group path § 10.16. For a collisionless plasma, to which (5.66) applies,

$$P'(f) = 2h'(f) = 2c\int_0^{z_0} \frac{\mathrm{d}z}{\mathscr{U}}, \tag{12.1}$$

$$h'(f) = \int_0^{z_0} n'\,\mathrm{d}z = \int_0^{z_0} \frac{1}{n}\,\mathrm{d}z, \tag{12.2}$$

and $h'(f)$ is called the 'equivalent height' of reflection. The f is again used to indicate the carrier frequency of the pulse. The integrand of (12.2) is infinite at the upper limit $z = z_0$, but the integral is bounded when z_0 is not at the maximum of an ionospheric layer, because $1/n$ behaves like $(z_0 - z)^{-\frac{1}{2}}$ and its integral converges. The total change of phase of a component wave of frequency f in the pulse is $2\pi f P/c$ where $P(f) = 2h(f)$ is the phase path. Now

$$h(f) = \int_0^{z_0} n\,\mathrm{d}z \tag{12.3}$$

is called the 'phase height' of reflection. On the real part of its path where the wave is propagated, $n \leqslant 1$ and (12.2), (12.3) then show that

$$h(f) < z_0(f) < h'(f). \tag{12.4}$$

Although z_0 is a function of f, this was not allowed for when deriving (12.2). To show that the result is correct it is instructive to derive (12.2) by another method. For any component frequency f in the pulse, the total phase change for the upward and downward path is $4\pi f h(f)/c$, where $h(f)$ is given by (12.3). Thus the field of the reflected wave reaching the ground is, compare (11.119):

$$E(t) = \int_{-\infty}^{\infty} M(f - f_1) \exp\left\{2\pi i f\left(t - \frac{2}{c}\int_0^{z_0(f)} n\,\mathrm{d}z\right)\right\}\mathrm{d}f. \tag{12.5}$$

Here $M(f - f_1)$ is the Fourier transform of the pulse envelope, shifted in frequency, as in (11.117), (11.118), so that it is centred on the carrier frequency of the pulse, here temporarily denoted by f_1. It is a slowly varying function of f. The time τ of arrival of the pulse at the ground is the value of t that makes the phase of the integrand

stationary for frequencies near f_1. Hence

$$\left[\frac{\partial}{\partial f}\left\{f\left(\tau-\frac{2}{c}\int_0^{z_0} n\,\mathrm{d}z\right)\right\}\right]_{f=f_1}=0 \tag{12.6}$$

so that

$$c\tau = 2\int_0^{z_0}\left\{\frac{\partial}{\partial f}(fn)\right\}_{f=f_1}\mathrm{d}z + 2f_1 n(z_0)\left(\frac{\partial z_0}{\partial f}\right)_{f=f_1} \tag{12.7}$$

Now $n(z_0)$ is zero and $\partial z_0/\partial f$ is bounded except when f_1 is near the penetration frequency of an ionospheric layer, and this case needs special treatment by full wave methods; see § 15.11. Apart from this the last term of (12.7) is zero, and (5.66) shows that the remaining term is the same as (12.2). Thus the dependence of z_0 on frequency does not affect the result. This is because, near the level of reflection, n is close to zero and the wavelength in the medium is very large. Hence small variations of reflection level do not affect the phases of the component waves.

Curves showing the form of the functions $h'(f)$, $z_0(f)$, $h(f)$ in specific cases have been given by Budden (1961a) and others. See also problems 12.1–12.3.

12.3. Effect of collisions on phase height $h(f)$ and equivalent height $h'(f)$

The uniform approximation solution (8.72) for vertically incident waves, and the W.K.B. solutions derived from it, are functions of the height z. The fields they represent can, of course, only be observed at real values of z. But in these solutions we can make z complex and require that the same differential equations are satisfied. This is the process of analytic continuation. It shows that (8.72) is still valid when continued analytically to use complex values of z, provided that the boundary condition at great heights is satisfied. When z is real and large the solution must represent a purely upgoing wave. It can be shown that this is so for the monotonic functions $N(z)$ considered in this section.

When collisions are allowed for the refractive index n is given by (4.9) and is complex. There is no real value z_0 of z that makes $n=0$, but there is a complex z_0. The reflected wave is still given by (7.153) with $C=1$ and $q=n$ for vertical incidence. The integrals are now contour integrals in the complex z plane. The factor i from (8.73) is now included although it is unimportant here. Thus the reflection coefficient measured at the ground is, from (7.151):

$$R = \mathrm{i}\exp\{-2\mathrm{i}kh(f)\} \tag{12.8}$$

where the phase height $h(f)$ is still given by (12.3) but is now complex. In (12.2) n' is no longer equal to $1/n$ so this equation must be rewritten

$$h'(f) = \int_0^{z_0} n'\,\mathrm{d}z = \int_0^{z_0}\frac{\partial}{\partial f}(fn)\mathrm{d}z \tag{12.9}$$

and this defines a complex equivalent height of reflection.

The quantities $2h(f)$ and $2h'(f)$ are special cases of the phase path $P(f)$ and the equivalent path $P'(f)$ respectively. The physical interpretation of complex values of these quantities was discussed in §11.16. It was shown that the time of travel of a wave packet is $\mathrm{Re}\{P'(f)\}/c$ so that the effective equivalent height of reflection is $\mathrm{Re}\{h'(f)\}$.

Usually in radio observations it is the magnitude $|R|$ of the reflection coefficient that is measured. Now (12.8) gives

$$\ln|R| = 2k\,\mathrm{Im}(h). \tag{12.10}$$

This is zero if there are no collisions. Introduce a small non-zero collision frequency ν and suppose that it is independent of height z. This is a useful assumption because $\nu(z)$ is much less strongly height dependent than $N(z)$; see fig. 1.2. There is then a small change of the phase height h:

$$\delta h \approx Z(\partial h/\partial Z)_{Z=0} = Z\left\{\int_0^{z_0} (\partial n/\partial Z)_{Z=0}\,\mathrm{d}z + n(z_0)\partial z_0/\partial Z\right\}. \tag{12.11}$$

But $n(z_0)$ is zero, and from (4.9):

$$(\partial n/\partial Z)_{Z=0} = -\tfrac{1}{2}\mathrm{i}X/n = \tfrac{1}{2}\mathrm{i}(n - n') \tag{12.12}$$

so that (12.11), with (12.3), (12.9), gives

$$\delta h \approx \tfrac{1}{2}\mathrm{i}Z(h - h'). \tag{12.13}$$

Then finally, from (12.10), the attenuation is

$$-\ln|R| \approx \frac{\nu}{c}(h' - h). \tag{12.14}$$

This formula was used for estimating the electron collision frequency ν (Appleton, 1935; Farmer and Ratcliffe, 1935). When the earth's magnetic field is allowed for the theory needs some modification. This is discussed in §13.10.

For one useful simple model of the ionosphere, there is free space in the height range $0 \leqslant z \leqslant h_0$ and above this the electron concentration $N(z)$ is proportional to $z - h_0$. The collision frequency is independent of z so that Z is a constant. For this model it can be shown (see problem 12.1) that the real parts of (12.3) and (12.9) are independent of Z. It can further be shown that (12.14) is true for any value of ν and is not restricted to small values.

In another model the electron concentration increases exponentially with height z. The values of $h(f)$ and $h'(f)$ are then as given in problem 12.2. If there is a constant collision frequency, the squared refractive index is

$$n^2 = 1 - F^2\mathrm{e}^{\alpha z}/(1 - \mathrm{i}Z). \tag{12.15}$$

It can then be shown that

$$-\ln|R| = k(\arctan Z)\,\mathrm{Re}\{h'(f) - h(f)\}. \tag{12.16}$$

The proof is left to the reader. If Z is small, this is the same as (12.14).

12.4. Equivalent height for a parabolic height distribution of electron concentration

Suppose that the electron concentration is given by the parabolic law

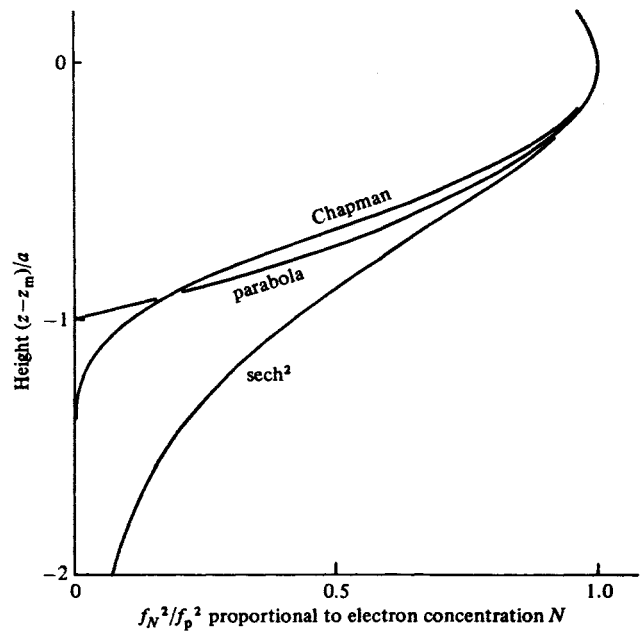
$$\begin{aligned} N(z) &= N_m\{1-(z-z_m)^2/a^2\} \quad \text{for} |z-z_m| \leqslant a, \\ N(z) &= 0 \qquad\qquad\qquad\quad \text{for} |z-z_m| \geqslant a. \end{aligned} \tag{12.17}$$

One curve in fig. 12.1 shows how N depends on z in this case. It has a maximum value N_m where $z = z_m$. The constant a is called the 'half thickness' of the parabolic layer. This example is of particular importance in the study of layers that have a maximum of electron concentration. Some important properties of any layer depend on the curvature of the $N(z)$ curve at the maximum. By studying the properties of a parabolic layer with the same curvature at the maximum it is possible to derive results for other layers.

