Cambridge University Press
978-0-521-11439-4 - Radio Waves in the Ionosphere: The Mathematical Theory of the Reflection of Radio Waves from Stratified Ionised Layers
K. G. Budden
Excerpt
More information

CHAPTER 1

INTRODUCTION

1.1 The composition of the ionosphere

The ionosphere consists of a number of ionised regions above the earth's surface, which play a most important part in the propagation of radio waves. Our knowledge of it is derived almost entirely from radio measurements, and it is therefore important to understand the processes by which the waves are reflected. The ionosphere is believed to influence radio waves mainly because of the presence of free electrons. The early experiments showed that the electrons must be arranged approximately in horizontally stratified layers, so that the number density is a function only of the height above the earth's surface. The ionosphere must be almost electrically neutral, for if there were any appreciable space charge, it would give rise to large electric forces which would prevent stable layers from forming. There must therefore be at least as many positive ions as electrons, per unit volume. Besides negative electrons there may also be heavy negative ions formed by the attachment of electrons to air molecules. Heavy ions of both signs might play a part in the propagation of radio waves, and this problem is discussed in §§3.9, 5.9 and 6.18. A heavy ion, whether positive or negative, has a mass approximately 60,000 times that of an electron, and it is shown that, for all frequencies above a few hundred cycles, ions must be about 60,000 times more numerous than electrons if they are to have a detectable effect. If this could happen at all, it would only be in the very lowest regions of the ionosphere, but there seems to be no evidence that heavy ions give any observable effect. It is therefore assumed, in most of this book, that only the free electrons can affect radio propagation.

1.2 Plane waves and spherical waves. The curvature of the earth

Radio waves travelling from a transmitter to a receiver near the earth's surface may take one of several possible paths. A wave may travel over the earth's surface, and it is then known as the ground wave. The earth is an imperfectly conducting, curved surface, and the theory of the propagation of the ground wave involves many problems of the greatest mathematical interest, but they are beyond the scope of this book.

Another wave may travel up to the ionosphere, be reflected there, and return to the receiver. It is with a single reflection of this kind that the present book is mainly concerned. The wave originates at a source of small dimensions so that the wave front is approximately spherical, but by the time it reaches the ionosphere the radius of curvature is so large that the wave can be treated as plane. This involves an approximation which is examined in §7.8, and it is shown that the error is negligible except in certain special cases rarely met with in practice. Similarly, the ionospheric layers are curved because of the earth's curvature, but in most problems this curvature can be neglected.

1.3 Effect of collisions and of the earth's magnetic field

The motion of an electron in the ionosphere is affected by the earth's magnetic field and by the collisions which the electron makes with other particles. It was shown by Lorentz that the collisions have the same effect as a retarding force proportional to the velocity. In most of this book the retarding force is included, but it may be neglected in some problems, which arise at high frequencies and can be treated by 'ray theory' methods (chs. 10 to 14). The retarding force or damping force is most important at low frequencies, and here the wavelength is so long that ray-theory methods are inapplicable, and a 'full-wave' treatment must be used. The second half of the book is devoted to this.

The effect of the earth's magnetic field is to make the ionosphere a doubly refracting medium. This is a great complication, and often leads to a differential equation of the fourth order. It is convenient to neglect the earth's magnetic field for many purposes, so that the differential equation to be solved is only of the second order. This is done for considerably more than half of the problems discussed. One reason for this is that only in this way can the differential equations be reduced to a form whose solution has been studied by mathematicians. But by neglecting the earth's magnetic field, principles can be established which can then be extended to more general cases where the field is included.

1.4 Relation to other kinds of wave-propagation

The theory of radio wave-propagation in the ionosphere is closely related to other branches of physics which deal with wave-propagation in media whose properties vary from place to place. For example, in wave-mechanics a study is made of the propagation of electron waves in a potential field. The variation of potential is analogous to the variation of the square of the refractive index for radio waves. But the potential is real in nearly all problems of wave-mechanics, so that there is nothing analogous to the damping forces which, for radio waves, lead to a complex value of the squared refractive index. Moreover, the differential equations in wave-mechanics are nearly always of the second order, which means that there is nothing analogous to double refraction. Some of the material of chs. 9, 10 and 15 is of great importance when applied to wave-mechanics.

