12 Propagation

IT HAS been said that, without an ionosphere, wireless telegraphy might have remained an interesting but commercially unrewarding experiment in physics. Yet, would you believe, there was no ionosphere before 1932?

That has to be a trick question, of course. For the truth is that Balfour Stewart deduced from a study of geomagnetic storm data in 1878 (which was several years before Hertz began his experiments with wireless transmission) that there had to be an electrically conducting layer high in the Earth's atmosphere.

In 1901, Marconi took a chance on it being so, and was rewarded for his faith by the achievement of the transatlantic 'first'. In the 'twenties, Appleton devised experiments to prove conclusively that such a layer existed; in fact, eventually he found more than one. And in 1932 Watson-Watt gave the region a general name - he called it the *ionosphere*.

Today, many people, including some who should know better, seem to think that we know all that we need to know about the ionosphere. But it preserves many a secret yet and there is still a place for careful experimentation by amateur operators.

This chapter can do no more than scratch the surface of a very wide-ranging topic. No attempt has been made in it to introduce divisions between LF, MF, HF, VHF and UHF because there are no such clear-cut divisions in the real world.

The story should really begin with a section about the Sun, because it is there that most of the direct influences on our signal paths have their origin. However, to appreciate that, we must first look at a few of the characteristics of our own atmosphere, and then run over some basic facts and figures.

Don't be put off by the appearance of the mathematics. There is nothing very difficult here and a simple scientific calculator or spreadsheet can handle the tasks with ease.

As you read, compare the text with your own experiences on the bands. In all probability you know already much of the 'whats', 'wheres' and 'whens' of radio propagation. With any luck, this chapter will provide you with the associated 'whys' and 'hows'. Be warned, though. This subject is strongly addictive and there is room for more up there among the 'whos'.

THE EARTH'S ATMOSPHERE

There is a certain amount of confusion surrounding any attempt to identify various portions of the Earth's atmosphere, and this has come about as a result of there being no obvious natural boundaries as there are between land and sea. Workers in different disciplines have different ideas about which functions ought to be separated and it is particularly awkward for the radio engineer that he has to deal with the *troposphere* and *stratosphere*, which are terms from one set of divisions, and the *ionosphere*, which is from another.

Fig 12.1 shows the nomenclature favoured by meteorologists and now widely accepted among physicists generally. It is based on temperature variations, as might be expected from its origin. Note, incidentally, that the height scale of kilometres is a logarithmic one, and that the atmospheric pressure has been scaled in hectopascals (hPa), which are now replacing millibars in scientific work by international agreement. However, the two units are identical in magnitude; only the name has been changed, so the numbers remain the same. See *Weather* (Royal Meteorological Society) May 1986, p172, for an explanation.

In the *troposphere*, the part of the atmosphere nearest the ground, temperature tends to fall off with height. At the *tropopause*, around 10km (although all these heights vary from day to day and from place to place), it becomes fairly uniform at first and then begins to increase again in the region known as the *stratosphere*.

This trend reverses at the *stratopause*, at an altitude of about 50km, to reach another minimum at the *mesopause* (ca 80km) after traversing the *mesosphere*. Above the mesopause temperatures begin to rise again in the *thermosphere*, soon surpassing anything encountered at lower altitudes, and levelling off at about +1200°C around 700km, where we must leave it in this survey.

The *ionosphere*, which has been defined as "the region above the Earth's surface in which ionisation takes place, with diurnal and annual variations which are regularly associated with ultraviolet radiation from the Sun, and sporadic variations arising from hydrogen bursts from sunspots" (*Chambers' Technical Dictionary*), overlaps the thermosphere, mesosphere and part of the stratosphere, but for practical purposes may be considered as lying between 60 and 700km. The name 'ionosphere' is perhaps misleading because it is the number of free electrons,

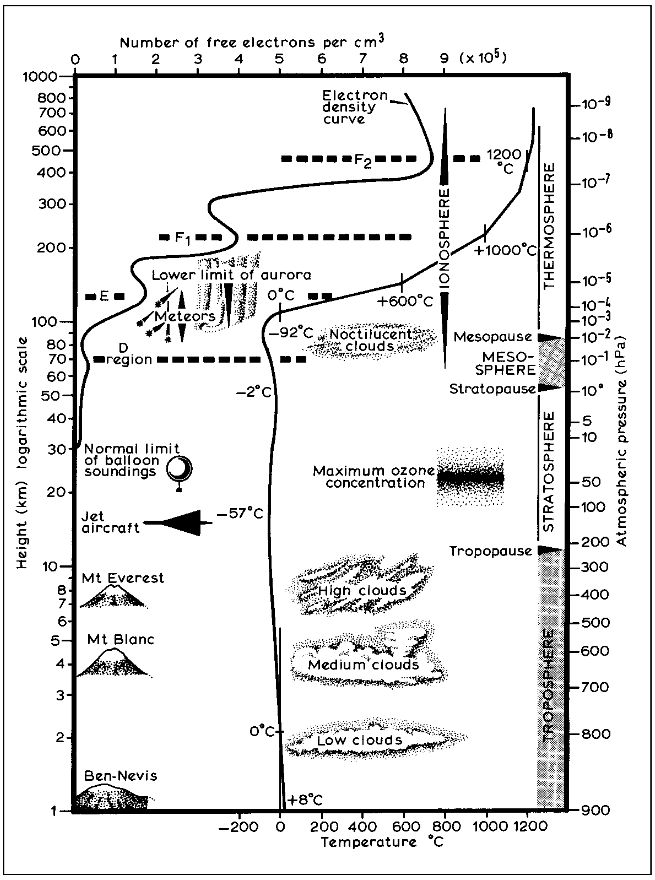


Fig 12.1: Some features of the Earth's atmosphere. The height scale is logarithmic, beginning at 1km above sea level. The equivalent pressure scale on the right is not regularly spaced because the relationship between pressure and height depends on temperature, which does not change uniformly with height

rather than the ions they have left behind, which principally determines the electrical properties of the region. The electron density curve in the diagram shows a number of 'ledges', identified by the letters D, E, F1 and F2, which are the concentrations of free electrons described as 'layers' later on, when dealing with propagation in the ionosphere.

They tend to act as mirrors to transmissions of certain wavelengths, while allowing others to pass through. The lowest layer is generally referred to as the *D-region* rather than the *D-layer* because, as we shall see, its principal role is one of absorption rather than reflection, and its presence is usually easier to infer than it is to observe, except at the lowest frequencies (LF).

It may be found helpful to refer to this diagram again when features of the ionosphere and troposphere are dealt with in later sections of this chapter.

FUNDAMENTAL CONSIDERATIONS

Radiation

The transmitted signal may be regarded as a succession of concentric spheres of ever-increasing radius, each one a unit of one wavelength apart, formed by forces moving outwards from the antenna. At great distances, these wavefronts, appear as plane surfaces as they approach the observer.

There are two inseparable fields associated with the transmitted signal, an *electric field* due to voltage changes and a *magnetic field* due to current changes, and these always remain at right-angles to one another and to the direction of propagation as the wave proceeds.

They always oscillate in phase and the ratio of their amplitudes remains constant. The lines of force in the electric field run in the plane of the transmitting antenna in the same way as would longitude lines on a globe having the antenna along its axis. The electric field is measured by the change of potential per unit distance, and this value is referred to as the *field strength*.

The two fields are constantly changing in magnitude and reverse in direction with every half-cycle of the transmitted carrier. As shown in **Fig 12.2**, successive wavefronts passing a suitably placed second antenna induce in it a received signal which follows all the changes carried by the field and therefore reproduces the character of the transmitted signal.

By convention the direction of the electric lines of force defines the direction of *polarisation* of the radio waves. Thus horizontal dipoles propagate horizontally polarised waves and vertical dipoles propagate vertically polarised waves. In free space, remote from ground effects and the influence of the Earth's atmosphere, these senses remain constant and a suitably aligned receiving antenna would respond to the whole of the incident field.

When the advancing wavefront encounters the surface of the Earth or becomes deflected by certain layers in the atmosphere, a degree of cross-polarisation may be introduced which results in signals arriving at the receiving antenna with both horizontal and vertical components present.

Circularly polarised signals, which contain equal components of both horizontal and vertical polarisation, are receivable on dipoles having any alignment in the plane of the wavefront, but the magnitude of the received signal will be only half that which would result from the use of an antenna correctly designed for such a form of polarisation (it must not be overlooked that there are two forms, differing only in their direction of rotation). Matters such as these are dealt with in detail in the chapter on VHF/UHF antennas.

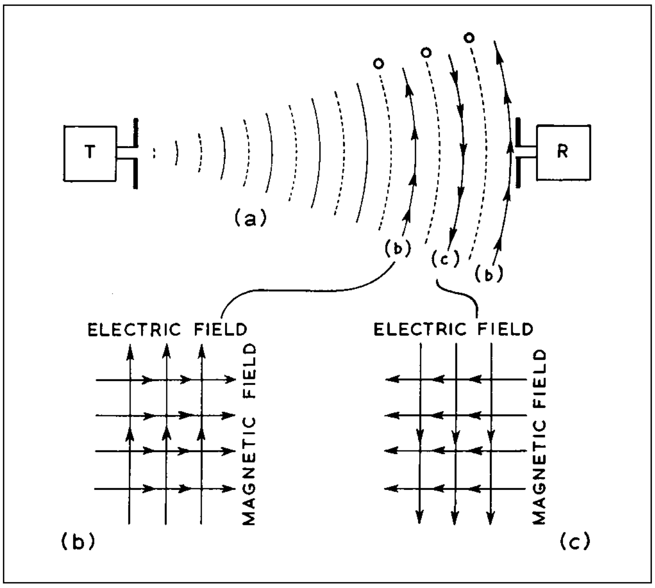


Fig 12.2: (a) A vertical section through the fields radiated from a vertically polarised transmitting antenna. The expanding spherical wavefront consists of alternate reversals of electric field, shown by the arrowed arcs; nulls are shown dotted. At right angles to the plane of the paper, but not seen here, would be simultaneous alternate reversals of magnetic field (b) and (c). These two 'snapshots' must be rotated through 90° in imagination, so that the magnetic field lines run into and out of the plane of the paper. The view is from R towards T

Field Strength

As the energy in the expanding wavefront has to cover an ever-increasing area the further it travels, the amplitude of the signal induced in the receiving antenna diminishes as a function of distance. Under free-space conditions an inverse square-law relationship applies, but in most cases the nature of the intervening medium has a profound, and often very variable, effect on the magnitude of the received signal.

The intensity of a radio wave at any point in space may be expressed in terms of the strength of its electric field at rightangles to the line of the transmission path. The units used indicate the difference of electric force between two points one metre apart, and **Fig 12.3** (which should not be used for general calculations as it relates only to free-space conditions) has been included here to give an indication of the magnitudes of fields likely to be met with in the amateur service in cases where a more-or-less direct path exists between transmitting and receiving antennas and all other considerations may be ignored. In practice signal levels very much less than the free-space values may be expected because of various losses en route, so that field strengths down to about 1µV/m need to be considered, or in the case of LF, down to the 1nV/m level.

Signal Input

In many calculations dealing with propagation, the parameter of interest is not the field strength at a particular place but the voltage which is induced by it across the input of a receiver. If a half-wave dipole is introduced into a field and aligned for maximum signal pick-up, the open-circuit EMF induced at its centre is given by the expression:

$$e = \frac{E\lambda}{\pi} \qquad (1)$$

where e = EMF at the centre of the dipole, in volts
E = the incident field strength in volts/metre

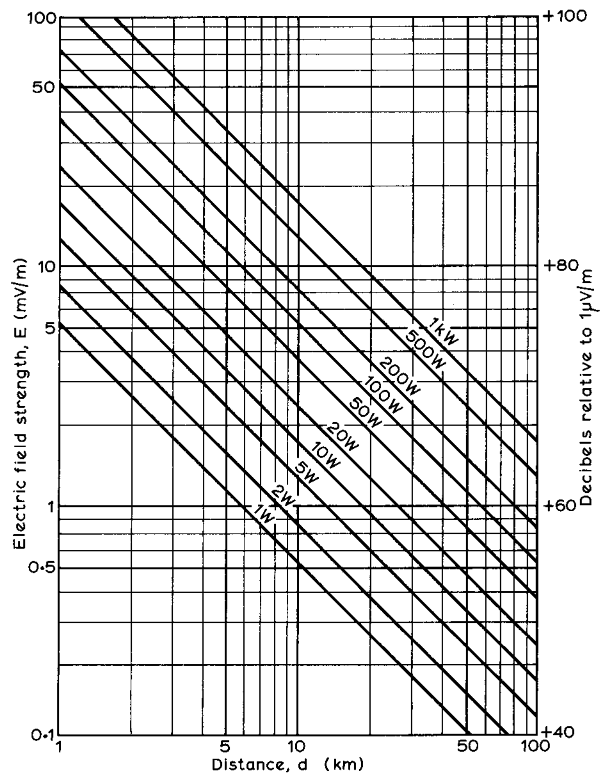


Fig 12.3: Field strength from an omnidirectional antenna radiating from 1kW to 1W into free space. This is an application of the inverse distance law

π = the wavelength of the transmitted signal in metres. When connected to a matched feeder correctly terminated at the receiver the input voltage available will be half this, or $e/2$. By substituting frequency for wavelength the equation may be reduced to the more practical form:

$$V_r = \frac{47.8}{f} E \qquad (2)$$

where V_r = microvolts of signal across the receiver input

f = frequency of the transmitted signal in megahertz

E = the incident field strength in microvolts/metre.