The plasma frequency that corresponds to the maximum electron concentration N_m will be denoted by f_p and is called the 'penetration frequency' because waves of frequency less than f_p are reflected below z_m and vertically incident waves of greater frequency travel right through the layer without reflection. It is convenient to use a new height variable

$$\zeta = (z - z_m)/a. \tag{12.18}$$

Fig. 12.1. Dependence of f_N^2/f_p^2, proportional to electron concentration $N(z)$, on height z for the parabolic distribution (12.17), the sech^2 distribution (12.26) and the Chapman layer (1.9), (1.10) with $H = \frac{1}{2}a$. All three distributions have a maximum at the same height $z = z_m$ and their curvatures there are the same.



Then (12.17) gives

$$f_N^2 = f_p^2(1-\zeta^2) \text{ for } |\zeta| \leqslant 1. \tag{12.19}$$

Collisions are neglected. The real refractive index n and group refractive index n' are given, from (4.10), (5.66), by

$$n = 1/n' = \{1-(1-\zeta^2)f_p^2/f^2\}^{\frac{1}{2}} \text{ for } |\zeta| \leqslant 1. \tag{12.20}$$

When $|\zeta| \geqslant 1$, f_N is zero and n, n' are both unity.

Two cases must now be studied. The first is for $f < f_p$ so that reflection occurs at a height z_0 below z_m. The value ζ_0 of ζ at the reflection level is

$$\zeta_0(f) = -\{1-f^2/f_p^2\}^{\frac{1}{2}}. \tag{12.21}$$

The equivalent height of reflection is

$$h'(f) = \int_0^{z_0} n'\,\mathrm{d}z = z_m - a + \tfrac{1}{2}a\frac{f}{f_p}\ln\left(\frac{f_p+f}{f_p-f}\right). \tag{12.22}$$

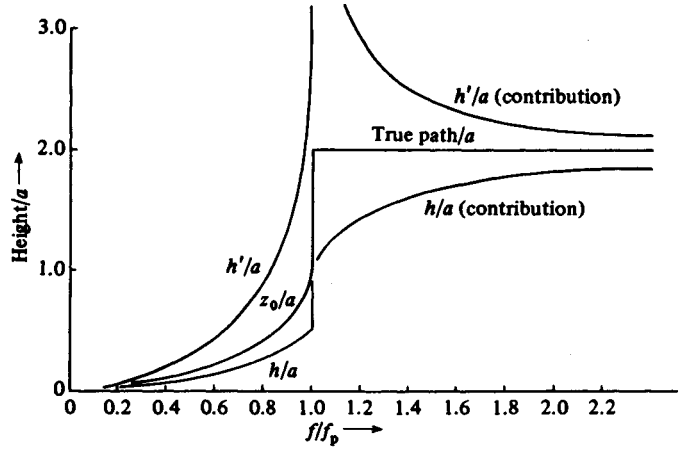
Similarly the phase height is

$$h(f) = \int_0^{z_0} n\,\mathrm{d}z = z_m - \tfrac{1}{2}a - \tfrac{1}{4}a\left(\frac{f_p}{f} - \frac{f}{f_p}\right)\ln\left(\frac{f_p+f}{f_p-f}\right). \tag{12.23}$$

The dependence on frequency f of the true height from (12.21), equivalent height (12.22) and phase height (12.23) are shown in the left half of fig. 12.2.

The second case is for $f > f_p$ so that the waves travel right through the layer without reflection. They may, however, be reflected from a higher layer, and the parabolic layer can contribute to the equivalent height of reflection. The contribution made by the region between the ground and the top of the parabolic layer will

Fig. 12.2. Left half shows how the equivalent height h', the true height of reflection z_0 and the phase height h depend on frequency f for the parabolic distribution (12.17) of $N(z)$. Right half shows the contributions to h' and h, and the true path for one passage through the layer.



now be calculated. The contribution to the true height is clearly $z_m + a$. The contribution to the equivalent height is

$$h'(f) = \int_0^{z_m+a} n' \, dz = z_m - a + \frac{f}{f_p} \ln\left(\frac{f+f_p}{f-f_p}\right). \tag{12.24}$$

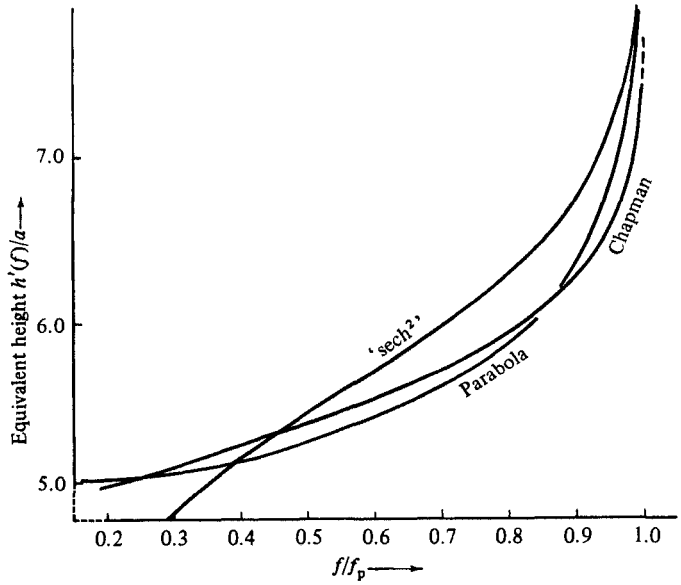
The contribution to the phase height is

$$h(f) = \int_0^{z_m+a} n \, dz = z_m + \tfrac{1}{2}a\left(\frac{f}{f_p} - \frac{f_p}{f}\right) \ln\left(\frac{f+f_p}{f-f_p}\right). \tag{12.25}$$

The dependence on frequency f of the contribution to the true height, the equivalent height and the phase height are shown in the right half of fig. 12.2.

The above results show that the equivalent height approaches infinity as the frequency f approaches the penetration frequency f_p. The group velocity of a pulse is smallest near the top of its trajectory. For a frequency just below penetration, the region of low group velocity is thicker than it is for lower frequencies, because the gradient of electron concentration is small near the level of reflection. This explains why the equivalent height is large for frequencies close to penetration. The pulse spends a relatively long time near the level of maximum electron concentration, and for a small range of frequencies below f_p the shape of the $h'(f)$ curve is determined predominantly by the curvature of the $N(z)$ curve at the maximum. The shape of the $N(z)$ curve at lower levels is less important for these frequencies, because the pulse spends relatively little time there. This property is illustrated in fig. 12.3 which shows

Fig. 12.3. Curves of $h'(f)$ for the three electron distributions of fig. 12.1. For all three curves, $z_m/a = 6$.



the $h'(f)$ curves for three different electron height distributions, all with the same curvature at the maximum. These are the parabolic layer, the 'sech^2' layer discussed below, and the Chapman layer given by (1.9), (1.10) with $H = \frac{1}{2}a$; see fig. 12.1.

The same property was further illustrated by Appleton and Beynon (1940, appendix). They plotted observed values of $h'(f)$ against $\frac{1}{2}(f/f_p)\ln[(f_p+f)/(f_p-f)]$. For a parabolic layer, (12.22) shows that the graph would be a straight line. For their actual ionospheric layers they fit to a straight line was good for the E-layer, and fairly good but with more scatter of the points for two examples of the F-layer. From graphs of this kind, the slope and the intercept can be used to get estimates of a and $z_m - a$ respectively.

The expression (12.2) for the equivalent height cannot be used for frequencies very close to the penetration frequency, and in this case the problem must be treated by full wave theory. This is done in § 15.11, where it is shown that the effective value of $h'(f)$ near the penetration frequency remains finite but attains a high maximum value. The amplitude of the reflected wave is then small, however, and the maximum is not easy to observe. In most practical cases there is no need to take account of the full wave correction.

The parabolic layer has discontinuities in the gradient of electron concentration at the top and bottom of the layer, $z = z_m \pm a$. In this respect it differs from actual ionospheric layers, in which the transitions are believed to be gradual. A distribution which does not have this drawback is given by

$$f_N^2 = f_p^2 \operatorname{sech}^2\left(\frac{z - z_m}{a}\right). \tag{12.26}$$

One curve in fig. 12.1 shows how the electron concentration N varies with height z in this case. As for the parabola there is a maximum of electron concentration where $z = z_m$, and the penetration frequency is f_p. The electron concentration is never zero; it must have a small non-zero value even at the ground. It is convenient to use the variable ζ defined by (12.18). Then the equivalent height, when $f < f_p$, is given by

$$h'(f) = a\ln\left[\left(\frac{f_p^2}{f^2} - 1\right)^{-\frac{1}{2}} \sinh\frac{z_m}{a} + \left\{\left(\frac{f_p^2}{f^2} - 1\right)^{-1} \sinh^2\left(\frac{z_m}{a}\right) - 1\right\}^{\frac{1}{2}}\right]. \tag{12.27}$$

The curvature of the distribution (12.26) at the maximum is the same as that of the parabolic layer (12.17) for the same values of f_p and a. Fig. 12.3 shows how $h'(f)$ in (12.27) varies with frequency, and the corresponding curve for the parabola is also shown. At high frequencies $h'(f)$ is greater for the 'sech^2' profile. Although the true height of reflection must be lower, the larger number of electrons at low levels causes greater group retardation, which more than compensates for the shorter true path. At low frequencies, however, $h'(f)$ is smaller for the 'sech^2' layer, because the true height falls rapidly as the frequency decreases.