In seismology some study has been made of the propagation of elastic waves in media whose properties vary gradually from place to place. For example, some parts of the ocean bed are horizontally stratified and, for sound waves, behave very like an inverted ionosphere. Many of the results of chs. 9, 10, 11 and 15 could be applied to this case.

In solids three kinds of elastic wave can be propagated, two transverse and one longitudinal, so that this medium might be considered to be triply refracting. But seismologists are interested mainly in propagation through homogeneous solids, and in the reflection and transmission which occurs at the sharp boundary between two media. (See, for example, Bullen, *Theory of Seismology*; Jeffreys, *The Earth* (ch. II); Musgrave, 1959.) There appears to have been very little study of propagation of elastic waves through solids whose properties vary continuously from place to place.

It is therefore probable that the theory of wave propagation in continuously variable media has advanced farthest in the field of radio waves in the ionosphere.

1.5 The variation of electron density with height. The Chapman layer

Before the reflecting properties of the ionosphere can be calculated, it is necessary to know how the number density of electrons, N, varies with height above the earth's surface. To study this problem, some assumption must be made about how the ionospheric layers are formed. A most important contribution to this problem was made by Chapman (1931 *a*, *b*, 1939), who derived a law for the variation of N with height, which is now known as the Chapman law. The full theory of the formation of ionospheric layers has been refined and extended by Chapman and others, and is beyond the scope of this book. Only the simplest version of the Chapman theory is given here.

Assume that the earth's atmosphere is constant in composition, and at a constant temperature. Then the air density $\mathfrak{d}$ at height z above the ground is

$$\mathfrak{d} = \mathfrak{d}_0 e^{-z/H}, \tag{1.1}$$

where $\mathfrak{d}_0$ is the density at the ground, $H = RT/Mg$, and R is the gas constant, M the mean molecular weight, T the absolute temperature, and g the gravitational acceleration. The curvature of the earth is neglected, and g is assumed constant. H is called the 'scale height' of the atmosphere, and is approximately 10 km at the ground. The sun's radiation enters the atmosphere at an angle χ from the zenith. Let the mass absorption coefficient of the air for the radiation be σ, and assume that the rate of production of electrons, q, is proportional to the rate of absorption of radiation per unit volume. Let the flux of energy in the incident radiation be I_0 outside the earth's atmosphere, and I at a height z. The energy flux I decreases as the radiation passes down through the atmosphere, and it is clear that

$$dI = I\sigma\mathfrak{d}\sec\chi\, dz. \tag{1.2}$$

This is combined with (1.1) and integrated, which gives

$$I = I_0 \exp\{-\sigma\mathfrak{d}_0 H \sec\chi\, e^{-z/H}\}. \tag{1.3}$$

Now it is convenient to take $z_0 = H\log(\sigma\mathfrak{d}_0 H)$ so that (1.3) becomes

$$I = I_0 \exp\left[-\sec\chi \exp\left\{-\frac{z-z_0}{H}\right\}\right]. \tag{1.4}$$

The rate of absorption of energy at height z is $\cos\chi(dI/dz)$, and since this is proportional to the rate of production of electrons q, we have from (1.4)

$$q = q_0 \exp\left[1 - \frac{z-z_0}{H} - \sec\chi \exp\left\{-\frac{z-z_0}{H}\right\}\right], \tag{1.5}$$

where q_0 is a constant, namely I_0/eH (here e is the exponential). q has the maximum value $q_0 \cos\chi$ when $(z-z_0)/H = \log(\cos\chi)$, so that q_0 is the maximum rate of electron production when $\chi = 0$.

Next it is necessary to consider how electrons are removed. It is now believed that a number of processes contribute to the removal, but the effect is the same as if the electrons simply recombined† with positive ions. Assume that the only ions present are electrons and positive ions, and that the number of each per unit volume is N. Then the rate of removal of electrons is αN^2, where α is a constant called the recombination coefficient. The variation of N with time t is then given by

$$\frac{dN}{dt} = q - \alpha N^2. \tag{1.6}$$

If α is large enough, the term dN/dt can be neglected. When this happens, the processes of formation and removal of electrons come into equilibrium in a negligibly small time, and we then have

$$N = (q/\alpha)^{\frac{1}{2}}. \tag{1.7}$$

---

† This may not be true for the F-layer where it is believed that electrons are removed according to the attachment law $dN/dt = q - \beta N$, and the constant β is called the attachment coefficient.