It should be noted that for a given frequency the first term becomes a constant factor. By a further rearrangement:

$$E = \frac{f}{47.8} V_r \qquad (3)$$

which enables the field to be estimated from the magnitude of the received signal in cases where the receiver is fed from a perfectly matched half-wave dipole.

It is an advantage to work in decibels in calculations of this nature. If a standard level of 1µV/m is adopted for field strength and of 1µV for signal level it is a simple matter to take into account the various gains and losses in a practical receiving system. **Fig 12.4** shows the terminated voltage at the receiver (V_0) in terms of decibels relative to 1µV for a normalised incident field strength of 1µV/m at the transmitter frequency, derived from expression (2). Then:

$$V_r(\text{dB}) = V_0(\text{dB}) + V_i(\text{dB}) + G_r(\text{dB}) - L_{fr}(\text{dB}) \qquad (4)$$

where V_r((dB) = input to receiver in decibels relative to 1mV

V_0(dB) = input in decibels relative to 1mV, for 1mV/m incident field strength (from graph)

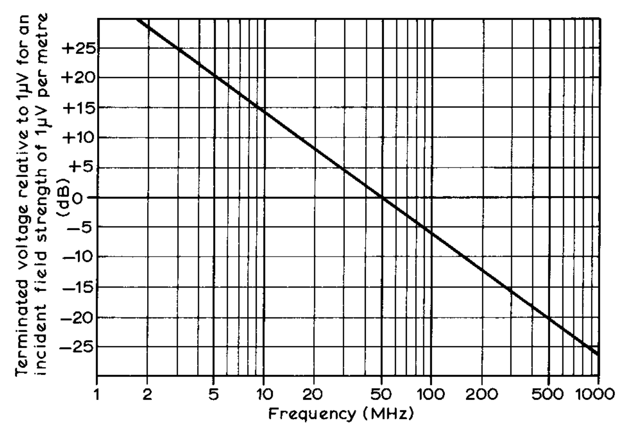


Fig 12.4: The relationship between the incident field strength on a half-wave dipole and the voltage at the receiver end of a correctly matched perfect feeder connected to it

V_i(dB) = the actual incident field strength in decibels relative to 1mV/m

G_r(dB) = the gain in decibels of the receiving antenna relative to a half-wave dipole

L_{fr}(dB) = loss in the feeder line between antenna and receiver

Example. The incident field strength of a 70MHz transmission is 100µV/m (or 40dB above 1µV/m). A three-element Yagi having a gain of 5dB over a half-wave dipole is connected to the receiver through 100ft of coaxial cable which introduces a loss of 2dB. From this, V_0 = -3.5, V_i = +40, G_r = +5 and L_{fr} = +2 (the fact that L_{fr} is a loss is allowed for in Eqn (4)), and V_r = -3.5 + 40 + 5 - 2dB = 39.5dB relative to 1µV, or a receiver input voltage of 94µV.

Table 12.1 may be found useful in converting the rather unfamiliar looking values of decibels met with in radio propagation work into their corresponding voltage or power ratios using only a cheap four-function calculator.

MODES OF PROPAGATION

Introduction

There are four principal modes by which radio waves are propagated. They are:

(a) *Free-space waves*, which are unaffected by any consideration other than distance;

(b) *Ionospheric waves*, which are influenced by the action of free electrons in the upper levels of the Earth's atmosphere;

(c) *Tropospheric waves*, which are subject to deflection in the lower levels by variations in the refractive index structure of the air through which they pass; and

(d) *Ground waves* which are modified by the nature of the terrain over which they travel.

Free-space waves propagate from point-to-point by the most direct path.

Waves in the other three categories are influenced by factors which make them tend to overcome the curvature of the Earth, either by refraction as with ionospheric waves, refraction as with tropospheric waves, or diffraction at the surface of the Earth itself as with ground waves.

Wavelength is the chief consideration which determines the mode of propagation of Earth-based transmissions.

Combine by multiplication Voltage ratios		Combine by addition (dB)	Combine by multiplication Power ratios	
up	down		up	down
1.01×10^{0}	9.89×10^{-1}	0.1	1.02×10^{0}	9.77×10^{-1}
1.02×10^{0}	9.77×10^{-1}	0.2	1.05×10^{0}	9.55×10^{-1}
1.03×10^{0}	9.66×10^{-1}	0.3	1.07×10^{0}	9.33×10^{-1}
1.05×10^{0}	9.55×10^{-1}	0.4	1.10×10^{0}	9.12×10^{-1}
1.06×10^{0}	9.44×10^{-1}	0.5	1.12×10^{0}	8.91×10^{-1}
1.07×10^{0}	9.33×10^{-1}	0.6	1.15×10^{0}	8.71×10^{-1}
1.08×10^{0}	9.23×10^{-1}	0.7	1.17×10^{0}	8.51×10^{-1}
1.10×10^{0}	9.12×10^{-1}	0.8	1.20×10^{0}	8.32×10^{-1}
1.11×10^{0}	9.02×10^{-1}	0.9	1.23×10^{0}	8.13×10^{-1}
1.12×10^{0}	8.91×10^{-1}	1	1.26×10^{0}	7.94×10^{-1}
1.26×10^{0}	7.94×10^{-1}	2	1.58×10^{0}	6.31×10^{-1}
1.41×10^{0}	7.08×10^{-1}	3	2.00×10^{0}	5.01×10^{-1}
1.58×10^{0}	6.31×10^{-1}	4	2.51×10^{0}	3.98×10^{-1}
1.78×10^{0}	5.62×10^{-1}	5	3.16×10^{0}	3.16×10^{-1}
2.00×10^{0}	5.01×10^{-1}	6	3.98×10^{0}	2.51×10^{-1}
2.24×10^{0}	4.47×10^{-1}	7	5.01×10^{0}	1.99×10^{-1}
2.51×10^{0}	3.98×10^{-1}	8	6.31×10^{0}	1.59×10^{-1}
2.82×10^{0}	3.55×10^{-1}	9	7.94×10^{0}	1.26×10^{-1}
3.16×10^{0}	3.16×10^{-1}	10	1.00×10^{1}	1.00×10^{-1}
1.00×10^{1}	1.00×10^{-1}	20	1.00×10^{2}	1.00×10^{-2}
3.16×10^{1}	3.16×10^{-2}	30	1.00×10^{3}	1.00×10^{-3}
1.00×10^{2}	1.00×10^{-2}	40	1.00×10^{4}	1.00×10^{-4}
3.16×10^{2}	3.16×10^{-3}	50	1.00×10^{5}	1.00×10^{-5}
1.00×10^{3}	1.00×10^{-3}	60	1.00×10^{6}	1.00×10^{-6}
3.16×10^{3}	3.16×10^{-4}	70	1.00×10^{7}	1.00×10^{-7}
1.00×10^{4}	1.00×10^{-4}	80	1.00×10^{8}	1.00×10^{-8}
3.16×10^{4}	3.16×10^{-5}	90	1.00×10^{9}	1.00×10^{-9}
1.00×10^{5}	1.00×10^{-5}	100	1.00×10^{10}	1.00×10^{-10}
1.00×10^{10}	1.00×10^{-10}	200	1.00×10^{20}	1.00×10^{-20}

Example: 39.5dB above 1µV (voltage ratio).
39.5dB = 30 + 9 + 0.5dB
Combining equivalents by multiplication
= $(3.16 \times 10^{1}) \times (2.82 \times 10^{0}) \times (1.06 \times 10^{0})$ = 94 times 1µV, or 94µV.

Table 12.1; Skeleton decibel table

The Spectrum of Electromagnetic Waves

The position of man-made radio waves in the electromagnetic wave spectrum is shown in **Fig 12.5**, where they can be seen to occupy an appreciable portion of a family of naturally-occurring radiations, all of which are characterised by inseparable oscillations of electric and magnetic fields and travel with the same velocity in free space. This velocity, 2.99790×10^{8}m/s (generally taken as 3×10^{8}m/s in calculations), is popularly known as *the speed of light* although visible light forms but a minor part of the whole range.

At the long-wavelength end the waves propagate in a manner which is similar in many respects to the way in which sound waves propagate in air, although, of course, the actual mechanism is different. Thus, reports of heavy gunfire in the 1914-18 war heard at abnormal ranges beyond a zone of inaudibility revealed the presence of a sonic skywave which had been reflected by the thermal structure of the atmosphere around 30km in height, and this has a parallel in the reflection of long wavelength radio sky-waves by the atomic structure of the atmosphere around 100km in height, which also leads to a zone of inaudibility at medium ranges.

Radio waves at the other end of the spectrum show characteristics which are shared by the propagation of light waves, from which they differ only in wavelength. For example, millimetre waves, which represent the present frontier of practical technology, suffer attenuation due to scattering and absorption by clouds, fog and water droplets in the atmosphere - the same factors which determine 'visibility' in the meteorological sense.

The radio wave portion of the spectrum has been divided by the International Telecommunication Union into a series of bands based on successive orders of magnitude in wavelength.

In **Table 12.2** an attempt has been made to outline the principal propagation characteristics of each band, but it must be emphasised that there are no clear-cut boundaries to the various effects described.

Wave Propagation in Free Space

The concept of *free-space* propagation, of a transmitter radiating without restraint into an infinite empty surrounding space, has been introduced briefly in the section on field strength where it was used to illustrate, in a general way, the relationship between the strength of the field due to a transmitter and the distance over which the waves have travelled.

It is only recently, with the advent of the Space Age, that we have acquired a practical opportunity to operate long-distance circuits under true free-space conditions as, for example, between spacecraft and orbiting satellites, and it is only recently that radio amateurs have had direct access to paths of that nature. Earth-Moon-Earth contacts are becoming increasingly popular, however, and reception of satellite signals commonplace, and for these the free-space calculations apply with only relatively minor adjustments because such a large part of the transmission paths involved lies beyond the reach of terrestrial influences.

In many cases, and especially where wavelengths of less than about 10m are concerned, the free-space calculations are even applied to paths which are subject to relatively unpredictable perturbations in order to estimate a convenient (and often unobtainable) ideal - a standard for the path - against which the other losses may be compared.

The basic transmission loss in free space is given by the expressions:

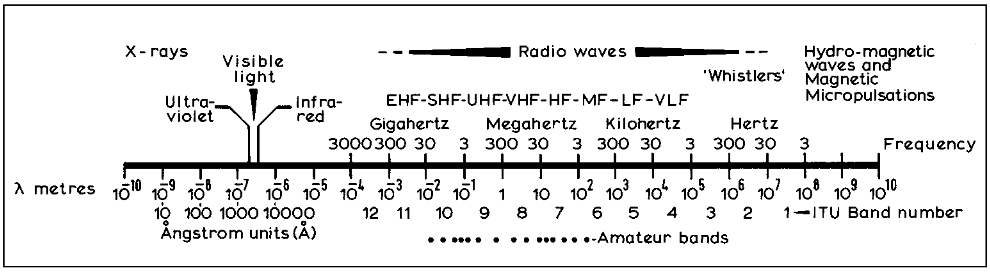


Fig 12.5: The spectrum of electromagnetic waves. This diagram shows on a logarithmic scale the relationship between X-rays, 'visible' and 'invisible' light, heat (infrared), radio waves and the very slow waves associated with geomagnetic pulsations, all of them similar in basic character