During the daytime there are two main layers in the ionosphere known as E and

F. It is now believed that they are not distinct but that the region between them is ionised with an electron concentration nearly as great as in the E-layer itself; see §1.8 and fig. 1.3. It is nevertheless instructive to assume that these two layers have separate parabolic distributions. Let f_{pE}, z_{mE}, a_E be the penetration frequency, height of maximum and half thickness respectively for the E-layer, and let f_{pF}, z_{mF}, a_F be the corresponding quantities for the F-layer above it. It is assumed that $f_{pF} > f_{pE}$ and that the layers do not overlap. Then if $f < f_{pE}$, $h'(f)$ is given by (12.22) with f_p, z_m and a replaced by f_{pE}, z_{mE} and a_E. If $f_{pE} < f < f_{pF}$,

$$h'(f) = z_{mF} - 2a_E - a_F + a_E \frac{f}{f_{pE}} \ln\left(\frac{f+f_{pE}}{f-f_{pE}}\right) + \tfrac{1}{2} a_F \frac{f}{f_{pE}} \ln\left(\frac{f_{pF}+f}{f_{pE}-f}\right). \quad (12.28)$$

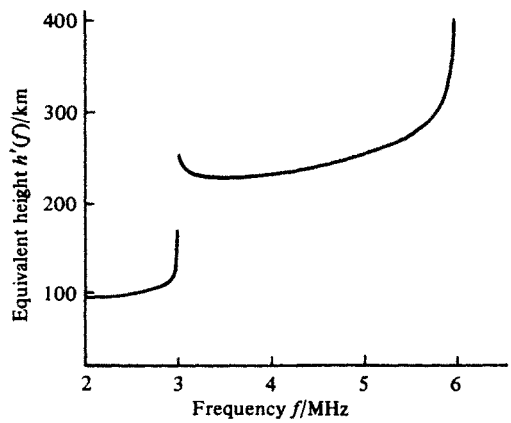
This function is shown plotted for typical values in fig. 12.4. The full wave correction (§ 15.11) near $f_p^{(E)}$ and $f_p^{(F)}$ is not used.

Although the earth's magnetic field has been neglected in deriving (12.28), the curve of fig. 12.4 is very similar in general form to some $h'(f)$ records obtained in practice. A characteristic feature of these records is the infinity at the penetration frequency f_{pE} of the E-layer. When such an infinity appears, it can usually be attributed to penetration of one layer. For frequencies just above it, the waves are being reflected from a higher layer, but are retarded in their passage through the layer that has just been penetrated. For greater frequencies the retardation is less because the group velocity in the lower layer is not so small as for frequencies near the penetration frequency.

12.5. Effect of a 'ledge' in the electron height distribution

A possible form for the electron height distribution $N(z)$ is shown in fig. 12.5(a). Here, for heights below the maximum, $N(z)$ increases monotonically as z increases,

Fig. 12.4. $h'(f)$ curve when the ionosphere consists of two separate parabolic layers. For the lower layer (E), $f_{pE} = 3$ MHz, $z_{mE} = 100$ km, $a_E = 10$ km, and for the upper layer (F), $f_{pF} = 6$ MHz, $z_{mF} = 240$ km, $a_F = 40$ km.



but there is a part of the curve where dN/dz is small. This feature is often called a 'ledge' in the distribution. Suppose that in the middle of the ledge the plasma frequency f_N is denoted by F_N. The form of the $h'(f)$ curve in this case is shown in fig. 12.5(b). For frequencies near F_N the gradient dN/dz is small at the level of reflection, and the group retardation is large, so that $h'(f)$ is large. For greater frequencies, the reflection level is higher and the group retardation in the ledge is less. Consequently $h'(f)$ is less. For still greater frequencies, $h'(f)$ increases again and becomes infinite at the penetration frequency.

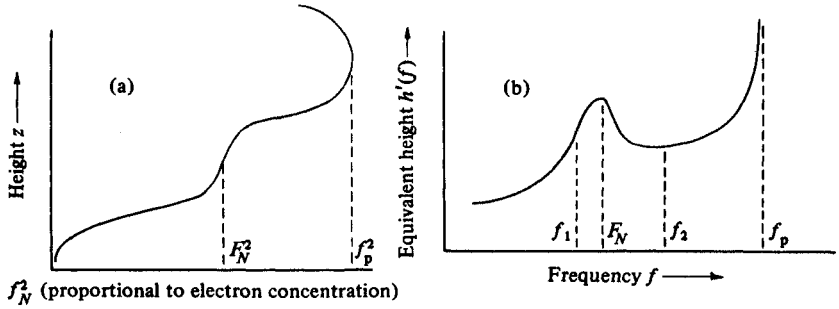
The effect of a ledge is therefore to give a maximum in the $h'(f)$ curve. The size of the maximum increases if the gradient dN/dz in the ledge decreases, but $h'(f)$ is always bounded as long as $N(z)$ increases monotonically. Hence, when $N(z)$ is a monotonically increasing function, the corresponding part of the $h'(f)$ curve is bounded and continuous. If $N(z)$ has a maximum, the $h'(f)$ curve shows an infinity at the penetration frequency, and is discontinuous there. When corrected by the full wave theory (§ 15.11), the $h'(f)$ curve does not go to infinity but has a maximum near the penetration frequency, although it is still discontinuous. The maximum is rarely observed on actual $h'(f)$ records, and for practical purposes the curve may be said to go to infinity. When the effect of the earth's magnetic field is allowed for, a discontinuity of a different kind can appear in the $h'(f)$ curve for the extraordinary wave at the gyro-frequency (see § 13.2).

The F1-layer often appears as a ledge below the F2-layer, and the $h'(f)$ curve for the whole F-layer is then similar to fig. 12.5(b). Sometimes the F1-layer may attain an actual maximum, and then the maximum in the $h'(f)$ record will go over into an infinity associated with penetration.

12.6. The calculation of electron concentration $N(z)$, from $h'(f)$

In the preceding sections some examples have been given of the calculation of $h'(f)$ when $N(z)$ is known. Of much greater practical importance is the calculation of $N(z)$

Fig. 12.5. (a) Shows an electron height distribution $N(z)$ with a ledge. (b) The resulting $h'(f)$ curve, not to scale, showing a maximum at a frequency near the plasma frequency F_N in the ledge.



when $h'(f)$ is known, since in the commonest form of radio sounding of the ionosphere it is $h'(f)$ that is observed. The relation between $h'(f)$ and $N(z)$ may be written

$$h'(f) = \int_0^{z_0} n'(N,f)\,\mathrm{d}z. \tag{12.29}$$

The problem is to solve this integral equation for $N(z)$. This can be accomplished analytically in the special case when the earth's magnetic field is neglected. The electron concentration N is proportional to the square of the plasma frequency $f_N(z)$. Hence $f_N(z)$ is to be found. The problem will actually be solved by finding the inverse function $z(f_N)$.

The group refractive index n' is given by (5.66) with ω replaced by f. Hence (12.29) may be written

$$\frac{1}{f}h'(f) = \int_0^{z_0} \frac{\mathrm{d}z}{(f^2 - f_N^2)^{\frac{1}{2}}} \tag{12.30}$$

where z_0 is that value of z for which $f_N = f$. Now it is convenient to use new variables

$$u = f_N^2, \quad v = f^2, \tag{12.31}$$

so that

$$v^{-\frac{1}{2}}h'(v^{\frac{1}{2}}) = \int_0^{z_0} \frac{\mathrm{d}z}{\{v - u(z)\}^{\frac{1}{2}}}. \tag{12.32}$$

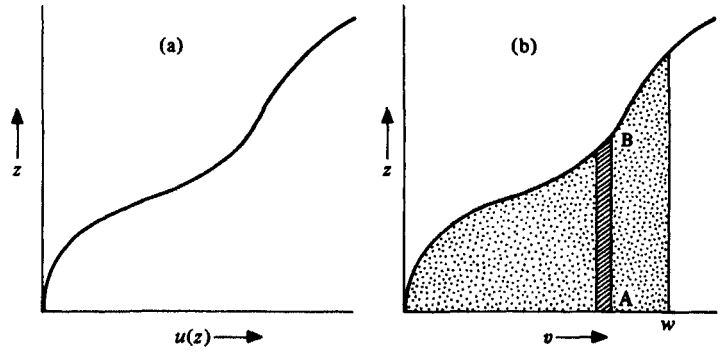
Here $u(z)$ is the unknown function that is to be found.