Cambridge University Press
978-0-521-11439-4 - Radio Waves in the Ionosphere: The Mathematical Theory of the Reflection of Radio Waves from Stratified Ionised Layers
K. G. Budden
Excerpt
More information

This is a very much oversimplified picture of the actual mechanism of formation and removal of electrons in the ionosphere, but it does at least give a general guide as to how the electron density might vary with height. Combination of (1.7) with (1.5) gives

$$N = N_0 \exp \frac{1}{2}\left[1 - \frac{z-z_0}{H} - \sec\chi \exp\left\{-\frac{z-z_0}{H}\right\}\right]. \qquad (1.8)$$

This expression will be called the 'Chapman law'. It assumes that the recombination coefficient α is independent of height. In Fig. 1.1 the expression (1.8) for N is plotted against the height z for various values of χ. It is seen that N has a maximum value $N_m = N_0(\cos\chi)^{\frac{1}{2}}$ at the level $z_m = z_0 + H\log(\sec\chi)$. It falls off quite steeply below this, and less steeply above it.

An alternative form of (1.8) is obtained by taking

$$\zeta = \frac{z-z_m}{H} = \frac{z-z_0}{H} - \log(\sec\chi), \qquad (1.9)$$

so that ζH is height measured from the level of maximum N. Then

$$N = N_m \exp \tfrac{1}{2}(1 - \zeta - e^{-\zeta}), \qquad (1.10)$$

which shows that the 'shape' of a Chapman layer is independent of the sun's zenith angle χ. The curvature of the $N(z)$ profile at the maximum is $N_m/2H^2$. This is the same as the curvature at the apex of a parabolic profile (see § 10.7) whose 'half thickness' is $2H$.

1.6 Approximations to the electron density profile

If N is assumed to vary with height z according to the Chapman law (1.8), then this expression would appear in the differential equation which has to be solved to find the reflection coefficient of the ionosphere. But such a differential equation would be so complicated that it could only be solved by numerical methods. Such methods have been used extensively, and are discussed in ch. 22. It is also useful, however, to select small ranges of z, and use approximate and simpler expressions for the electron density. This permits the differential equations to be reduced to simple standard forms whose solutions have well-known properties. For example, it is often possible to choose a range of z so small that the variation of N may be assumed to be linear. This case is of the greatest importance, and is the subject of chs. 15 and 16. Near the maximum values of N in Fig. 1.1, the linear law is not satisfactory, but the profile can be treated as a parabola. In the lower part of a Chapman layer it is often useful to treat the variation as approximately exponential over a small range. Other laws of variation of N with z are discussed because of their mathematical interest. An especially important case is the homogeneous medium with a sharp lower boundary, which is discussed in ch. 8, and which, for very low frequencies, may be a fair approximation to the true ionosphere.

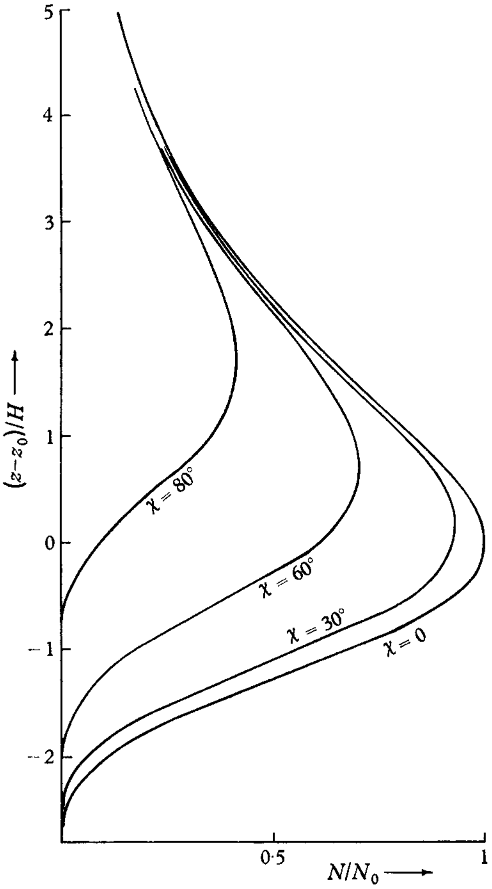

Fig. 1.1. Curves showing how the electron number density N varies with height z according to the simple Chapman theory for a flat earth, for various values of the sun's zenith angle χ.