ITU Band No	Metric name of band and limits by wavelength	Alternative name of band and limits by frequency	UK amateur bands by frequency (and usual description based on wavelength)	Principal propagation modes	Principal limitations
4	Myriametric 100,000-10,000m	Very low frequency (VLF) 3-30kHz	-	Extensive surface wave Ground to ionosphere space acts as a waveguide	Very high power and very large antennas required. Few channels.
5	Kilometric 10,000-1000m	Low frequency (LF) 30-300kHz	135.7-137.8kHz	Surface wave and reflections from lower ionosphere	High power and large antennas required. Limited number of channels available. Subject to fading where surface wave and sky wave mix
6	Hectometric 1000-100m	Medium frequency (MF) 300-3000kHz	1810-2000kHz (160m band, also known as topband)	Surface wave only during daylight. At night reflection from decaying E-layer	Strong D-region absorption during day. Long ranges possible at night but signals subject to fading and considerable co-channel interference
7	Decametric 100m-10m	High frequency (HF) 3-30MHz	3.50-3.80MHz (80m) 7.00-7.20MHz (40m) 10.10-10.15MHz (30m) 14.00-14.35MHz (20m) 18.068-18.168MHz (17m) 21.00-21.45MHz (14m) 24.89-24.99MHz (12m) 28.00-29.70MHz (10m)	Short-distance working via E-layer. Nearly all long-distance working via F2-layer	Daytime attenuation by D-region, E and F1-layer absorption. Signal strength subject to diurnal, seasonal, solar-cycle and irregular changes
8	Metric 10m-1m	Very high frequency (VHF) 30-300MHz	50.00-52.00MHz (6m) 70.00-70.50MHz (4m) 144.00-146.00MHz (2m)	F2 occasionally at LF end of band around sunspot maximum. Irregularly by sporadic-E and auroral-E. Otherwise maximum range determined by temperature and humidity structure of lower troposphere	Ranges generally only just beyond the horizon but enhancements due to anomalous propagation can exceed 2000km
9	Decimetric 1m-10cm	Ultra high frequency (UHF) 300-3000MHz	430-440MHz (70cm) 1240-1325MHz (23cm) 2310-2450MHz (13cm)	Line-of-sight modified by tropospheric effects	Atmospheric absorption effects noticeable at top of band
10	Centimetric 10cm-1cm	Super high frequency (SHF) 3-30GHz	3.400-3.475GHz (9cm) 5.650-5.850GHz (6cm) 10.00-10.50GHz (3cm) 24.00-24.25GHz (12mm)	Line-of-sight	Attenuation due to oxygen, water vapour and precipitation becomes increasingly important
11	Millimetric 1cm-1mm	Extra high frequency (EHF) 30-300GHz	47.00-47.20GHz (6mm) 75.50-76.00GHz (4mm) 142.00-144.00GHz (2mm) 248.00-250.00GHz (1.2mm)	Line-of-sight	Atmospheric propagation losses create pass and stop bands. Background noise sets a threshold
12	Decimillimetric (sub-millimetric)	- 300-3000GHz	-	Line-of-sight	Present limit of technology

Table 12.2: A survey of the radio-frequency spectrum

$$L_{bf} = 32.45 + 20 \log f(\text{MHz}) + 20 \log r(\text{km}) \quad (5)$$

$$L_{br} = 36.6 + 20 \log f(\text{MHz}) + 20 \log r(\text{miles}) \quad (6)$$

where r is the straight line distance involved.

If transmitter and receiver levels are expressed in either dBW (relative to 1W) or dBm (relative to 1mW) - it does not matter which providing that the same units are used at both ends of the path - with other relevant parameters similarly given in terms of decibels, it is a relatively simple matter to determine the received power at any distance by adding to the transmitter level all the appropriate gains and subtracting all the losses. Thus:

$$P_r(\text{dBW}) = P_t(\text{dBW}) + G_t(\text{dB}) - L_{ft}(\text{dB}) - L_{bf}(\text{dB}) + G_r(\text{dB}) - L_{fr}(\text{dB}) \quad (7)$$

$$Pr(\text{dBm}) = Pt(\text{dBm}) + Gt(\text{dB}) - Lft(\text{dB}) - Lbf(\text{dB}) + Gr(\text{dB}) - Lfr(\text{dB}) \quad (8)$$

where P_r = Received power level (dBm or dBW)
P_t = Transmitted power level (dBm or dBW)
G_t = Gain of the transmitting antenna in the direction of the path, relative to an isotropic radiator
L_{ft} = Transmitting feeder loss
L_{bf} = Free-space transmission loss
G_r = Gain of the receiving antenna in the direction of the path, relative to an isotropic radiator
L_{fr} = Receiving feeder loss

The free-space transmission loss may be estimated approximately from **Fig 12.6** which perhaps conveys a better idea of its relationship to frequency (or wavelength) and distance than the nomogram generally provided. Unless a large number of calculations have to be made, it is no great hardship to use the formula for individual cases should greater accuracy be desired. It should be noted that antenna gains quoted with respect to a half-wave dipole need to be increased by 2dB to express them relative to an isotropic radiator.

Example. A 70MHz transmitter radiates 100W ERP in the direction of a receiver 550km away. The receiving antenna has

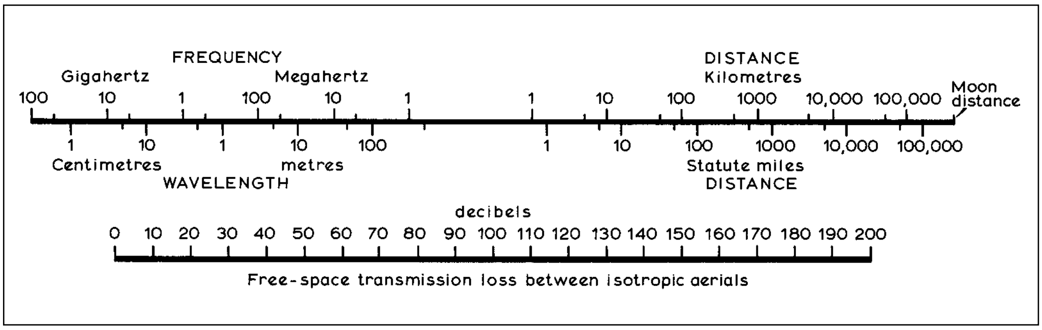


Fig 12.6: The effects of frequency (or wavelength) and distance on the free-space transmission loss between isotropic antennas. The length on the diagram between points corresponding to frequency and distance on the upper scale corresponds to the free-space transmission loss measured from the zero decibel mark on the lower scale

a gain of 5dB over a half-wave dipole (2dB more over an isotropic radiator), and there is a 2dB loss in the feeder. In this case the effective radiated power is known, which takes the place of the terms P_t, G_t and L_{ft}.

	P_{erp}	=	100W = 105mW = 50dBm
	L_{bf}	=	124 from Fig 12.6 or by use of the expressions (5) or (6)
	G_r	=	7dB over an isotropic radiator
	L_{fr}	=	2dB
Then	P_t	=	$P_{erp} - L_{bf} + G_r - L_{fr}$
		=	50 - 124 + 7 - 2
		=	-69dBm, or 69dB below 1mW
		=	12.6×10^{-6}mW
		=	12.6×10^{-9}W

If this power is dissipated in an input impedance of 70 ohms the voltage appearing across the receiver V_r is $\sqrt{(P_r Z_{in})}$, in this case $\sqrt{(12.6 \times 10^{-9} \times 70)}$ which is 94µV. This example deals with the same situation which was considered earlier in connection with field strength and offers an alternative method of calculation

Wave Propagation in the Ionosphere

It has been shown that a transmitted signal may be considered as consisting of a succession of spherical wavefronts, each one a wavelength apart, and they approximate to plane surfaces at great distances. At certain heights in the upper atmosphere concentrations of negatively charged free electrons occur, and these are set into oscillatory motion by the oncoming waves, which causes them to emit secondary wavelets having a phase which is 90ø in advance of the main wave. It is only in the forward direction that the original waves and their dependent wavelets combine coherently and their resultant consists of a wave in which the maxima and minima occur earlier than in the projection of the originating wave - to all intents and purposes the equivalent of the wave having travelled faster in order to arrive earlier. The amount of phase advancement is a function of the concentration of electrons and the change of speed is greatest at long wavelengths, decreasing therefore as the signal frequency increases.

The advancing wave-front, travelling, let us say, obliquely upwards from the ground, meets the layer containing the accumulation of free electrons in such a way that its upper portion passes through a greater concentration of charge than does a portion lower down. The top of the wavefront is therefore accelerated to a greater extent by the process just outlined than are the parts immediately below, which results in a gradual swing-round until the wavefront is being returned towards the ground as though it had experienced a reflection.

The nearer the wavefront is to being vertically above the transmitter the more quickly must the top accelerate relative to the bottom, and the more concentrated must be the charge of electrons. It may be that the density of electrons is sufficiently high to turn even wavefronts propagating vertically upwards (a condition known as *vertical incidence*), although it must be appreciated that deeper penetration into the layer will occur as the propagation angle becomes steeper.

Consider the circumstances outlined in **Fig 12.7**, where T indicates the site of a transmitting station and R1, R2 and R3 three receiving sites. For a given electron concentration there is a *critical frequency* f_0 which is the highest to return from radiation directly vertically upward. Frequencies higher than this will penetrate the layer completely and be lost in space, but their reflection may still be possible at *oblique incidence* where waves have to travel a greater distance within the electron concentration. This is not the case at point A in the diagram, so that reception by sky-wave is impossible at R1 under these circumstances, but at a certain angle of incidence to the layer (as at point B) the ray bending becomes just sufficient to return signals to the ground, making R2 the nearest location relative to the transmitter at which the sky-wave could be received. The range over which no signals are possible via the ionosphere is known as the *skip distance*, and the roughly circular area described by it is called the skip zone. Lower-angle radiation results in longer ranges, for example to point R3 from a reflection at C, and a second 'hop' may result from a further reflection from the ground. The longest ranges at HF are achieved this way - and it is possible for an HF signal to travel right round the world using a succession of hops.

If the angle φ is very large and conditions are favourable, the transmission path may lead from one point on the ionosphere to another without intermediate ground reflection. This is known as *chordal hop* propagation; signal strength is usually higher than normal because there are no ground reflection points or extra transits of the absorbing D-layer to introduce losses.

From the point of view of the operator, it is φ, the angle of signal take-off relative to the horizon (**Fig 12.8**, inset diagram), that is mainly of interest in this application, not the angle of incidence at the ionosphere. The two curves in the main diagram show the distances covered for various values of θ at representative heights of 120km (E-layer) and 400km (F2-layer).

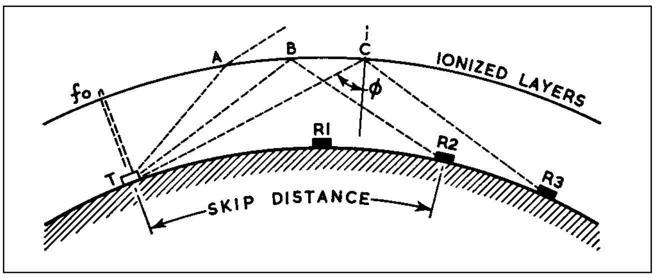


Fig 12.7: Wave propagation via the ionosphere. T is the site of a transmitter, R1, R2 and R3 are the sites of three receivers. The significance of the ray at vertical incidence (dotted) and the three oblique rays is explained in the text

For oblique incidence on a particular path (eg the ray from T to R3 via point C in the ionosphere) there is a *maximum usable frequency* (MUF), generally much higher than the critical frequency was at vertical incidence, and this is approximately equal to $f_0/\cos\varphi$, where φ is the angle of incidence of the ray to the point of reflection, as is shown in the figure. The limiting angle which defines the point at which reflections first become possible is called the *critical wave angle*, and it is this function which determines the extent of the skip zone.

This mechanism is effective for signals in excess of about 100kHz, for which the concentration of electrons appears as a succession of layers of increasing electron density having the effect of progressively bending the rays as the region is penetrated.

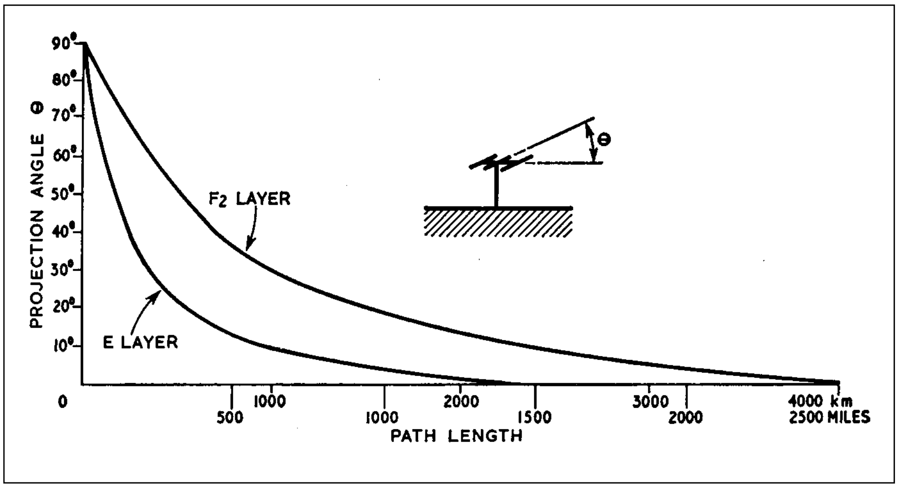


Fig 12.8: Relationship between angles of take-off and resulting path length. The two ionospheric layers are assumed to be at heights of 120km and 400km respectively

Below 50kHz the change in concentration occurs within a distance which is small compared to the wavelength and which therefore appears as an almost perfect reflector. Waves are propagated in that way over great distances by virtue of being confined between two concentric spheres, one being the lower edge of the layer and the other the surface of the Earth, sometimes referred to as *waveguide mode*.

During the hours of daylight the quantity of free electrons in the lower ionosphere becomes so great that the oscillations set up by incident waves are heavily damped on account of energy lost by frequent collisions with the surrounding neutral air particles. Medium-wave broadcast band signals are so much affected by this as to have their sky-waves completely absorbed during the day, leading to the familiar rapid weakening of distant stations around dawn and their subsequent disappearance at the very time when the reflecting layers might otherwise be expected to be most effective. The shorter wavelengths are less severely affected, but suffer attenuation nevertheless. On the 136kHz band, signals can be enhanced at extreme range by ‘reflection’ from the D layer, the effect peaking at mid-path.