When $u(z)$ is a monotonically increasing function of z, (12.32) can be solved by a standard method (see, for example, Whittaker and Watson, 1927, §11.8). An alternative method using Laplace transforms has been given by Manning (1947, 1949). Multiply both sides by $(1/\pi)(w - v)^{-\frac{1}{2}}$, where

$$w = g^2 \tag{12.33}$$

and g is the value of f_N for which the true height $z(g)$ is to be found. Integrate with

Fig. 12.6. (a) The unknown function $u(z)$ proportional to electron concentration. (b) The same curve plotted in the v–z plane.



respect to v from 0 to w. Then (12.32) gives

$$\frac{1}{\pi}\int_0^w v^{-\frac{1}{2}}(w-v)^{-\frac{1}{2}}h'(v^{\frac{1}{2}})\,\mathrm{d}v = \frac{1}{\pi}\int_0^w\left[\int_0^{z_0}\frac{\mathrm{d}z}{(w-v)^{\frac{1}{2}}(v-u)^{\frac{1}{2}}}\right]\mathrm{d}v. \tag{12.34}$$

Now suppose that the unknown function $u(z)$ is as shown by the curve of fig. 12.6(a). The integral on the right is a double integral with respect to z and v. It is to be integrated first with respect to z while v is kept constant, and the top limit for z is the value that makes $u = v$. Fig. 12.6(b) is a diagram of the $v - z$ plane, and the curve of fig. 12.6(a) is copied in it. The first integration is over the shaded strip AB. The result is then to be integrated with respect to v from $v = 0$ to $v = w$. It is thus clear that the whole integration extends over the dotted region of fig. 12.6(b). The order of integration may now be changed provided that the limits are suitably modified. Then the right side of (12.34) becomes

$$\frac{1}{\pi}\int_0^{z_0(g)}\left[\int_u^w\frac{\mathrm{d}v}{(w-v)^{\frac{1}{2}}(v-u)^{\frac{1}{2}}}\right]\mathrm{d}z, \tag{12.35}$$

where $z_0(g)$ is the value of z that makes $u = w$. The integration with respect to v may be effected by elementary methods (by the substitution $v = w\cos^2\theta + u\sin^2\theta$) and gives simply π. Hence the expressions (12.34) and (12.35) are equal to $z_0(g)$. On the left of (12.34) put $v = w\sin^2\alpha$, which gives

$$z_0(g) = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} h'(w^{\frac{1}{2}}\sin\alpha)\,\mathrm{d}\alpha. \tag{12.36}$$

Now (12.33) shows that $w^{\frac{1}{2}} = g$. We now put $g = f_N$ so that (12.36) may be written

$$z(f_N) = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} h'(f_N\sin\alpha)\,\mathrm{d}\alpha, \tag{12.37}$$

which is the required solution.

The equation (12.32) was solved by Abel for a different problem. A particle is projected horizontally with velocity V on to a frictionless slope, for which the height $H(x)$ is an unknown function of the horizontal distance x. The slope is assumed to be so small that the horizontal component of the velocity may be taken to be the same as the actual oblique velocity. The time T taken for the particle to come to rest and then return to its starting point is given by

$$T(V) = 2\int_0^{x_0}\frac{\mathrm{d}x}{\{V^2 - 2gH(x)\}^{\frac{1}{2}}} \tag{12.38}$$

where g is the gravitational acceleration, and x_0 is the value of x for which $H(x) = \frac{1}{2}V^2/g$. The function $T(V)$ can be determined experimentally. Abel's problem was to find the function $H(x)$. Clearly (12.38) is of the same form as (12.32) and can be solved in the same way (see de Groot, 1930). Ratcliffe (1954) has used this analogy by constructing a mechanical model in which a steel ball-bearing is projected up an incline shaped like the electron distribution $N(z)$ in the ionosphere. This exhibits

many of the features of the reflection of a pulse of radio waves projected up into the ionosphere.

Suppose now that $u(z)$ is not a monotonically increasing function of z, but has one maximum and one minimum, as shown in fig. 12.7(a). At the maximum let $z = z_1$, $u = u_p$ where

$$u_p = f_p^2, \tag{12.39}$$

and f_p is the penetration frequency. Let $z = z_3$ be the height where u again has the value u_p (see fig. 12.7(a)). It was shown in § 12.4 that the $h'(f)$ curve must have an infinity where $f = f_p$, so that the curve is as in fig. 12.8(a). It is now assumed that this curve is given, and we require to find the curve of fig. 12.7(a). This problem cannot be solved completely. Only the part of the curve below z_1 can be found unambiguously.

The equation (12.32) still holds. To solve it, we apply the method of the last section. Multiply both sides by $(1/\pi)(w - v)^{-\frac{1}{2}}$ where w is given by (12.33). Integrate

Fig. 12.7. (a) Similar to fig. 12.6(a) but the distribution now has a valley. (b) Similar to fig. 12.6(b).

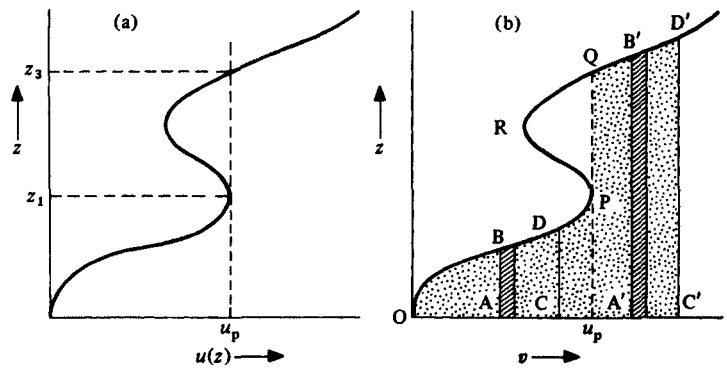
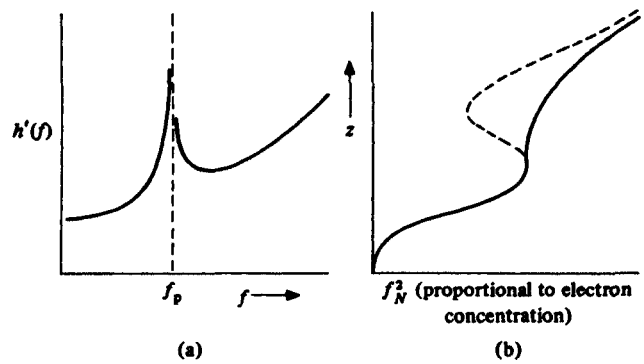


Fig. 12.8. (a) The $h'(f)$ curve from which $u(z)$ is to be found. (b) Result of applying Abel's method to (a). The unknown curve $u(z)$ is shown for comparison, as a broken line.



with respect to v from 0 to w, which gives (12.34), whose right side is a double integral with respect to z and v. Fig. 12.7(b) is a diagram of the v–z plane and the curve of fig. 12.7(a) is copied in it. Suppose first that w is less than u_p and is given by the ordinate CD in the figure. Then the double integral extends over the area OACDB. The order of integration may be changed as before and the final result is

$$z(f_N) = \int_0^{\frac{1}{2}\pi} h'(f_N \sin\alpha)\,\mathrm{d}\alpha \text{ for } f_N < f_p, \tag{12.40}$$

which is the same as (12.37), and enables the part of the curve below $z = z_1$ in fig. 12.7(a) to be found. Next suppose that w is greater than u_p, and is given by the ordinate C'D' in fig. 12.7(b). In the double integral on the right of (12.34) the first integration with respect to z extends from $z = 0$ up to the curve. When v is just less than u_p, the upper limit is on the part DP of the curve. When v becomes just greater than u_p, the upper limit suddenly increases to the part QB' of the curve. It is thus clear that the double integral extends over the dotted area of fig. 12.7(b), and the area PQR is not included. Now the order of integration may be changed, as in the last section, but if the right side of (12.34) were written in the form of (12.35), the integration would extend over the dotted area plus the area PQR. Equation (12.35) is therefore no longer correct. The following integral must be subtracted from it:

$$\frac{1}{\pi}\int_{z_1}^{z_3}\left[\int_u^{u_p}\frac{\mathrm{d}v}{(w-v)^{\frac{1}{2}}(v-u)^{\frac{1}{2}}}\right]\mathrm{d}z. \tag{12.41}$$

This is the integral over the area PQR. Now (12.35) can be integrated as before, and gives simply $z_0(f_N)$. The integration with respect to v in (12.41) can be effected in the same way as before, and gives for this term

$$\zeta(g) = \frac{1}{\pi}\int_{z_1}^{z_3}\arccos\left(\frac{w-2u_p+u}{w-u}\right)\mathrm{d}z. \tag{12.42}$$

Hence (12.34) becomes

$$z(f_N) = \frac{2}{\pi}\int_0^{\frac{1}{2}\pi} h'(f_N \sin\alpha)\,\mathrm{d}\alpha + \zeta(f_N). \tag{12.43}$$

Hence Abel's integral (12.37) gives the correct value for $z(f_N)$ when $f_N < f_p$, but gives a result which is too small by $\zeta(f_N)$ when $f_N > f_p$. This is illustrated in fig. 12.8(b). Thus the expression (12.42) is the error which is contributed by the part of the ionosphere between $z = z_1$ and $z = z_3$.

In the range $z_1 < z < z_3$ it is clear that $u < u_p$. If $w \gg u_p$, the angle arccos (...) in (12.42) must be very small, and hence the error $\zeta(f_N)$ is small. This means that the error is not serious for frequencies which greatly exceed the penetration frequency f_p. In particular the E-layer (for which f_p is of the order 2 to 4 MHz in the daytime) would not seriously affect results for the densest part of the F-layer which may reflect frequencies of order 5 to 10 MHz. There is now good evidence

to show that the minimum of the electron density between the E- and F-layers is very shallow, and this would mean that the error ζ is always small.