1.7 The variation of collision-frequency with height

The average number of collisions ν which an electron makes per unit time with the air molecules depends upon the number density of the molecules, and therefore on the density and composition of the air. It also depends on the velocity of the electron, but for many purposes it is permissible to neglect this effect. Then, in an atmosphere which is constant in composition and temperature:

$$\nu = \nu_0 \exp(-z/H), \tag{1.11}$$

where H is the scale height defined on p. 4, and ν_0 is constant. In practice H takes different values at different levels, and the law can only be expected to hold over ranges of z so small that H may be treated as constant. A useful summary of the factors which affect the value of ν have been made by Nicolet (1953). Fig. 1.2 shows how ν depends on the height z according to the best estimates at present available.

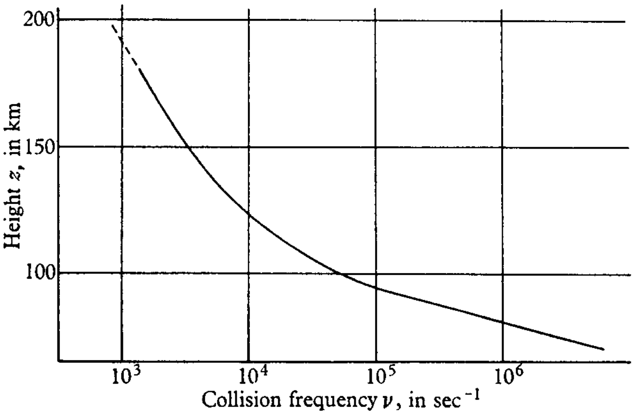

Fig. 1.2. The dependence of electron collision-frequency ν upon height z. This curve is based partly on the work of Crompton, Huxley and Sutton (1953). The author is greatly indebted to Dr K. Weekes for supplying the data from which it was plotted.

It is found that changes of the value of ν affect the propagation of radio waves far less than changes of the electron number density N. For many purposes it is therefore permissible to treat ν as constant over a small range of height z. This is especially true at high frequencies (greater than about 1 Mc/s), where the wavelength is small compared with the scale height H, which is about 10 km.

The dependence of ν on electron velocity gives rise to the phenomena of wave-interaction, which is beyond the scope of this book. Here ν is treated as a constant at each level.

1.8 The structure of the ionosphere

A great deal of information has been accumulated on the detailed structure of the ionosphere, but the topic is beyond the scope of this book, and only a brief outline is given here of the major features. For further details see Appleton (1935) and Rawer (1953).

The ionosphere consists of two main layers known as *E* and *F*. The *E*-layer has its maximum electron density at a height of about 110 km. Radio measurements have been used to study its height and penetration-frequency and their variation with time of day and season. The electron density just above the maximum of the *E*-layer is not easy to investigate, but it is probable that it is only slightly less than the maximum value for a range of height extending right up to the base of the *F*-layer. Thus the *E*- and *F*-layers are not really distinct. For the purposes of this book, however, it will often be sufficiently accurate to assume that during daylight hours the *E*-layer is a Chapman layer given by (1.8) with a scale height $H = 10$ km, $z_0 = 115$ km, and $N_0 = 2{\cdot}8 \times 10^5\,\text{cm}^{-3}$. This means that the penetration-frequency (see for example, § 10.7) is $4{\cdot}7 \times (\cos\chi)^{\frac{1}{4}}$ Mc/s. The actual behaviour of the *E*-layer is very much more complicated than this. At night the *E*-layer has a penetration-frequency of the order of 0·5 Mc/s, corresponding to an electron density of $3000\,\text{cm}^{-3}$. Much less is known about its structure at night.