A certain amount of cross-polarisation occurs when ionospheric reflections take place so that the received signals generally contain a mixture of both horizontal and vertical components, irrespective of which predominated at the transmitting antenna.

Ionospheric scatter propagation does not make use of the regular layers of increased electron density. Instead forward scattering takes place from small irregularities in the ionosphere comparable in size with the wavelength in use (generally around 8m or about 35MHz). With high powers and very low angles of radiation, paths of some 2000km are possible and this mode has the advantage of being workable in auroral regions where conventional HF methods are often unreliable, but only a very small proportion of the transmitted power is able to find its way in the desired direction.

Wave Propagation in the Troposphere

The *troposphere* is that lower portion of the atmosphere in which the general tendency is for air temperature to decrease with height. It is separated from the *stratosphere*, the region immediately above, where the air temperature tends to remain invariant with height, by a boundary called the *tropopause* at around 10km. The troposphere contains all the well-known cloud forms and is responsible for nearly everything loosely grouped under the general heading of 'weather'.

Its effect upon radio waves is to bend them, generally in the same direction as that taken by the Earth's curvature, not by encounters with free electrons or layers of ionisation, but as a result of successive changes in the refractive index of the air through which the waves pass. In optical terms this is the mechanism responsible for the appearance of mirages, where objects beyond the horizon are brought into view by raybending, in that case resulting from temperature changes along the line-of-sight path. In the case of radio signals the distribution of water vapour also plays a part, often a major one where anomalous propagation events are concerned.

The refractive index, n, of a sample of air can be found from the expression $n = 1 + 10^{-6}N$, where N is the refractivity expressed by:

$$N = \frac{77.6P}{T} + \frac{4810e}{T^2}$$

where P = the atmospheric pressure in millibars or hectopascals

e = the water-vapour pressure, also in millibars or hectopascals

T = the temperature in kelvin or degrees absolute

Substituting successive values of P, e and T from a standard atmosphere table shows that there is a tendency for refractive index to decrease with height, from a value just above unity at the Earth's surface to unity itself in free space. This gradient is sufficient to create a condition in which rays are normally bent down towards the Earth, leading to a similar state of affairs as that which would result from the radio horizon being extended to beyond the optical line-of-sight limit by an average of about 15%.

The distance d to the horizon from an antenna of height h is approximately $\sqrt{(2ah)}$, where a is the radius of the Earth and h is small compared with it. The effect of refraction can be allowed for by increasing the true value of the Earth's radius until the ray paths, curved by the refractive index gradient, become straight again. This modified radius a′ can be found from the expression:

$$\frac{1}{a'} = \frac{1}{a} + \frac{dn}{dh}$$

where dn/dh is the rate of decrease of refractive index with height. The ratio a'/a is known as the *effective Earth radius factor k*, so that the distance to the radio horizon becomes $d = \sqrt{(2kah)}$.

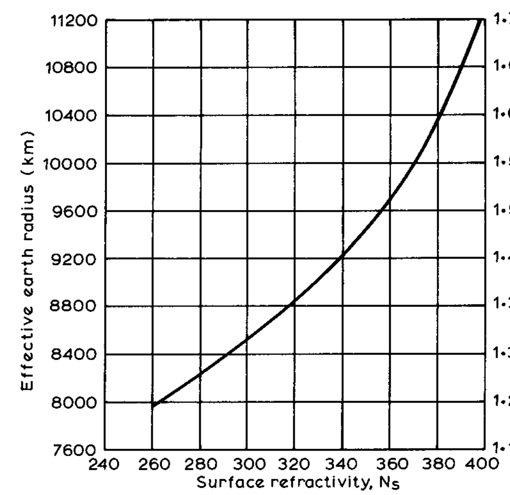


Fig 12.9: Effective Earth radius corresponding to various values of surface refractivity. A curved ray-path drawn over an Earth section between terminals may be rendered straight by exaggerating the Earth's curvature. An average value often used is equal to 1.33 times the actual radius; this is sometimes referred to as the four-thirds Earth approximation. (Based on CCIR Report 244.)

An average value for k, based on a standard atmosphere, is 1.33 or 4/3 (whence a common description of this convention, *the four-thirds Earth*). Notes relating to the construction of path profiles using this convention will be found in a later section of this chapter. When the four-thirds Earth concept is known to be inappropriate an estimate of a suitable value for k can be obtained from the surface refractivity N_s (obtained from **Fig 12.9**), using for N the value obtained from ground-level readings of pressure, temperature and vapour pressure but this method may be misleading if marked anomalies are present in the vertical refractive index structure.

Waves of widely separated length are liable to be disturbed by the troposphere in some way or other, but it is generally only those shorter than about 10m (over 30MHz) which need be considered. There are two reasons for this; one is that usually ionospheric effects are so pronounced at the longer wavelengths that attention is diverted from the comparatively minor enhancements due to the refractive index structure of the troposphere, and the other that anomalies, when they occur, are often of insufficient depth to accommodate waves as long as, or longer than, 10m. For example, it might be that the decrease of refractive index with height becomes so sharp in the lower 100m of the troposphere, that waves are trapped in an *atmospheric duct*, within which they remain confined for abnormally long distances. The maximum wavelength which can be trapped completely in a duct of 100m thickness is about 1m (corresponding to a frequency of 300MHz), for example, so that the most favourable conditions are generally found in the VHF and UHF bands or above. The relationship between maximum wavelength λ and duct thickness t is shown in the expression:

$$t = 500\ \lambda^{2/3}$$

where both t and λ are expressed in centimetres.

At centimetre wavelengths signals propagated through the troposphere suffer rapid fluctuations in amplitude and phase due to irregular small-scale variations in refractive index which give rise to continuous changes known as *scintillations*, and they are also attenuated by water in the form of precipitation (rain, snow, hail, etc) or as fog or cloud. As **Table 12.3** shows, this effect increases both with radio frequency and with either the rate of rainfall or the concentration of water droplets. Precipitation causes losses by absorption and by random scattering from the liquid (or solid in the case of ice) surfaces and this scattering becomes so pronounced as to act as a 'target' for weather radars which use these *precipitation echoes* to detect rain areas.

At even shorter wavelengths resonances occur within the molecules of some of the gases which make up the atmosphere. Only one of these approximates to an amateur band, the attenuation at 22.23GHz due to water vapour. The principal oxygen resonances at around 60 and 120GHz are in the upper reaches of amateur activities, but are likely to prove to be limiting factors where amateur and professional work at millimetre wavelengths involves paths passing through the atmosphere, as opposed to outer space working.

Wave Propagation near the Ground

Diffraction is an alteration in direction of the propagation of a wave due to change in velocity over its wavefront. Radio waves meeting an obstacle tend to diffract around it, and the surface of the Earth is no exception to this. Bending comes about as a result of energy being extracted due to currents induced in the ground. These constitute an attenuation by absorption, having the effect of slowing down the lower parts of the wavefront, causing it to tilt forward in a way which follows the Earth's curvature. The amount of diffraction is dependent on the ratio of the wavelength to the radius of the Earth and so is greatest when the waves are longest. It also depends on the electrical characteristics of the surface, namely its relative permeability (generally regarded as unity for this purpose), its dielectric constant, ε (epsilon), and conductivity, σ (sigma). This diffracted wave is known as the *surface wave*.

Moisture content is probably the major factor in determining the electrical constants of the ground, which can vary considerably with the type of surface as can be seen from **Table 12.4**. The depth to which the wave penetrates is a function also of frequency, and the depth given by the δ (delta) value is that at which the wave has been attenuated to 1/*e* (or about 37%) of its surface magnitude.

At VHF and at higher frequencies the depth of penetration is relatively small and normal diffraction effects are slight. At all frequencies open to amateur use, however, the ground itself appears as a reflector, and the better the conductivity of the sur-

Frequency band (GHz)	Precipitation (mm/h)					Fog or cloud water content (g/m3) (at 0°C)		
	100	50	25	10	1	2.35	0.42	0.043
3.400-3.475	0.1	0.02	0.01	-	-	-	-	-
5.650-5.850	0.6	0.25	0.1	0.02	-	0.09	-	-
10-10.5	3.0	1.5	0.6	0.2	0.01	0.23	0.04	-
21-22	13.0	6.0	2.5	1.0	0.1	0.94	0.17	0.02
24	17.0	8.0	3.8	1.5	0.1	1.41	0.25	0.03
48-49	30.0	17.0	9.0	4.0	0.6	4.70	0.84	0.09

100mm/h = tropical downpour; 50mm/h = very heavy rain; 25mm/h = heavy rain: 10mm/h = moderate rain, 1mm/h = light rain. 2.35g/m3 visibility of 30m; 0.42g/m3 100m; 0.043g/m3 500m.

Table 12.3: The attenuation in decibels per kilometre to be expected from various rates of rainfall and for various degrees of cloud intensity

Type of surface	Dielectric constant ε	Conductivity σ (mho/m) 1MHz	10MHz	100MHz	1000MHz	Depth of penetration δ (m) 1MHz	10MHz	100MHz	1000MHz
Sea water (20°C)	70	5	5	5	5	0.25	0.08	0.02	0.001
Fresh water (10°C)	80	0.003	0.003	0.005	0.2	15	10	4	0.3
Very moist soil	30	0.01	0.01	0.02	0.2	5	3	1.5	0.2
Average ground	15	0.001	0.001	0.002	0.04	25	16	6	0.5
Very dry ground	3	0.0001	0.0001	0.0001	0.0002	90	90	90	40

(Top) Table 12.4: Typical values of dielectric constant, conductivity and depth of wave penetration for various types of surface at various frequencies

Distance (km)	Free-space field (dB) rel to 1µV/m for a 1kW transmitter	Sea, σ = 4 1.8	3.5	7.0	Land, σ = 3 10 2 1.8	3.5	7.0	Land, σ = 3 10 3 1.8	3.5	7.0	Land, σ = 10 3 1.8	3.5	7.0
3	100	1	1	1	1	3	12	8	18	30	19	28	36
10	90	1	1	1	6	8	22	18	29	42	30	38	47
30	80	1	2	2	8	18	34	28	41	51	40	48	56
100	70	2	3	7	18	33	51	44	56	68	58	63	73
300	60	10	12	23	43	61	88	68	85	-	78	-	-
1000	50	48	55	-	-	-	-	-	-		-	-	-

(Bottom) Table 12.5: The number of decibels to be subtracted from the calculated free-space field in order to take into account various combinations of ground conductivity and distance. The values are shown in each case for 1.8, 3.5 and 7.0MHz. Vertical polarisation is assumed

face the more effective the reflection (Table 12.5). It is this effect which makes generalisations of wave propagation near the ground difficult to make for frequencies greater than about 10MHz where the received field is, more often than not, due to the resultant of waves which have travelled by different paths.

Further aspects of propagation near the ground will be dealt with in the section on multiple-path propagation where it will be necessary to consider the consequences of operating with antennas at heights of one to several wavelengths above the ground.

MULTIPLE-PATH EFFECTS

Introduction

The preceding descriptions of the various modes of propagation do not necessarily paint a very realistic picture of the way in which the signals received at a distant location depend on the radio frequency in use and the distance from the transmitter, excluding ionospheric components. The reason for this is that the wave incident on the receiving antenna is rarely only the one which has arrived by the most direct path but is more often the resultant of two or more waves which have travelled by different routes and have covered different distances in doing so. If these waves should eventually arrive in phase they would act to reinforce one another, but should they reach the receiving antenna in antiphase they would interfere with one another and, if they happened to be equal in amplitude, would cancel one another completely.

These alternative paths may arise as a result of reflections in the horizontal plane (as in Fig 12.10(a), where a tall gasholder intercepts the oncoming waves and deflects them towards the receiving site) or in the vertical plane (as in Fig 12.10(b), where reflection occurs from a point on the ground in line-of-sight from both ends of the link). If the reflecting surface is stationary, as it ought to be in the two cases so far considered, the phase difference (whatever it may be) would be constant and a steady signal would result.

It may happen that the surface of reflection is in motion as it would be if it was part of an aeroplane flying along the transmission path. In that case the distance travelled by the reflected wave would be changing continually and the relative phases would progressively advance or retard through successive cycles (effectively an increase or decrease in frequency - the *Doppler effect*), leading to alternate enhancements and degradations as the two waves aid or oppose one another. This performance is one which is particularly noticeable on analogue TV receivers sited near an airport, and even non-technical viewers can instantly diagnose as *aircraft flutter* the fluctuations in picture brilliance which result.