When $f_N > f_p$, the integral in (12.43) covers a range which includes the infinity at $f = f_p$ in the $h'(f)$ curve (fig. 12.8(a)). This infinity arises from a logarithm term of the form given by (12.22) and (12.24). If the integral is evaluated analytically in this case, the range of integration must be divided into the two parts $0 \leqslant \sin\alpha \leqslant (f_p/f_N) - \varepsilon$ and $(f_p/f_N) + \varepsilon \leqslant \sin\alpha \leqslant 1$ where ε is an arbitrarily small quantity. The two parts of the integral are evaluated separately and added, and ε is then allowed to become indefinitely small. In this way it can be shown that the sum approaches a definite limit known as the 'principal value' of the integral, which is the value to be used in (12.43) (for further details see, for example, Whittaker and Watson, 1935, § 4.5).

12.7. Ray paths at oblique incidence

We now study some examples of the paths of obliquely travelling wave packets in a horizontally stratified isotropic ionosphere with negligible collision frequency. Without loss of generality the path can be assumed to be in the plane $y = 0$, and to start from a transmitter at the origin. Then its equation, from (10.14) is

$$x = S \int_0^z \frac{\mathrm{d}z}{q}, \tag{12.44}$$

$$q^2 = n^2 - S^2 \tag{12.45}$$

where $S = \sin\theta$, and θ is the angle between the predominant wave normal and the vertical in free space.

Consider, first, the simple model ionosphere of § 12.3 in which there is free space where $0 \leqslant z \leqslant h_0$, and above this $N(z)$ is proportional $z - h_0$. Let $C = \cos\theta$. Then

$$\begin{aligned} q^2 &= C^2 - \alpha(z - h_0)/f^2 \quad \text{when } z \geqslant h_0, \\ q &= C \qquad\qquad\qquad\quad \text{when } z \leqslant h_0. \end{aligned} \tag{12.46}$$

where α is a constant. Evaluation of the integral in (12.44) is often most easily achieved by changing to q as the variable of integration. If (12.46) is used, it gives, for $z \geqslant h_0$

$$x = h_0 \tan\theta + 2\frac{f^2 S}{\alpha}(C - q) = h_0 \tan\theta + \frac{f^2 \sin 2\theta}{\alpha} - \frac{2f^2 \sin\theta}{\alpha}\left\{\cos^2\theta - \frac{\alpha(z - h_0)}{f^2}\right\}^{\frac{1}{2}} \tag{12.47}$$

which shows that within the ionosphere the ray path is a parabola; compare (10.26) which gives the same result for $h_0 = 0$. It is now of interest to find the total horizontal range D traversed by the wave packet when it returns to the ground. The top of its trajectory is where $\mathrm{d}x/\mathrm{d}z = \infty$, that is where $q = 0$ and at this point x is $\frac{1}{2}D$. Hence

from (12.47)

$$D = 2h_0 \tan\theta + 2f^2(\sin 2\theta)/\alpha. \tag{12.48}$$

The first term is contributed by the two straight sections of the path below the ionosphere and the second term is the part within the ionosphere.

Fig. 12.9 shows how D/h_0 depends on θ for some typical cases. It can easily be shown that if $f > 4\alpha h_0$ the curve has a maximum and a minimum so that for some values of D (12.48) is satisfied by three different values of θ. This means that wave packets can travel from the transmitter to the receiver by three different paths. This is only possible if $D > 6h_0\sqrt{3}$. The proof is left as an exercise for the reader.

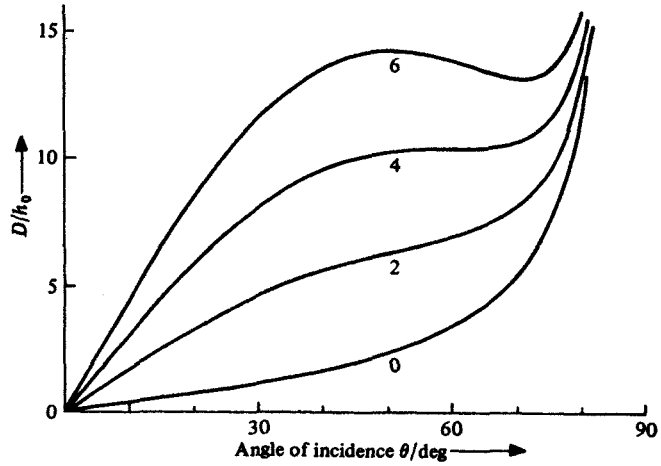
Suppose that the horizontal line $D/h_0 =$ constant in fig. 12.9 just touches the curve at either a maximum or a minimum. Then two of the ray paths have moved to coincidence which means that a caustic surface meets the ground at the range D; see §§ 8.22, 10.18. If the touching point is at a minimum, the range D is a skip distance; see § 10.22 and fig. 10.17. If it is at a maximum, D is a reversed skip distance. Diagrams showing the shape of the rays and of the caustics for this and other cases have been given by Maslin (1976a, b).

Suppose, next, that the electron concentration is given by the parabolic law (12.17), and use ζ from (12.18). Then

$$\begin{aligned} q^2 &= C^2 - (1-\zeta^2)f_p^2/f^2 \quad \text{for } |\zeta| \leqslant 1, \\ q &= C \qquad\qquad\qquad\quad \text{for } |\zeta| \geqslant 1, \end{aligned} \tag{12.49}$$

where f_p is the penetration frequency of the layer. The equation of a ray path within

Fig. 12.9. Shows how the horizontal range D depends on angle of incidence θ for the model ionosphere with free space below the height h_0, and $f_N^2 = \alpha(z - h_0)$ above it. The earth's curvature is neglected. The numbers by the curves are the values of $f^2/\alpha h_0$.



the ionosphere is, from (12.44)

$$x = h_0 \tan\theta + aS \int_{-1}^{\zeta} \{C^2 - (1-\zeta^2) f_p^2/f^2\}^{-\frac{1}{2}} \, d\zeta \tag{12.50}$$

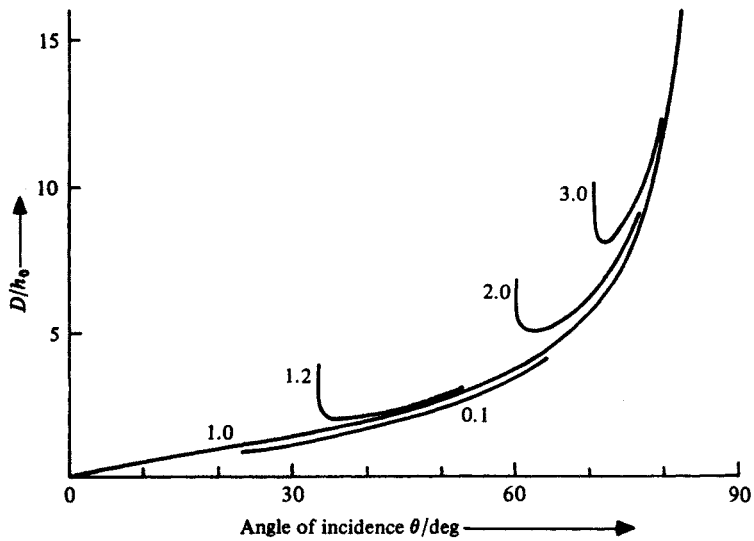
where $h_0 = z_m - a$. The horizontal range D is found, as in (12.48), by finding the part between the transmitter and the top of the trajectory and doubling it. This gives

$$D = 2h_0 \tan\theta + a \sin\theta \frac{f}{f_p} \ln\left(\frac{f_p + f\cos\theta}{f_p - f\cos\theta}\right). \tag{12.51}$$

The first term is contributed by the two straight sections of the path below the ionosphere, and the second term is the part within the ionosphere.

Fig. 12.10 shows how D/h_0 depends on θ for some typical cases. When $f \leqslant f_p$ any wave packet is reflected whatever the value of θ, and D varies between 0 and ∞. When $f > f_p$, however, wave packets that are near vertical incidence can penetrate the ionosphere, and do not return to the ground. For reflection it is necessary that $\cos\theta < f_p/f$ which gives a minimum value of θ, and (12.51) shows that D tends to ∞ as $\cos\theta$ tends to f_p/f and also as $\cos\theta$ tends to zero. This behaviour is clearly shown by the three curves for which $f > f_p$ in fig. 12.10. Between these values there is a minimum value of D, and for any greater value of D wave packets can travel from the transmitter to the receiver by two different paths. The path for which θ is the smaller is known as the Pedersen ray (Pedersen, 1927, § XI 2c). It is shown in § 12.8 that this ray has the greater time of travel.

Fig. 12.10. Shows how the horizontal range D depends on angle of incidence θ for a typical parabolic layer. In this example $a = \frac{1}{5}h_0$. The numbers by the curves are the values of f/f_p.



When D has a minimum value, this is an example of a skip distance already discussed in § 10.22. No ray that goes via the ionosphere can reach the ground at a shorter range. But near the skip distance the simple ray theory is inadequate. A caustic surface meets the ground there and the dependence of signal on range is given by an Airy integral function (10.69). The scale of this function is given by (10.70), and uses the curvature $\mathrm{d}^2D/\mathrm{d}\theta^2$ of the $D(\theta)$ curve at its minimum.