The *F*-layer is more heavily ionised, with its maximum of electron density in the range 200–400 km. Its diurnal and seasonal variations are more complex than those of the *E*-layer. During daylight the curve of electron number density versus height, for the *F*-layer, often shows a subsidiary bulge below the maximum of electron density (see Fig. 1.3). This is known as the F_1-layer, and it may occasionally attain an actual maximum. The main maximum above it is known as the F_2-layer. There is now some evidence to show that the formation of the whole *F*-layer can be explained by a single ionising agency which would form a simple layer like a Chapman layer if the attachment coefficient β (see footnote to § 1.5) were constant for all heights. Bradbury (1937, 1938) has put forward the hypothesis that β decreases as the height z increases, so that the maximum of ionisation in the F_2-layer arises, not from a fast rate of production, but because of a slow rate of removal of electrons. In temperate latitudes the penetration frequency of the *F*-layer ranges from about 2 Mc/s at night, to 8 Mc/s in a summer day. These correspond to electron densities $5 \times 10^4\,\text{cm}^{-3}$ and $8 \times 10^5\,\text{cm}^{-3}$ respectively.

In chs. 10 and 12 some account is given of methods of finding the function $N(z)$ in the *F*-layer from radio observations. Work of this kind shows that, at the maximum, the curvature of the $N(z)$ curve is about the same as that at the apex of a parabolic layer of half-thickness 100 km, or that of a Chapman layer of scale height about 50 km. The

F-layer is therefore much thicker than the E-layer, as well as more heavily ionised.

There is some evidence that in the daytime there may be another layer lower down with its maximum of electron density near 80 km. This has been called the 'D-layer'. The evidence is not strong, and it is possible that even though $N(z)$ is enhanced below the E-layer, it is still a monotonically increasing function, as shown in Fig. 1.3. It is better to use the term 'D-region' to mean the part of the ionosphere below about 90 km.

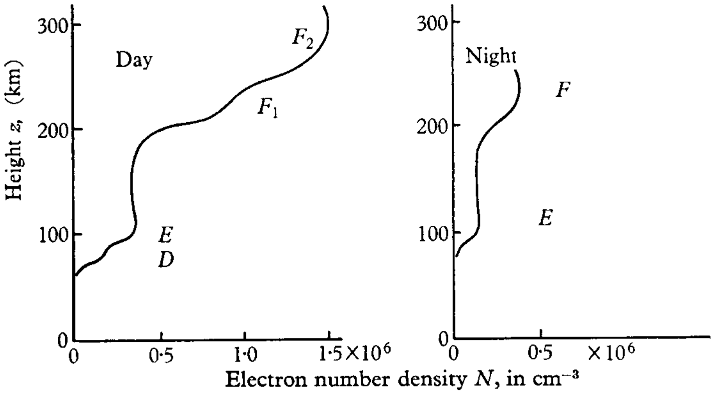

Fig. 1.3. The curves show very roughly how the electron number density N is believed to depend on height z. Actual electron density profiles vary over wide ranges and depend markedly on time of day, season, sun-spot number and whether or not the ionosphere is disturbed.

The experimental study of this region comes largely from radio observations at very low frequencies, and here the wavelength is so long that interpretation of the observations is much less direct than at high frequencies. One of the most important features of the full wave theory given in this book is that it may help to disentangle the numerous radio observations at very low frequencies. Here the procedure is to assume some profile for the electron density and work out its reflecting properties. If these do not agree with observations, some other profile must be tried until a satisfactory result is obtained. The process is difficult because of the complexity of the mathematics and the variability of the radio observations with time of day and season, and with ionospheric disturbances. Some more direct information about the structure of the D-region has been provided recently through the work of Gardner and Pawsey (1953) and Fejer and Vice (1959).

1.9 Horizontal variations and irregularities

Many of the results of radio observations can be explained by assuming that the ionosphere is horizontally stratified, that is, that the electron density and collision frequency are functions only of the height z. This assumption is adopted throughout this book.

Recent experimental work has been shown that there must also be horizontal variations of electron density in the ionosphere. These are often irregular, and are subject both to steady movements and random change. The irregularities permit measurements to be made of steady motion (winds) in the ionosphere. For reviews of these topics, and a bibliography, see Briggs and Spencer (1954), Ratcliffe (1956).