Any 'ghost' image on an analogue TV picture is evidence of a second transmission path, and the amount of its horizontal displacement from the main picture is a measure of the additional transmission distance involved. So, with the reflection from the moving aircraft, the displacement of the ghost picture will change as the path length changes, and its brilliance will reach a maximum every time the difference between the direct path and the reflected path is exactly a whole number of wavelengths. An analytical treatment of the appearance of aircraft reflections on pen recordings of distant signals has appeared in

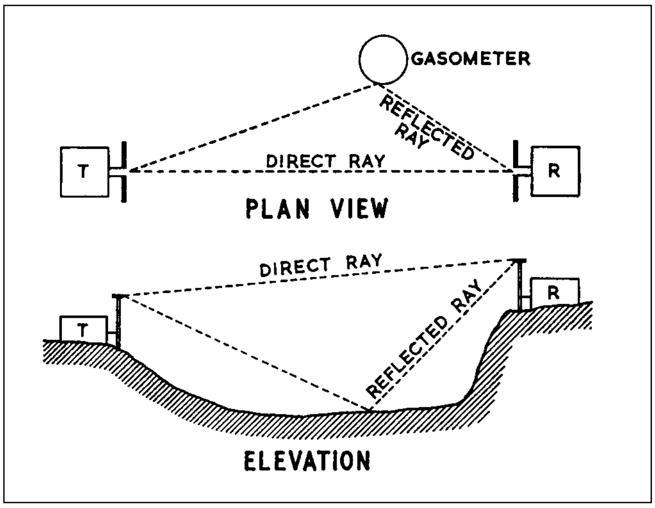


Fig 12.10: Multipath effects brought about by reflections in (a) the horizontal plane, and (b) the vertical plane

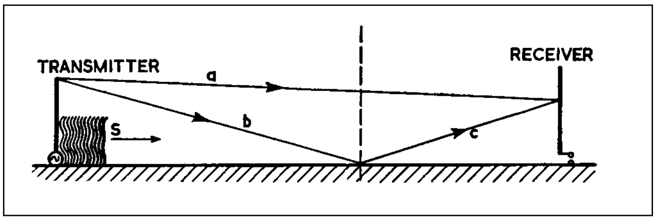


Fig 12.11: The diffracted surface wave S together with the space wave, composed in turn of a direct wave (a) and a reflected wave (b) and (c)

the pages of *Radio Communication* [1]. UHF and low microwave band operators have used reflections from aircraft to make radio contacts.

Because the waves along the reflected path repeat themselves after intervals of exactly one wavelength, it is only the portion of a wavelength 'left over' which determines the phase relationship in comparison with the direct-path wave. This suggests that relatively small changes in the position of a receiving antenna could have profound effects on the magnitude of the received signal when multiple paths are present, and this is indeed found to be the case, particularly where the point of reflection is near at hand.

Ground-wave Propagation

It should now be evident why it was not possible to generalise on the relationship between distance and received signal strength when dealing with propagation near the ground. The surface wave, influenced by the diffraction effects considered earlier, is only one of the possible paths. If the spacing of the transmitting and receiving antennas is such that they are not hidden from one another by the curvature of the Earth there will also be a space wave, made up of a direct wave and a ground-reflected wave as suggested by **Fig 12.11**.

The combination of this space wave and the diffracted surface wave form what is called the *ground wave*, and it may sometimes be difficult (and often perhaps unnecessary) to try to separate this into its three components.

Beyond the radio horizon the direct and reflected rays are blocked by the bulge of the Earth, and the range attained is then determined by the surface wave alone. This diffracted wave is strong at low frequencies (including the amateur 160m band, and more particularly at 136kHz) but becomes less so as the carrier frequency increases and may be considered negligible at VHF and beyond. When occasional signals are received well beyond the horizon the dominant mechanism may be forward scatter.

The strength of the reflected component of the space wave depends largely on the conductivity and smoothness of the ground at the point of reflection, being greatest where oversea paths are involved, and least over dry ground and rock. An extensive treatment of the various factors concerned will be found in the Society's journal [2]. If perfect reflection is presumed, the received field strength due to the interaction of one reflected ray with the direct ray can be estimated from the expression:

$$E = \frac{2E_0}{d} \sin\left(2\pi \frac{h_t h_r}{\lambda d}\right)$$

where E is the resultant received signal strength, E_0 is the direct-ray field strength, h_t and h_r are the effective antenna heights above a plane tangential to the Earth at the point of reflection, d is the distance traversed by the direct ray, and λ the wavelength, all units being consistent (eg metres).

It can be seen that the magnitude of the received signal depends on the relative heights of the two antennas, the distance between them, and, of course, the frequency.

This relationship suggests that doubling the antenna height has the same effect on the received signal strength as halving the length of the path. In view of the respective distances involved it will be appreciated that an increase in the height of one of the antennas has a greater effect than a comparable horizontal movement towards the transmitter.

The Effect of Varying Height

A few moments of experiment with two pieces of cotton representing the two alternative ray paths will provide a convincing demonstration of the effect of altering the height of one or both of the antennas. An increase in height of (say) the receiving antenna has little effect on the length of the direct path, but the ground-reflected ray has to travel further to make up the additional distance and it therefore arrives at a later point on its cycle. The consequence is even more pronounced when it is realised that at very low angles of incidence, when the two antennas are at ground level, the indirect ray may well have experienced a phase change of 180 degrees upon reflection so that the two components arrive roughly equal in magnitude but opposite in phase, so tending to cancel. As the height of one or both antennas is increased the space wave increases in magnitude and the field becomes the vector sum of the diffracted surface wave and the space wave. At even greater heights the effect of the surface wave can be neglected, while the intensity of the space wave continues to increase.

In practice these considerations apply only to antennas carrying VHF, UHF and above. This is because it is not practicable to raise antennas to the necessary heights at the longer wavelengths, and in any case the reception of ionospherically propagated waves imposes different requirements as regards angle of arrival.

As with other functions which depend on Earth constants for their effectiveness there is a marked difference between overland and over-sea conditions. There is more to be gained by raising an antenna over land than over sea, for high frequencies than for low, and for horizontally polarised waves rather than vertically polarised ones.

The ratio of the received field at any given height above ground to the field at ground level due to the surface wave alone (presuming that the two components of the space wave have cancelled one another) is known as the *height-gain factor*. This can be expressed either as a multiplier, or as the corresponding equivalent in decibels, using the voltage scale of relationships.

Over flat ground there is little to be gained from raising antennas for frequencies below about 3MHz, unless it is to clear local obstacles, but it should not be overlooked that it may be desirable to raise antennas no matter what their frequency of operation for reasons unconnected with height-gain benefits - to increase the distance to the radio horizon, for example.

The result of changing the receiving antenna height is by no means as predictable as some authorities would have us believe, and the subject is still a matter thought worthy of further investigation at some research establishments. The following figures summarise the gains to be expected after raising a receiving antenna from a height of 3m to a height of 10m above the ground, according to a current CCIR report [3], primarily concerned with television broadcasting frequencies but relevant nevertheless:

50-100MHz	Median values of height-gain 9-10dB.
180-230MHz	Median values of height-gain 7dB in flat terrain and 4-6dB in urban or hilly areas

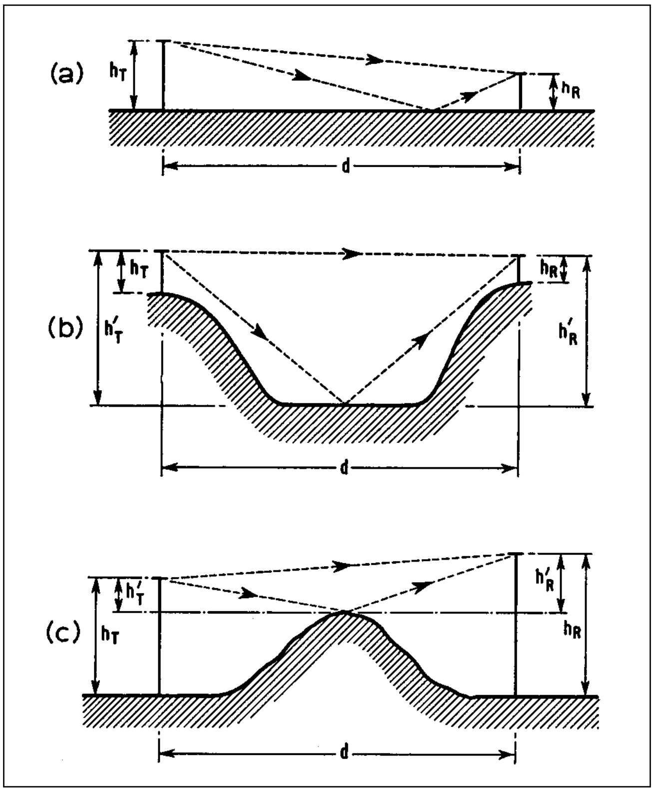


Fig 12.12: The effect of terrain on direct and indirect rays, showing the effective heights of the antennas when undertaking height-gain calculations.

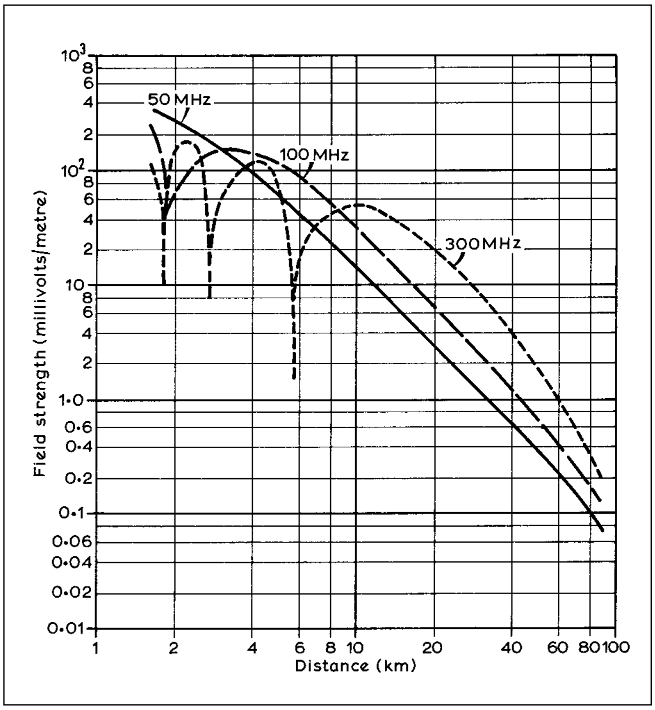


Fig 12.13: Relationship between field strength and distance at VHF and UHF

450-1000MHz	Median values of height-gain very dependent on terrain irregularity. In suburban areas the median is 6-7dB, and in areas with many tall buildings 4-5dB

A simple rule-of-thumb often adopted by radio amateurs is to reckon on a height-gain of 6dB for each time that the antenna height is doubled (eg if 12dB at 3m height, then expect 18dB at 6m, 24dB at 12m, etc), but the presence of more component waves than the two considered can lead to wide departures from this relationship, particularly in urban areas.

If the terrain is not flat the result of altering the antenna height depends largely on the position in the vertical plane of the reflection points relative to the two terminals. Thus in Fig 12.12 the situation shown in (a) corresponds to the one already considered.

Should the two antennas be sited on hills, or separated by a valley, as at (b), there will be large differences in path length between the direct and ground-reflected rays which alterations in antenna height will do little to alter, so that elevating it is unlikely to have very much effect on the received signal strength.

On the other hand, the presence of high ground between transmitter and receiver, as at (c), may make communication between them difficult at low antenna heights, and in that case there would be a great deal to be gained from raising them. In cases (b) and (c) the two antennas should be considered as having effective heights of h'_T and h'_R respectively when dealing with height-gain calculations.

If the intervening high ground has a relatively sharp and well-defined upper boundary, such as would be the case with a mountain ridge, the receiving antenna height at which signals cease might be much lower than would be expected from line-of-sight considerations, even when refractivity changes are taken into account. This is because of an effect known as *knife-edge diffraction*, which often enables 2m operators situated in the Scottish Highlands (to cite an instance) to receive signals from other stations which are apparently obscured from them by surrounding mountains.

The Effect of Varying Distance

The effect on the field strength of varying the distance between transmitter and receiver is shown in Fig 12.13, where the result is again due to the interference between the direct wave and the ground-reflected wave passing through successive maxima and minima as the path difference becomes an exact odd or even number of halfwavelengths. (It must be remembered that a low-angle ground-reflection itself introduces a phase change of very nearly 180 degrees).

The spacing of the maxima (which are greater in magnitude than the free-space value) is closer the higher the frequency of operation and the shorter the path for a given frequency. The most distant maximum will occur when the path difference is down to one half wavelength; beyond that the difference tends towards zero and the two waves progressively oppose one another, the field rapidly falling below the free-space estimate. If the antenna heights are raised the patterns move outwards.

As with all these matters involving ground reflection there is a difference between the behaviour of horizontally and vertically polarised waves, and the foregoing description favours the former.

The reflection coefficient and phase shift at the reflection point vary appreciably with the ground constants when vertical polarisation is employed. In practice, whichever is used, the measured field strength may vary considerably from the calculated value because of the presence of other components due to local reflections. A fairly reliable first estimate for VHF and UHF paths up to about 50km unobstructed length is just to allow for a possible increase or decrease of 10dB on the free-space figure.