12.8. Equivalent path P' at oblique incidence

For a wave packet travelling from the origin to the point $(x, 0, z)$, the equivalent path is given by (10.11) which, with $S_2 = 0$, $y = 0$, $S_1 = S$, becomes

$$P' = Sx + \int_0^z \frac{\partial(fq)}{\partial f}\mathrm{d}z. \tag{12.52}$$

This gives the time of travel $t = P'/c$. Now for the isotropic collisionless plasma, (4.10) and (10.12) show that

$$q^2 = C^2 - f_N^2/f^2 \tag{12.53}$$

from which (12.52) gives

$$P' = Sx + C^2 \int_0^z \frac{\mathrm{d}z}{q}. \tag{12.54}$$

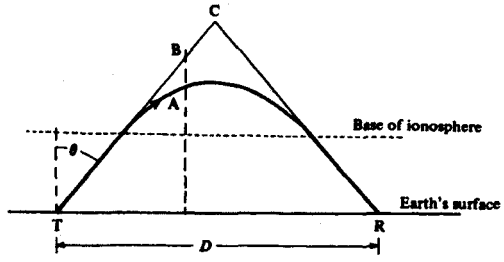
Now (12.44) shows that the integral is simply x/S. Hence

$$P' = x/S \tag{12.55}$$

which means that when a wave packet starts out at an angle θ to the vertical and has travelled along some path TA, fig. 12.11, its equivalent path is TB obtained by extending its initial path along a straight line to a point B vertically above A. In particular if the wave packet returns to the ground at a distance $x = D$, the equivalent path is the sum of the two oblique sides TC, CR of the triangle formed by extending the initial and final directions of the wave packet until they intersect at C. Hence

$$P' = D/S. \tag{12.56}$$

Fig. 12.11. Ray path in the ionosphere and the associated triangulated path.



This is Breit and Tuve's theorem (1926). The fictitious path formed by the two straight lines TC, CR is called the triangulated path. The theorem is true only when the earth's magnetic field and curvature are neglected and when the electron collision frequency is very small.

An alternative proof of Breit and Tuve's theorem is based on the result that for an isotropic plasma the ray and the wave normal have the same direction. Hence, if ψ is the inclination of the ray and wave normal to the vertical, Snell's law gives $n \sin\psi = S$. Let δs be an element of the ray path and δx its horizontal projection. Then $\delta x = \delta s \sin\psi = S\delta s/n = S\delta s n'$ from (5.66). The wave packet travels with the group velocity $\mathscr{U} = c/n'$, and hence $n'\delta s = c\delta t$ where δt is the time taken to travel the distance δs. Hence $\delta t = \delta x/Sc$ which on integration gives $P' = x/S$, the same as (12.55).

Equations (12.53), (12.54) may be combined to give

$$P'(f) = \frac{1}{C}\int_0^z \frac{\mathrm{d}z}{\{1 - f_N^2/(fC)^2\}^{\frac{1}{2}}} \tag{12.57}$$

where P' is measured at the particular frequency f. Now the integrand is the group refractive index $n'(fC)$ for the frequency fC. A wave of frequency fC vertically incident on the ionosphere would be reflected at the level z_0 where $f_N = fC$ and this is also the level of reflection of the frequency f at oblique incidence. Suppose that the vertically incident wave packet and the wave packet described by (12.57) both return to the ground. Then the upper limit of the integral must be set equal to z_0 and a factor 2 introduced to allow for the upward and downward paths. This gives

$$P'(f) = \frac{2}{C}\int_0^{z_0} n'(fC)\,\mathrm{d}z. \tag{12.58}$$

Here the integral is the equivalent height of reflection $h'(fC)$, (12.2), for a frequency fC at vertical incidence. Hence

$$CP'(f) = 2h'(fC). \tag{12.59}$$

In words: 'For wave packets that return to the ground, the equivalent path for a frequency fC at vertical incidence is the vertical projection of the triangulated path for a frequency f at oblique incidence.' This is Martyn's theorem for equivalent path (Martyn, 1935). It must not be confused with Martyn's theorem for attenuation, § 12.11. Some modifications of the theorem, to allow for the earth's curvature, were discussed by Millington (1938b).

The ionosonde technique, § 1.6, uses vertically incident pulses of radio waves for probing the ionosphere and supplies curves of equivalent height $h'(f)$ versus frequency f. It is sometimes used also at oblique incidence, and supplies curves of equivalent path $P'(f)$ versus frequency. It is therefore of interest to calculate $P'(f)$ curves for various models of the ionosphere. For typical observed $P'(f)$ curves at

oblique incidence see for example Wieder (1955), Agy and Davies (1959), Davies (1969).

The range D where a wave packet returns to the ground is found from (12.44) by putting $z = z_0$, the height of reflection, and doubling the right-hand side, as explained in § 12.7. This gives D as a function of the angle of incidence θ. Now $\sin\theta = D/P'$ from (12.56), and on substitution for θ this gives an equation for finding P' when D and f are known.

Consider again the simple model ionosphere of (12.46) with free space below $z = h_0$ and a linear $N(z)$ above this. Then D is given by (12.48) and from (12.56) the required equation is

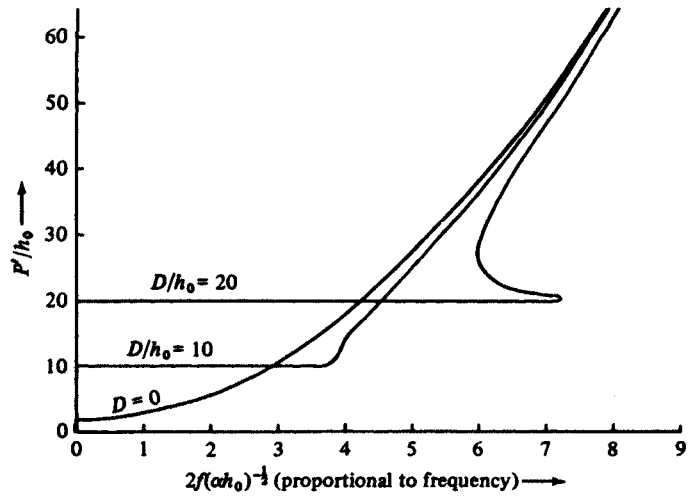
$$\frac{1}{h_0}(P'^2 - D^2)^{\frac{1}{2}} = 2 + \frac{4f^2}{\alpha h_0}\{1 - (D/P')^2\} \tag{12.60}$$

which may be used to calculate f for various values of P'. Some typical curves are shown in fig. 12.12. It was mentioned in § 12.7 that if $D > 6h_0\sqrt{3}$, three different rays can reach the receiver. This property can be seen in fig. 12.12.

For the parabolic model ionosphere of (12.49), D is given by (12.51). From (12.56) $\sin\theta$ is found and substituted in (12.51) to give an equation relating f/f_p and P'. It is a bit complicated because of the logarithm but can be solved quite quickly on a pocket calculator by using an iterative process. Some typical results are shown in fig. 12.13. The curve for $D = 0$ is the same as for vertical incidence, fig. 12.3, and is included for comparison.

For all the curves, $P' \to \infty$ at the penetration frequency $f = f_p$. For a range of

Fig. 12.12. Typical $P'(f)$ curves for oblique incidence and the same model ionosphere as in fig. 12.9. D is the horizontal range from transmitter to receiver.



frequency greater than f_p there are two different values of P', which shows that signals arrive at the receiver by two different ray paths. The ray with the longer path and the smaller θ is the 'Pedersen ray' already mentioned in § 12.7. At the maximum frequency of this range, the two curves of P' move together. This frequency is called the 'maximum usable frequency', or by its abbreviation 'MUF', and it is here denoted by f_M. It is where the range D is at the skip distance. It is studied in the following section.

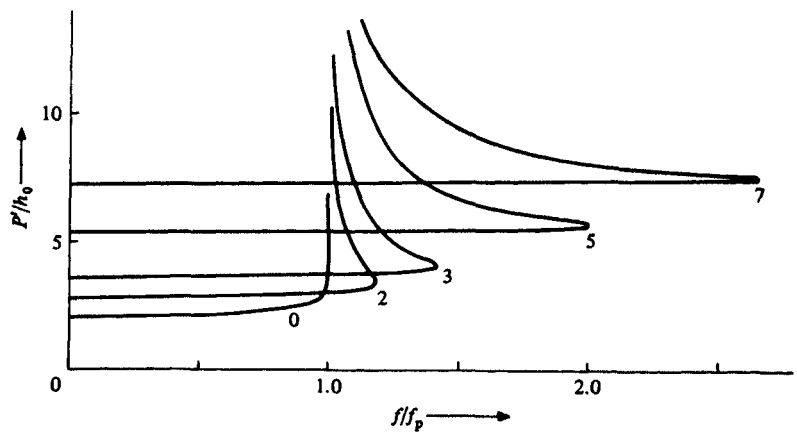
It was shown in § 10.22 that, for frequencies near the MUF, the dependence of signal amplitude on frequency is given by an Airy integral function (10.69), (10.72), and shows an interference effect. But this occurs for a very small range of $f - f_M$ and is not usually seen on oblique incidence ionograms. It is sometimes found that appreciable signals are received at frequencies well beyond the point where the two branches of the $P'(f)$ curve come together, that is at the extreme right of the curves in fig. 12.13. These cannot be explained in terms of wave propagation near the caustic at the skip distance. They might possibly be explained by scattering of the waves from irregularities of N, either at the reflecting level or at some lower level through which the waves pass.