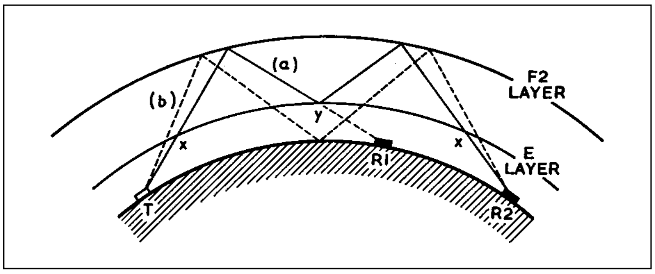


Fig 12.14: Fading due to multipath reception. In case (a) the frequency in use is higher than the E-layer MUF at the points marked 'x', but lower at 'y', where the ray will suffer reflection. In case (b) the frequency is higher than the MUF of the E-layer at all four contacts with it

Fresnel Zones

In all the explanations so far it has been presumed that reflection at a surface occurs at the point which enables the reflected ray to travel the shortest distance.

Because the surfaces considered in radio propagation work are neither plane nor perfect reflectors, the received waveform is the resultant of signals which have been reflected from an area, the size of which is determined by the frequency and by the separation of the terminal antennas, so that the individual reflected path lengths differ by no more than half a wavelength from one another. The locus of all the points surrounding the direct path which give exactly half a wavelength path difference is described by an ellipsoid of revolution having its foci at the transmitting and receiving antennas respectively. A cross-section of this volume on an intersecting plane of reflection encloses an area known as the *first Fresnel zone*.

The radius, *R*, of the first Fresnel zone at any point, *P*, is given by the expression:

$$R^2 = \frac{\lambda d_1 d_2}{d_1 + d_2}$$

where d_1 and d_2 are the two distances from *P* to the ends of the path, and λ is the wavelength, all quantities being in similar units. The maximum radius occurs when d_1 and d_2 are equal. Thus, on an 80km path at a wavelength of 2m, the radius of the first Fresnel zone at the midpoint is 200m, and this represents the clearance of the line-of-sight ray (corrected for refraction) necessary if the path is to be considered 'unobstructed'.

Higher orders of Fresnel zone surround the regions where similar relationships occur after separations of one wavelength, two wavelengths etc, but the conditions for reflection in the required direction rapidly become less favourable and Fresnel zones other than the first are rarely considered.

Ionospheric Multi-path Effects

The multiple-path effects so far considered have been mainly associated with VHF, UHF and SHF where very low angles of radiation and reception are generally involved, and where the wavelengths concerned are sufficiently small to enable optimum heights and favourable positions to be found for the antennas.

When ionospherically propagated signals are of interest it is generally sufficient to regard the ground wave as a single entity, without attempting to separate it into its three components. When multi-path effects occur (as they frequently do) they may be between the ground wave and an ionospheric wave, between two ionospheric waves which may have been propagated by different layers or, in the case of long range transmissions, by signals which have followed different paths entirely in different directions around the Earth's curvature, perhaps in several 'hops' between the ionosphere and the ground. Whatever the cause the result is inevitably a fading signal.

The ionosphere is not a perfect reflector; it has no definite boundaries and it is subject to frequent changes in form and intensity. These deficiencies appear on the received signal as continual small alterations in phase or frequency as the effective path length alters and, when only a single ionospheric wave is present, can pass almost unnoticed by the average HF listener, who is remarkably tolerant of imperfections on distant transmissions. However, when a second signal from the same source is present, which may be either the relatively steady ground wave or another ionospheric component, these phase or frequency changes become further emphasised by appearing as changes in amplitude as the waves alternately reinforce and interfere, and by distortion of modulated signals if the various sideband frequencies do not resolve back into their original form.

PC data recordings of signal strength generally show very clearly the period of fading which results when two modes of propagation begin to interact, continuing until the second predominates. This could occur as a result of circumstances similar to those shown in **Fig 12.14**, where the receiver at R2 may receive signals either by double-hop off the ground (as in path (b)) or double-hop off an intermediate layer, path (a), depending on the relationship between the signal frequency and the maximum usable frequency at point y. The transition between one propagation mode and another is generally accomplished within a relatively short time.

Occasionally very long-distance transmissions may be heard with a marked echo on their modulation. As the two signal components responsible have obviously travelled routes of markedly differing length they probably require very different azimuths at both ends of the path. This effect is most noticeable on omnidirectional broadcast transmissions and is minimised by the use of narrow-beam antennas for transmission and reception.

Fading

Fading is generally, though not exclusively, a consequence of the presence of multiple transmission paths. For that reason it is appropriate to include a summary of its causes here, although more properly some of the comments belong elsewhere in this chapter.

Fading is a repetitive rise and fall in signal level, often described as being *deep* or *shallow* when referring to the range of amplitudes concerned, and *slow* or *rapid* when discussing the period. It is sometimes *random*, usually *periodic*, but occasionally *double periodic*, as when a signal with a rapid fade displays slow changes in mean level. Generally the fading rate increases with frequency because a particular motion in the ionosphere causes a greater phase shift at the shorter wavelengths.

At VHF and UHF the fading rate is often closely associated with the pattern of atmospheric pressure at the surface, tending to become slower during periods of high pressure. This can be particularly noticeable on a pen recording of signal strength taken while a ridge of high pressure moves along the transmission path, the slowest rate occurring as the ridge crosses the mid-point.

Interference fading, as its name implies, is caused by interference between two component waves when one or both path lengths are changing, perhaps due to fluctuations in the ionosphere or troposphere, or due to reflections from a moving surface. The period is relatively short, usually up to a few seconds. Fast interference fading is often called *flutter*. Auroral flutter comes in this category, being caused by motion of the reflecting surfaces.

Polarisation fading is brought about by continuous changes in polarisation due to the effect of the Earth's magnetic field on the ionosphere. Signals are at a maximum when they arrive with the same polarisation as the receiving antenna. Period again up to a few seconds.

Absorption fading, generally of fairly long period, is caused by inhomogeneities in the troposphere or ionosphere. Period up to an hour, or longer.

Skip fading occurs when a receiver is on the edge of a skip zone and changes in MUF cause the skip distance to shorten and lengthen. Highly irregular as regards period of fade.

Selective fading is the name given to a form of fading characterised by severe modulation distortion in which the path length in the ionosphere varies with frequency to such an extent that the various sideband frequencies are differently affected. It is most severe when ground and sky waves are of comparable intensity.

Scintillations are rapid fluctuations in amplitude, phase and angle-of-arrival of tropospheric signals, produced by irregular small-scale variations in refractive index. The term is also used to describe irregular fluctuations on HF signals transmitted through the ionosphere from satellites and other sources outside the Earth.

Diversity Reception

The effects of fading can be countered to a certain extent by the use of more than one receiving system coupled to a common network which selects at all times the strongest of the outputs available.

There are three principal versions in common use:

(a) *space diversity*, obtained by using antennas which are so positioned as to receive different combinations of components in situations where multi-path conditions exist;

(b) *frequency diversity*, realised by combining signals which have been transmitted on different frequencies; and

(c) polarisation diversity, where two receivers are fed from antennas having different planes of polarisation.

It is unlikely that any of these systems have any application in normal amateur activities, but they may be of interest in connection with research projects relating to the amateur hands.

SOLAR AND MAGNETIC INFLUENCES

The Sun

Our Sun is at the centre of a complex system consisting of nine major planets (including ours), five of which have two or more attendant moons, of several thousand minor planets (or asteroids), and of an unknown number of lesser bodies variously classed as comets or meteoroid swarms.

It is a huge sphere of incandescence of a size which is equivalent to about double our moon's orbit around the Earth, but, despite appearances to the contrary, it has no true 'edge' because nearly the whole of the Sun is gaseous and the part we see with apparently sharp boundaries is merely a layer of the solar atmosphere called the *photosphere* which has the appearance of a bright surface, preventing us from seeing anything which lies beneath.

Beyond the photosphere is a relatively cooler, transparent layer called the *chromosphere*, so named because it has a bright rose tint when visible as a bright narrow ring during total eclipses of the Sun. From the chromosphere great fiery jets of gas, known as *prominences*, extend. Some are slow changing and remain suspended for weeks, while others, called *eruptive prominences*, are like narrow jets of fire moving at high speeds and for great distances.

Outside the chromosphere is the *corona*, extending a distance of several solar diameters before it becomes lost in the general near-vacuum of interplanetary space. At the moment of totality in a solar eclipse it has the appearance of a bright halo surrounding the Sun, and at certain times photographs of it clearly show it being influenced by the lines of force of the solar magnetic field.

The visible Sun is not entirely featureless - often relatively dark *sunspots* appear and are seen to move from east to west, changing in size, number and dimensions as they go. They are of interest for two reasons: one is that they provide reference marks by which the angular rotation of the Sun can be gauged, the other that by their variations in number they reveal that the solar activity waxes and wanes in fairly regular cycles. The Sun's rotation period has been found to vary with latitude, with its maximum angular speed at the equator. The *mean synodic rotation period* - the time required for the Sun to rotate until the same part faces the Earth is 27.2753 days.

The Sun's rotations have been numbered since the year 1853 (the start of rotation 2039 is dated 18 January 2006), and observations of solar features are referred to an imaginary network of latitudes and longitudes which rotates from east to west as seen from the Earth. Remember, when looking at the mid-day Sun from the UK, north is at the top, the east limb is on the left side and the west limb on the right. A long-persistent feature which first appears on the east limb is visible for about 13.5 days before it disappears from sight over the west limb.

For many purposes it is more convenient to refer the positions of noteworthy features to a related, but stationary, set of co-ordinates, the *heliocentric latitude and longitude*, in which locations are described with respect to the centre of the visible disc. An important statistic relating to sunspots is the time of their *central meridian passage* (CMP).

Radio telescopes detect features which are usually situated in the vicinity of the solar corona, and some of them reveal disturbances beyond the limbs of the visible disc. They often travel across the face of the Sun at a faster rate than any spots beneath.

The apparent diameter of the Sun varies with the choice of radio frequency. Because the lower frequencies come from the outer parts of the corona, the width of their Sun appears to be large. At high frequencies the sources are situated below the level which provides the visible disc so that the width of the Sun then appears to be small. The optical and radio frequency diameters are similar at a frequency of 2800MHz (10.7cm wavelength), so that was chosen to provide the daily *solar flux* measurements. The daily flux figure, which can vary between about 65 at solar minimum and 300 at maximum, is now more frequently used than the sunspot number as a measure of the Sun's ionising capability. The higher the figure the better in terms of HF propagation (subject to the level of geomagnetic activity). A 90-day average of the daily figures has been found to be a good basis for home computer prediction programs.

Sunspots

Sunspots are the visible manifestation of very powerful magnetic fields; adjacent spots often have opposing polarities. These intense magnetic fields also produce *solar flares*, which are emissions of hydrogen gas. They are responsible also for the ejection of streams of charged particles and X-rays.

Sunspot numbers have been recorded for over 200 years and it has been found that their totals vary over a fairly regular cycle occupying around 22 years where magnetic polarity is the criterion, or 11 years if only the magnitude of activity is considered.

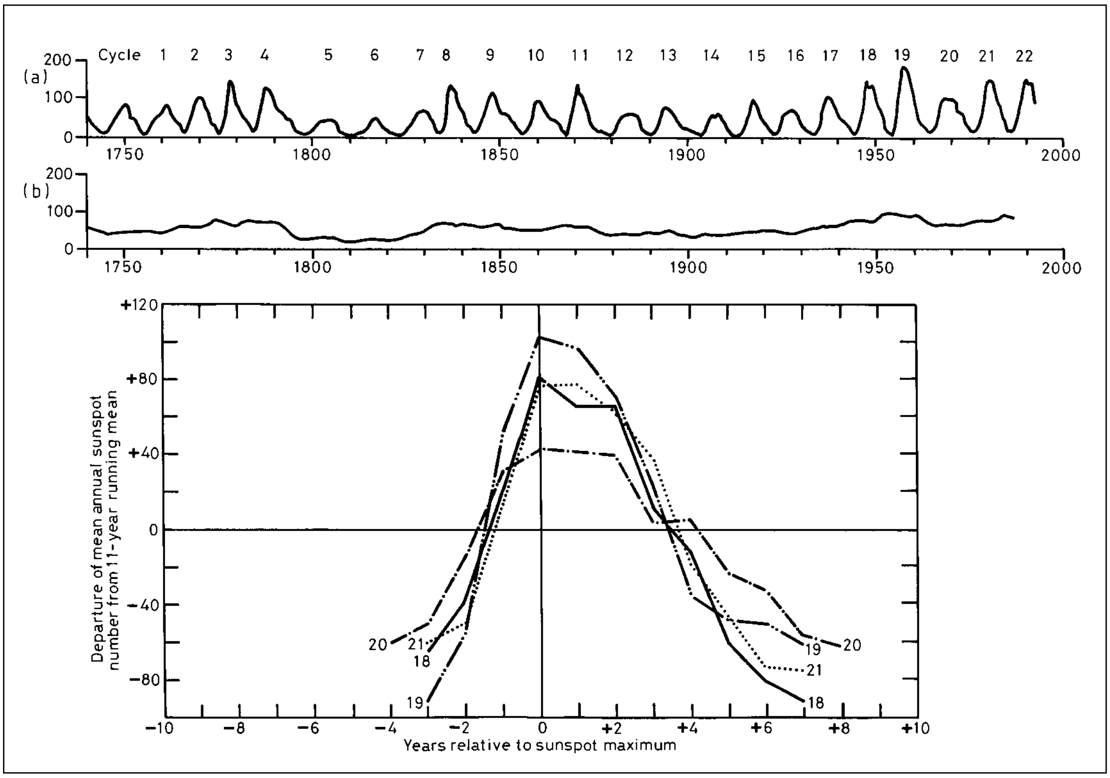


Fig 12.15: (a) Annual relative sunspot numbers, the means of monthly means of daily values. (b) The 11-year running means of the annual numbers. The succession of points reveals the underlying trend in solar activity. (c) A comparison of the four cycles most recently completed. The steep rise to maximum and the relatively slow decline thereafter are characteristic of all sunspot cycles

The 11-year peaks are known as *sunspot maxima*; the intervening troughs are *sunspot minima*.

The rise and fall times are not equal, though. Four years and seven years respectively are typical, although each cycle differs from the others in both timing and maximum value, as may be seen in **Fig 12.15**. At sunspot minimum the Sun may be completely spotless for weeks or months - or even, during the Maunder minimum in the 17th century, for years.

Tables of daily *relative sunspot numbers* are prepared monthly at the Sunspot Index Data Centre (SIDC) in Brussels, from information supplied by a network of participating observatories. It is widely supposed by radio amateurs that those figures indicate the number of visible spots but that is not so. In accordance with a formula devised by Dr Wolf in Zurich (hence the description *Wolf number*, still used professionally) the relative sunspot number, R, is found from the expression:

$$R = k\,(10g + t)$$

where k is a regulating factor that keeps the series to a uniform standard, g is the number of spot groups, and t is the total number of spots.

Daily figures obtained from the Ursigram messages (see later) use unity as the value of k, so, for example, Boulder figures obtained from that source record just 10 times the number of groups plus the total number of spots. Note also that those figures are provisional because they will have been prepared in haste to meet a deadline.

When the figures from the participating observatories have been combined in Brussels, a value of k which is less than unity will have been applied. That figure does not appear in the tables and it may well be subject to frequent variation, but it is currently around 0.7.

The Brussels figures are issued twice, first provisionally as soon as possible, then definitively after more careful scrutiny. From the monthly means of the definitive values a smoothed index R_{12} is obtained. This is the arithmetic mean of 12 successive monthly means, the result being ascribed to the period at the centre of the sample. In order to make that fall in the middle of the month, rather than between months, 13 months are taken but the first and the last are given only half weight in the calculations. From the nature of the 12-month running average - for that is what it is - it must be evident that R_{12} (which is the 'sunspot number' called for in ionospheric prediction programs) never reaches the peaks and troughs of the individual monthly means and it falls far short of the maxima of the daily values.

To put all these different versions of the 'sunspot number' into perspective we have only to look at a specific example, say the month which contained the peak of solar cycle 22, which was June 1989. The Boulder figures on the Ursigrams reached 401, and their monthly mean was 297. The SIDC Brussels definitive figure at maximum was 265, the monthly mean 196. However, the smoothed figure R_{12} for the month was only 158 and, remember, it is that one that you need for prediction programs, not the 401 from Boulder. You should be aware also that the latest smoothed figure available is always six months behind the current date, so a figure for this month has been a forecast six months ahead.

It is also of interest to observe that the peak Boulder figure was made up of 18 groups which, between them, contained a total of 221 spots. Put those figures in the formula and you come up with 401, the figure reported. The three largest groups accounted for 86, 53 and 26 spots respectively, and none of the others contained more than 7.

The Solar Wind

The solar corona was described in the last section as extending outwards until it becomes lost in interplanetary space. In fact it does not become lost at all, but turns into a tenuous flow of hydrogen which expands outwards through the solar system, taking with it gases evaporating from the planets, fine meteoric dust and cosmic rays. It becomes the *solar wind.*

Near the Sun the corona behaves as a static atmosphere, but once away it gradually accelerates with increasing distance to speeds of hundreds of kilometres per second. The gas particles take about nine days to travel the 150,000,000km to the Earth, carrying with them a magnetic field (because the gas is ionised) which assumes a spiral form due to the Sun's rotation. It is the solar wind, rather than light pressure alone, which is responsible for comets' tails flowing away from the Sun, causing them to take on the appearance of celestial wind-socks.

The existence of the solar wind was first detected and measured by space vehicles such as the Luniks, Mariner II and Explorer X, which showed that its speed and turbulence are related to solar activity. Regular measurements of solar wind velocity are now routine.

There is thus a direct connection between the atmosphere of the Sun and the atmosphere of the Earth. In the circumstances it is hardly surprising that solar events, remote though they may at first seem, soon make their effects felt here on Earth.

The Earth's Magnetosphere

It is well known that the Earth possesses a magnetic field, for most of us have used a compass, at some time or another, to help us to get our 'bearings'. We know from such experiences that the field appears to be concentrated at a point somewhere near the north pole (and are prepared to believe that there is another point of opposing polarity somewhere near the south pole). Popular science articles have familiarised us with a picture of field lines surrounding the Earth like a section of a ring doughnut made up of onion-like layers.

Because the particles carried by the solar wind are charged their movement produces a magnetic field which interacts with the geomagnetic field. A blunt shock-wave is set up, called the *magnetosheath*, and the wind flows round it, rejoining behind where the field on the far side is stretched in the form of a long tail, the overall effect being reminiscent of the shape of a pear with its stalk pointing away from the Sun. The region within the magnetosheath, into which the wind does not pass, is called the *magnetosphere*.

On the Earthward side the magnetosphere merges into the ionosphere. Inside the magnetosphere there are regions where charged particles can become trapped by geomagnetic lines of force in a way which causes them to oscillate back and forth over great distances. Particles from the solar wind can enter these regions (often called *Van Allen belts* after their discoverer) in some way, as yet not perfectly understood.

The concentration of electrons in the magnetosphere can be gauged from the ground by observations on *whistlers*, naturally-occurring audio-frequency oscillations of descending pitch which are caused by waves radiated from the electric discharge in a lightning flash. These travel north and south through the ionosphere and magnetosphere from one hemisphere to another along the magnetic lines of force. The various component frequencies propagate at different speeds so that the original flash (which appears on an ordinary radio receiver as an *atmospheric*) arrives at the observer considerably spread out in time. The interval between the reception of the highest and lowest frequencies is a function of the concentration of electrons encountered along the way.

These trapped particles move backwards and forwards along the geomagnetic field lines within the Van Allen belts and some collide with atoms in the ionosphere near the poles where the belts approach the Earth most closely. Here they yield up energy either as ionisation or illumination and are said to have been dumped. These dumping regions surround the two poles, forming what are called the *auroral zones*. The radius of the circular motion in the spirals (they are of a similar form to that of a helical spring) is a function of the strength of the magnetic field, being small when the field is strong. Electrons and protons perform their circular motions in opposite senses, and the two kinds of spiralling columns drift sideways in opposite directions, the electrons eastward, the protons westward, around the world. Because of the different signs on the two charges these two drifts combine to give the equivalent of a current flowing in a ring around the Earth from east to west.

This ring current creates a magnetic field at the ground which combines with the more-or-less steady field produced from within the Earth. We shall see later the sort of effects that solar disturbances have on the magnetosphere, the ionosphere, and the total geomagnetic field.

The Quiet Ionosphere

With the stage set to follow the antics of the ions and electrons deposited by dumping we must pause again, this time to examine the normal day-to-day working of the ionosphere, which is dependent for its chemistry on another form of incoming solar radiation.

The gas molecules in the Earth's upper atmosphere are normally electrically neutral, that is to say the overall negative charges carried by their orbiting electrons exactly balance the overall positive charges of their nuclei. Under the influence of ultraviolet radiation from the Sun, however, some of the outer electrons can become detached from their parent atoms, leaving behind overall positive charges due to the resulting imbalance of the molecular structure. These ionised molecules are called ions, from which of course stems the word 'ionosphere'.

This process, called *disassociation*, tends to produce layers of free electrons brought about in the following manner. At the top of the atmosphere where the solar radiation is strong there are very few gas molecules and hence very few free electrons. At lower levels, as the numbers of molecules increase, more and more free electrons can be produced, but the action progressively weakens the strength of the radiation until it is unable to take full advantage of the increased availability of molecules and the electron density begins to decline. Because of this there is a tendency for a maximum (or peak) to occur in the production of electrons at the level where the increase in air density is matched by the decrease in the strength of radiation. A peak formed in this way is known as a *Chapman layer*, after the scientist who first outlined the process.

The height of the peak is determined not by the strength of the radiation but by the density/height distribution of the atmosphere and by its capability to absorb the solar radiation (which is a function of the UV wavelength), so that the layer is lower when the radiation is less readily absorbed. The strength of the radiation affects the rate of production of electrons at the peak, which is also dependent on the direction of arrival. The electron density is greatest when the radiation arrives vertically and it falls off as a function of zenith distance, being proportional to cos χ, where χ (the Greek letter chi) is the angle between the vertical and the direction of the incoming radiation. As cos χ decreases (ie when the Sun's altitude declines) a process of recombination sets in, whereby the free electrons attach themselves to nearby ions and the gas molecules revert to their normal neutral state.

Experimental results have led to the belief that the E-layer (at about 120km) and the F1-layer (at about 200km) are formed according to Chapman's theory as a result of two different kinds of radiation with perhaps two different atmospheric constituents involved.

The uppermost layer F2 (around 400km), which normally appears only during the day, does not follow the same pattern and is thought to be formed in a different way, perhaps by the

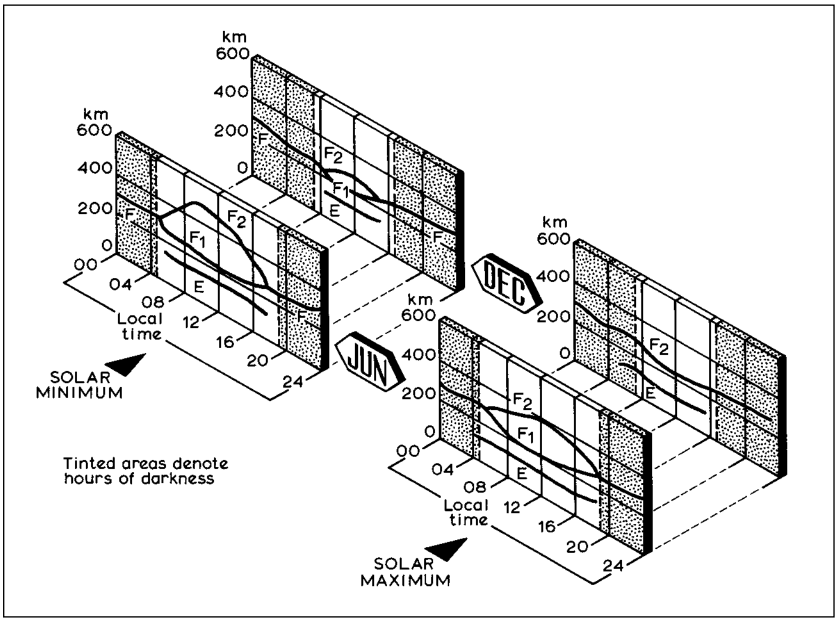


Fig 12.16; Typical diurnal variations of layer heights for summer and winter at minimum and maximum states of the solar cycle

diffusion of ions and electrons, but there are still a number of anomalies in its behaviour which are the subject of current investigation. These include the *diurnal anomaly*, when the peak occurs at an unexpected time during the day; the *night anomaly*, when the intensity of the layer increases during the hours of darkness when no radiation falls upon it; the *polar anomaly*, when peaks occur during the winter months at high latitudes, when no illumination reaches the layer at all; the *seasonal anomaly*, when magnetically quiet days in summer (with a high Sun) sometimes show lower penetration frequencies than quiet days in winter (with a low Sun); and a *geomagnetic anomaly* where, at the equinoxes, when the Sun is over the equator, the F2-layer is most intense at places to the north and south separated by a minimum along the magnetic dip equator. It is thought that topside sounding from satellites probing the ionosphere from above the active layers may help to explain some of these anomalies in F2 behaviour.

Regular Ionospheric Layers

The various regular ionospheric layers were first defined by letters by Sir Edward Appleton who gave to the one previously known as the *Kennelly-Heaviside layer* the label 'E' because he had so marked it in an earlier paper denoting the electric field reflected from it, and to the one he had discovered himself the letter 'F', rather than call it the *Appleton layer*, as some had done. To the band of absorption below thus naturally fell the choice of the letter 'D', although this was generally referred to as a 'region' rather than a 'layer' because its limits are less easy to define.