12.9. Maximum usable frequency, MUF

As a simple illustration of the calculation of an MUF, consider the result (12.51) for the parabolic model ionosphere when the earth's curvature is neglected. It gives the range D for a ray that leaves the transmitter at an angle θ to the vertical and returns to the ground. If f is the MUF for a receiver at range D, then D is the skip distance, so that when θ is varied in (12.51), D is a minimum. The condition $\partial D/\partial\theta = 0$ gives

$$2h_0 + a\cos^3\theta \frac{f}{f_p} \ln\left(\frac{f_p + f\cos\theta}{f_p - f\cos\theta}\right) = \frac{2af^2 \sin^2\theta \cos^2\theta}{f_p^2 - f^2\cos^2\theta}. \tag{12.61}$$

Fig. 12.13. As for fig. 12.12 but for the parabolic model ionosphere used in fig. 12.10, with $a = \frac{1}{5}h_0$. The numbers by the curves are the values of D/h_0.



If θ were eliminated from (12.51), (12.61) it would give an equation relating the maximum usable frequency $f = f_M$ and the range D. This, however, is algebraically complicated. Appleton and Beynon (1940) first assumed a value for $\cos\theta\, f/f_p$ and used it in (12.61) to calculate θ. Thus values of f/f_p and θ are found that satisfy (12.61). These are then used in (12.51) to find D. In this way a curve of D versus $f = f_M$ is plotted.

12.10 The forecasting of MUF

In planning the best use of a given radio communication link, it is important to know in advance what frequencies can be used. We therefore need to know the maximum usable frequency. In forecasting this there are two stages. First, some information is needed about the electron height distribution $N(z)$. This depends on time of day, time of year, position on the earth and sunspot number. A possible method for this is to use the International Reference Ionosphere; see § 1.8. Second, this information is used to estimate the MUF for given positions of the transmitter and receiver. Two methods of doing this have been in use for many years, the method of Appleton and Beynon (1940, 1947), and the method of N. Smith (1937, 1938, 1939, 1941). The problem is complicated so that approximations are necessarily used. In the original forms of both methods the earth's magnetic field was neglected, but corrections that allow for it have been used with later versions. The original forms, however, illustrate well the basic principles used in forecasting MUF and a brief summary is given here. For fuller accounts see the original papers cited above and Millington (1938b), Davies (1969, 1981). More recent methods are described in reports of the CCIR (see end of this section). The author is greatly indebted to Dr P.A. Bradley of the Rutherford Appleton Laboratory, Chilton, England, for information on this subject.

The distance D_G between transmitter and receiver is usually of the order of several thousand kilometres, so that the curvature of the earth cannot be neglected. D_G is measured over the earth's curved surface. It is related to the straight line distance D thus

$$D = 2\rho \sin(D_G/2\rho) \tag{12.62}$$

where ρ is the radius of the earth's surface.

In Appleton and Beynon's (1940, 1947) method the parabolic distribution (12.17) for $N(z)$ is used as an approximation for the forecast $N(z)$. One method of estimating the half thickness a, and the height $h_0 = z_m - a$ of the base, uses the $h'(f)$ curve at vertical incidence and was described in § 12.4. Other methods have been given by Booker and Seaton (1940) and by Ratcliffe (1951). Forecasts of a, z_m and f_p are sometimes available so that a forecast of the complete $N(z)$ is not needed. Because of the earth's curvature, the ray path in the ionosphere must be found by using the Bouger law (10.18) instead of Snell's law (6.4). It leads to an expression for D_G as a function of f/f_p and of the angle i_0 between the ray where it enters the base of the

ionosphere and the vertical. This is analogous to (12.51). It is algebraically complicated, but is simplified by assuming that h_0/ρ is small and neglecting squares and higher powers. Now f is the MUF, so that D_G is the skip distance and therefore $\partial D_G/\partial i_0 = 0$. The resulting equation and the equation for D_G are now combined by the method described in §12.9. In this way curves of various kinds can be plotted. In one form of these, for a given value of a/h_0, a set of curves is plotted of f/f_p versus D_G, each curve for a fixed value of $z_m = h_0 + a$. Separate diagrams are used for each value of a/h_0. Thus when a, h_0 and f_p are known, the value of f/f_p for a given D_G can be read off, and f is the required MUF.

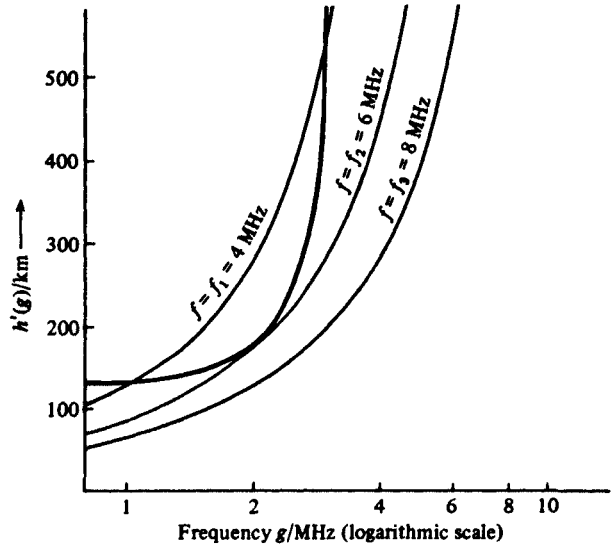
Newbern Smith's (1937, 1938, 1939, 1941) method does not use the assumption that $N(z)$ is the parabolic distribution (12.17). Instead it uses directly the $h'(f)$ curve for vertical incidence. This must first be found for the forecast $N(z)$. The method is described here for the case where the earth's curvature is neglected.

A combination of Breit and Tuve's theorem (12.56) and Martyn's theorem (12.59) gives

$$h'(fC) = \tfrac{1}{2}D(f)\cot\theta \tag{12.63}$$

where $D(f)$ is the range of a ray of frequency f that leaves the transmitter at an angle θ to the vertical, and $h'(fC)$ is the equivalent height of reflection at vertical incidence

Fig. 12.14. Simplified version of Newbern Smith's method for finding the MUF when the earth's curvature is neglected. The thick line is an example of an observed $h'(f)$ curve at vertical incidence, showing penetration at 3 MHz. The thin lines are the right-hand side of (12.64) with $D = 1000$ km. In this example the MUF is 6 MHz.



for a frequency fC. Let $fC = g$. Then (12.63) may be written

$$h'(g) = \tfrac{1}{2}D(f)\frac{g/f}{\{1-(g/f)^2\}^{\frac{1}{2}}}. \tag{12.64}$$

Here g may be called the 'equivalent frequency at vertical incidence'. Now the $h'(g)$ curve is given. Suppose it is plotted as in fig. 12.14. In the same figure let a series of curves be plotted each for a fixed f, and for a chosen value of D, to show how the right-hand side of (12.64), as ordinate, depends on g. Consider the curve for $f = f_1$. It cuts the $h'(g)$ curve at two points, which shows that (12.64) is satisfied for two different frequencies. On the other hand, the curve for the greater frequency $f = f_3$ does not cut the $h'(g)$ curve, so that (12.64) cannot be satisfied. Between them is the curve for an intermediate frequency $f = f_2$, which just touches the $h'(g)$ curve, and therefore f_2 must be the MUF for the chosen D.

This method can be applied more simply as follows. The $h'(g)$ curve is plotted with $\ln g$ as abscissa, and the right-hand side of (12.64) is plotted on a sheet of transparent material with $\ln(g/f)$ as abscissa, so that it has a vertical asymptote where $\ln(g/f) = 0$. The curves are superimposed, so that the horizontal axes coincide, and the transparent sheet is slid horizontally until the two curves just touch. The point where the line $\ln(g/f) = 0$ cuts the g axis of the $h'(g)$ diagram is then noted. At this point $g = f$, so that it gives the MUF, which is simply read off from the horizontal axis. In this way one curve on the transparent sheet can be used for all values of f, but it is necessary to have a different curve for each value of the horizontal range D.

Refinements of this method to allow for the earth's curvature and the earth's magnetic field have been described by Newbern Smith (1938). For a full description of the method see National Bureau of Standards Circular 462 (1948, ch. 6).

When the earth's magnetic field is allowed for, Martyn's theorem and Breit and Tuve's theorem are not valid, and there is no simple method of computing the MUF. It is needed for large values of the range D, so that the earth's curvature cannot be ignored. The problem can be tackled by using general ray tracing methods, ch. 14, and some work on these lines has been done (for example Capon 1962, Herring, 1984). It is necessary first to adopt some model for the ionosphere. In recent work a model that has proved useful is known as the Bradley–Dudeney model (Bradley and Dudeney, 1973; Dudeney, 1978, 1983), and another useful model was proposed by Booker (1977). Then, for a given frequency and bearing, several extraordinary rays are traced, each with a different initial elevation angle, and the horizontal ranges are found. There is, in general, a minimum range or skip distance. The extraordinary ray is used because it gives the smaller skip distance. The calculations are repeated for other frequencies. Then, for a receiving station with the chosen bearing, the MUF is the greatest frequency for which the skip distance is less than the range

of the station. For a receiving station with a different bearing, a new set of calculations is needed.

A program of this kind is extremely demanding even for the fastest modern computers, and the uncertainties in predicting the parameters of the chosen model of the ionosphere are so great that it would not be worth while for regular use in forecasting MUFs.

The standardisation of methods for forecasting MUF is one of the functions of the Comité Consultatif International des Radiocommunications (CCIR) which is a committee of the International Telecommunications Union (ITU). The journal of the ITU is the Telecommunications Journal and is published monthly in several languages. It gives addresses, and has bulletins containing data for forecasting purposes. Details of prediction methods are contained in CCIR reports. Usually these are not published in scientific journals, but can be obtained from the ITU.

12.11. Martyn's theorem for attenuation of radio waves

In this chapter, except in § 12.3, electron collisions have been neglected. If the electron collision frequency ν is small enough, most of the foregoing theory can be used without modification and some new results can be established. When electron collisions are allowed for, q from (4.9) and (12.45) is given by

$$q^2 = C^2 - \frac{X}{1+Z^2} - \frac{\mathrm{i}XZ}{1+Z^2} \tag{12.65}$$

so that q is complex and may be written $q_\mathrm{r} + \mathrm{i}q_\mathrm{i}$ where q_r and q_i are real and q_r is positive. If $Z = \nu/\omega$ is so small that its square and higher powers may be neglected, (12.65) gives

$$q_\mathrm{r}^2 - q_\mathrm{i}^2 \approx C^2 - X, \quad 2q_\mathrm{r}q_\mathrm{i} \approx -\mathrm{i}XZ, \tag{12.66}$$

whence

$$q_\mathrm{r} \approx (C^2 - X)^{\frac{1}{2}}, \quad q_\mathrm{i} \approx -\mathrm{i}\tfrac{1}{2}XZ/q_\mathrm{r}. \tag{12.67}$$

Thus q_r is the same as the value of q used when collisions are neglected. The field in a wave packet is given by (10.4). This now includes an extra factor

$$\exp\left(k\int_0^z q_\mathrm{i}\,\mathrm{d}z\right). \tag{12.68}$$

Since q_i in (12.67) contains a factor Z and is therefore very small, the equations giving the path and time of travel P'/c of a wave packet are not appreciably affected. The factor (12.68) shows that the wave packet is attenuated. When it returns to the ground its amplitude is reduced by a factor $|R|$ where

$$\ln|R| = 2k\int_0^{z_0} q_\mathrm{i}\,\mathrm{d}z. \tag{12.69}$$

When (12.67) for q_i and (3.6), (3.15) for X and Z are used, this gives

$$-\ln|R(f)| = \frac{2\pi}{c}\int_0^{z_0}\frac{f_N^2\nu}{f^2 q_r}\,dz$$

$$= \frac{2\pi C}{c}\int_0^{z_0}\frac{f_N^2\nu}{C^2f^2(1-f_N^2/C^2f^2)^{\frac{1}{2}}}\,dz \tag{12.70}$$

which gives the attenuation for a wave packet of frequency f obliquely incident on the ionosphere at an angle arccos C. If a wave packet of frequency $g = fC$ were vertically incident on the ionosphere, its attenuation would be given by

$$-\ln|R_0(g)| = \frac{2\pi}{c}\int_0^{z_0}\frac{f_N^2\nu}{g^2(1-f_N^2/g^2)^{\frac{1}{2}}}\,dz \tag{12.71}$$

where R_0 denotes the reflection coefficient at vertical incidence. The levels of reflection z_0 are the same in the two cases. Hence

$$\ln|R(f)| = C\ln|R_0(fC)|. \tag{12.72}$$

This is Martyn's theorem for attenuation (Martyn 1935; see also Millington, 1938b; Appleton and Beynon, 1955). It shows that the logarithm of the reflection coefficient for frequency f at oblique incidence is C times the logarithm of the reflection coefficient for frequency fC at vertical incidence. It applies only when the effect of the earth's magnetic field is neglected and when the collision frequency ν is so small that $Z \lesssim 0.1$.

PROBLEMS 12

12.1. In a simple model of the ionosphere there is free space from $z = 0$ to $z = h_0$. Above this the plasma frequency f_N is given by $f_N^2 = \alpha(z - h_0)$. The collision frequency $\nu = 2\pi f Z$ is independent of height. The earth's magnetic field is neglected. Show that for vertically incident waves of frequency f, the phase height, true height of reflection and equivalent height are respectively

$$h = h_0 + \frac{2f^2}{3\alpha}(1 - iZ), \quad z_0 = h_0 + \frac{f^2}{\alpha}(1 - iZ), \quad h' = h_0 + \frac{2f^2}{\alpha}(1 - \frac{2}{3}iZ).$$

Show that the reflection coefficient R satisfies

$$\ln|R| = 2k\,\mathrm{Im}(h) = -\frac{4\nu f^2}{3c\alpha} = -\frac{\nu}{c}\{\mathrm{Re}(h') - \mathrm{Re}(h)\}$$

(compare (12.14)).

12.2. In a model ionosphere the electron concentration increases exponentially with height z above the ground so that $f_N^2 = F^2 e^{\alpha z}$. Collisions are negligible. Show that for vertically incident waves of frequency f, the equivalent height, true height and phase height of reflection are given respectively by

$$h'(f) = \frac{2}{\alpha}\ln\left[\frac{f}{F} + \left\{\left(\frac{f}{F}\right)^2 - 1\right\}^{\frac{1}{2}}\right], \quad z_0 = \frac{2}{\alpha}\ln\frac{f}{F}, \quad h(f) = h'(f) - \frac{2}{\alpha}\left\{1 - \left(\frac{F}{f}\right)^2\right\}^{\frac{1}{2}}.$$

Note that there must be a non-zero electron concentration even at the ground $z = 0$, but for a practical model it is very small so that F/f is an extremely small number. Then

$$h'(f) \approx \frac{2}{\alpha} \ln (2f/F), \quad z_0 = \frac{2}{\alpha} \ln \frac{f}{F}, \quad h(f) \approx \frac{2}{\alpha} \{\ln (2f/F) - 1\}.$$

12.3. In problem 12.2 suppose that the electron collision frequency is independent of height. Show that its effect is to increase the equivalent height of reflection $\mathrm{Re}\{h'(f)\}$ by $Z^2/\alpha + \mathrm{O}(Z^4)$ where $Z = \nu/2\pi f$.

12.4. What electron height distribution functions $N(z)$ would give the following $h'(f)$ curves if electron collisions and the earth's magnetic field are neglected?

(a)

$$h'(f) = h_0 \text{ when } f \leqslant f_1; \; h'(f) = h_0 + \frac{2f}{\alpha}(f^2 - f_1^2) \text{ when } f \geqslant f_1.$$

(b)

$$h'(f) = h_0 + 2f^2/\alpha \qquad \text{when } f \leqslant f_1,$$

$$= h_0 + 2f^2/\alpha + 2f(f^2 - f_1^2)^{\frac{1}{2}}\left(\frac{1}{\beta} - \frac{1}{\alpha}\right) \text{ when } f \geqslant f_1.$$

12.5. Suppose that the refractive index n in the ionosphere is given by $n^2 = 1 - \alpha(z - h_0)$ for $z \geqslant h_0$; $n^2 = 1$ for $z \leqslant h_0$, where α is a real constant, and the earth's magnetic field is neglected. A point transmitter sends up a wave packet at an angle θ to the vertical. Find the distance at which it returns to the earth's surface. (Neglect the earth's curvature.)

The transmitter emits a pulse of radio waves uniformly in all directions. Show that the pulse first returns to the earth at a distance $4h_0(2/h_0\alpha - 1)^{\frac{1}{2}}$ from the transmitter, provided that $\alpha < 2/h_0$. (From Math. Tripos, 1957, Part III.)

12.6. A radio wave packet with predominant frequency f leaves the ground at an angle θ to the vertical. It is reflected from the ionosphere and returns to the ground at distance D, taking a time $P'(f)/c$. There is free space up to height $z = h_0$ and above this the electron concentration is proportional to $(z - h_0)^2$. Electron collisions and the earth's curvature and magnetic field are to be neglected. Prove that

$$P'(f) \sin\theta = D = 2h_0 \tan\theta + Kf \sin\theta \text{ where } K \text{ is a constant.}$$

(Adapted from Math. Tripos 1958, Part III.)

12.7. The ionosphere is a very thin layer, for example a sporadic E-layer ($\mathrm{E_S}$), at height h_0, with penetration frequency f_p. Collisions and the earth's magnetic field are to be neglected.

(a) If the earth's curvature is neglected and if the receiver is at range D from the transmitter, show that the maximum usable frequency is $f_\mathrm{M} = f_\mathrm{p}(1 + D^2/4h_0^2)^{\frac{1}{2}}$.

(b) If the earth's surface has radius ρ, and if the range is D_G measured over the curved surface, show that

$$f_M = f_p\{(h_0+\rho)^2+\rho^2-2\rho(\rho+h_0)\cos\Phi\}^{\frac{1}{2}}/(h_0+\rho-\rho\cos\Phi)$$

where $\Phi = \frac{1}{2}D_G/\rho$. Since Φ is small, show that this leads to

$$f_M = f_p\left(1+\frac{D_G^2}{4h_0^2}-\frac{D_G^2}{4\rho h_0}\right)^{\frac{1}{2}}.$$

(For the solution see Appleton and Beynon, 1940, § 4).