From comments already made it will be appreciated that the regular ionospheric layers which these letters define exhibit changes which are basically a function of day and night, season and solar cycle.

Most of our knowledge of the ionosphere comes from regular soundings made at vertical incidence, using a specialised form of radar called an *ionosonde* which transmits short pulses upwards using a carrier which is continuously varied in frequency from the medium-wave broadcast band through the HF bands to an upper limit of about 20MHz, but beyond if conditions warrant.

Reflections from the various layers are recorded photographically in the form of a graph called an *ionogram*, which displays *virtual height* (the apparent height from which the sounding signal is returned) as a function or signal frequency.

Critical frequencies, where the signals pass straight through the layers, are read off directly. There are many such equipments in the world; the one serving the United Kingdom is located near Slough. It is under the control of the Rutherford Appleton Laboratory, which houses one of a number of World Data Centres to which routine measurements of the ionosphere are sent from most parts of the world.

The two sets of diagrams (Fig 12.16 and Fig 12.17) summarise the forms taken by the diurnal variations in height and critical frequency for two seasons of the year at both extremes of the sunspot cycle. The actual figures vary very considerably from one day to the next, but an estimate of the expected monthly median values of maximum usable frequency and optimum working frequency between two locations at any particular year, month and time of day can be obtained from predictions .

The critical frequencies of the E and F1-layers are a function of R12, the smoothed SIDC Brussels (formerly Zurich) relative sunspot number (which is predicted six months in advance for this purpose), and the cosine of the zenith distance χ, the angle

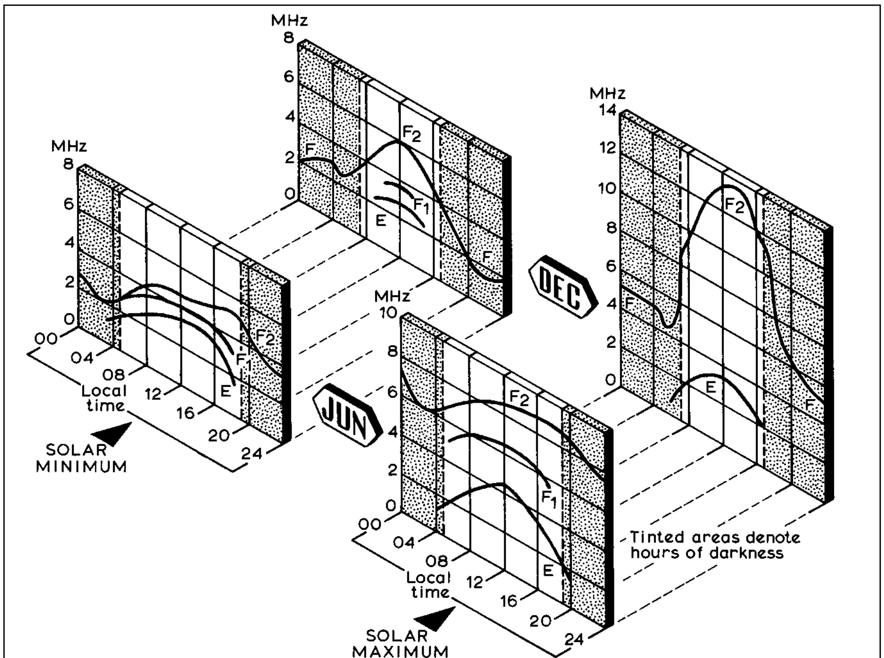


Fig 12.17: Typical diurnal variations of F-layer critical frequencies for summer and winter at the extremes of the solar cycle

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
$R_3 = 0$	5.3	5.1	4.75	4.80	5.10	5.03	4.72	4.75	4.90	5.69	5.58	5.32
$R_3 = 150$	12.17	12.42	11.67	9.88	8.23	7.70	7.73	7.68	8.81	11.26	12.93	12.53

Table 12.6: F2-layer critical frequencies at Slough. The two rows show mean median value of f_0F2 in megahertz for three month weighted mean sunspot numbers of 0 and 150

	Distance			
Layer	1000km	2000km	3000km	4000km
Sporadic E	4.0	5.2	-	-
E	3.2	4.8	-	-
F1	2.0	3.2	3.9	-
F2 winter	1.8	3.2	3.7	4.0
F2 summer	1.5	2.4	3.0	3.3

Table 12.7: MUF (maximum usable frequency) factors for various distances assuming representative heights for the principal layers

the Sun makes with the local vertical, and thus, to a first approximation:

$$f_0E = 0.9\,[(180 + 1.44R)\cos\chi]^{0.25}$$

(usually to within 0.2MHz of the observed values), and:

$$f_0F_1 = (4.3 + 0.01R)\cos^{0.2}\chi$$

which is less accurate because of uncertainty in the value of the exponent which varies with location and season.

The F2-layer is the most important for HF communication at a distance, but is also the most variable. It is subject to geomagnetic control which impresses a marked longitudinal effect on the overall world pattern. causing it to lag behind the sub-solar point so as to give maximum values in critical frequency during the local afternoon.

The F2 critical frequency f_0F2 varies with the solar cycle, as shown in **Table 12.6**, which shows monthly median values for Slough, applicable to sunspot numbers of 0 and 150. In recent years it has been found possible to predict the behaviour of the F2-layer by extrapolation several months ahead, using an index known as *IF2*, which is based on observations made at about 10 observatories.

The MUF which can be used on a particular circuit may be calculated from the critical frequency of the appropriate layer by applying the relationship:

$$\text{MUF} = F/\sec\varphi$$

where φ is the angle that the incident ray makes with the vertical through the point of reflection at the layer. The factor (sec φ) is called the *MUF factor*; it is a function of the path length if the height of the layer is known. **Table 12.7** shows typical figures obtained by assuming representative heights.

To an operator, the *optimum working frequency* (OWF) is the highest (of those available) which does not exceed the MUF. As will be seen later, both MUF and OWF take on different meanings in the context of ionospheric predictions.

There is a lower limit to the band of frequencies which can be selected for a particular application. This is set by the *lowest usable frequency* (LUF), below which the circuit becomes either unworkable or uneconomical due to the effects of absorption and the level of radio noise. Its calculation is quite a complicated process beyond the scope of this survey.

It is often useful to be able to estimate the radiation angle involved in one- or two-hop paths via the E and F2-layers. **Fig 12.18**, also prepared for average heights, accomplishes this. It is a useful rule-of-thumb to remember that the maximum one-

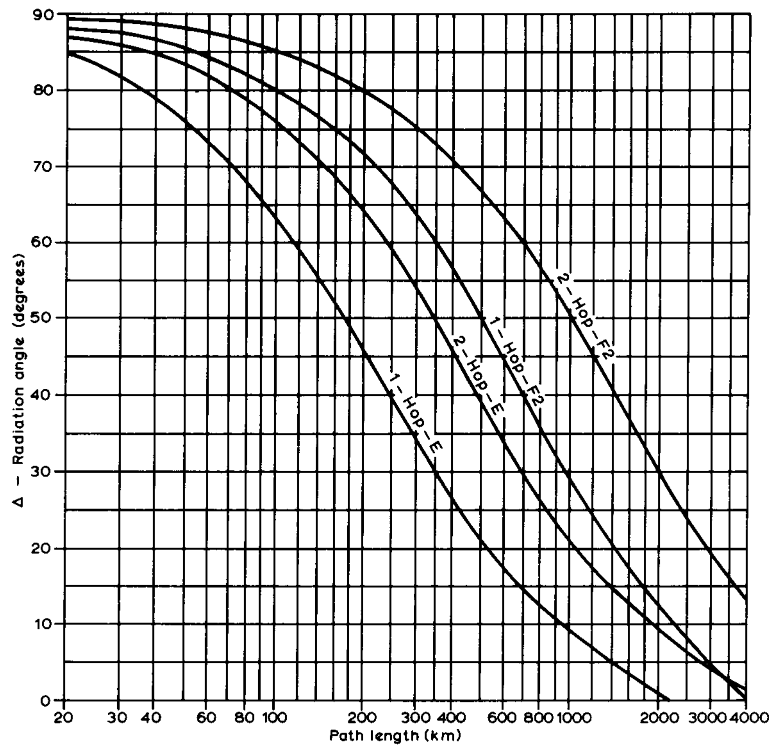


Fig 12.18: Radiation angle involved in one-hop and two-hop paths via E and F2-layers. (From NBS publication *Ionospheric Radio*

hop E range is 2000km and that the useful two-hop E range, twice that (4000km), is also the one-hop F2 range. Of course, all extreme ranges require very low angles of take-off, almost unachievable for radio amateurs.

Irregular Ionisation

Besides the regular E, F1 and F2-layers there are often more localised occurrences of ionisation which make their contribution to radio propagation. They generally occur around the heights associated with the E-layer and often the effects extend well into VHF, although the regular E-layer can never be effective at frequencies of 30MHz or more.

Sporadic E (Es) has been observed at HF on ionospheric sounding apparatus since the early 'thirties. It has been shown to take the form of clouds of high density of ionisation, forming sheets perhaps a kilometre deep and some 100km across in a typical instance, and appearing at a height of 100-120km.

However, strangely, over the years not one of the participating observatories has ever found evidence of a layer having sufficient electron density to support propagation at 144MHz - yet radio amateurs make use of something that behaves as though it was sporadic E many days of the year.

It is mainly a summertime phenomenon in May to August (the months without an 'r' in them) so far as the Northern Hemisphere is concerned, but there is also some activity during the latter part of the year. Distances worked are not often less than 500km, mainly between 1000 and 2000km, and usually with good-quality, steady, strong signals. Two-hop Es has been observed on occasion, but that is more likely to involve two small clouds rather than one big one. Also on record are cases where 2000km has been exceeded by the aid of tropospheric enhancement at one end of the path. (Generally, VHF Es and tropospheric modes are quite separate and a list of their distinguishing features will be found in **Table 12.8**.)

The numbers and durations of sporadic-E events decrease with frequency. To date, the highest recorded frequency appears to be 220MHz but that may not necessarily be the absolute limit. Due to the mystery surrounding the mechanism of the mode at VHF, many observers make a long-term specific study of Es, and there has long been an International Amateur Radio Union co-ordinator whose task is to guide national societies into setting up useful co-operative research projects.

Auroral E (Ar) is closely connected with geomagnetic disturbances. At times of high activity (popularly known as *magnetic storms*), the regions around the north and south poles where visual aurora are commonplace (the *auroral ovals*) expand towards the equator, taking with them the capability of returning VHF signals that have been directed towards them.

It used to be thought that antennas had to be turned towards the north (some even said the magnetic north) in order to take advantage of this mode but, thanks to careful observing by a group of dedicated amateurs over a period of many years, it is now known that optimum bearings can and do change considerably during an auroral event, and that from the UK a gradual swing towards the east may be expected as the activity develops. Radar measurements have shown that the reflecting regions are usually around 110km in height, but it should be noted that visual aurora extends very much beyond that. However, there is a quite good general relationship between the radio and visual forms of aurora, although attempts to match details on pen recordings of received signals against observations of changes in the structure of auroral forms seen from the transmitting site have been disappointing.

The signal paths are of necessity angled, often with one leg much longer than the other, the two antennas being directed towards a common reflection point that is at a latitude higher than that of either station (forward scatter is unlikely). Stations in central Europe are able to work with beam headings considerably west of north; that sector is of little use to UK operators because there are no available contacts in the North Atlantic.

Tropospheric propagation	Sporadic-E propagation
May occur at any season	Mainly May, June, July and August
Associated with high pressure, or with paths parallel to fronts	No obvious connection with weather patterns
Gradual improvement and decline of signals	Quite sudden appearance and disappearance
Onset and decay times similar over a wide range of frequencies	Begins later and ends earlier as radio frequency increases
Observed at VHF, UHF, SHF	Rarely above 200MHz
Area of enhancement relatively stable for several hours at a time	Area of enhancement moves appreciably in a few hours
May last a week or more	Duration minutes or hours, never days
Wide range of distances with enhanced signals at shorter ranges	Effects mainly at 1000-2000km. No associated enhancement at short ranges

Table 12.8: Comparative characteristics of tropospheric and sporadic-E propagation

In Europe the signals have a raw, rasping tone that is readily recognised again, once identified. It is said to differ in character from the tone of auroral signals met with in North America.

A radio auroral event typically begins in the afternoon and may appear to be over by the end of the afternoon, but many events exhibit two distinct phases, and the evening phase is often the better in terms of DX worked, partly due perhaps to the greater number of stations likely to be active at that time. The event will frequently finish with dramatic suddenness, just as some of the longest paths are being achieved.

A book outlining the theory of auroral-E propagation, together with an analysis of observations made over much of the time since the second world war, will be found in the bibliography at the end of this chapter.

Trans-equatorial propagation (TEP) is one of the success stories of amateur radio research. Much of the pioneering work was carried out at 50MHz, but higher-frequency working, eg 70MHz, is possible. Instances of TEP tend to favour the years of high solar activity, but it is present, on a reduced time scale, even near sunspot minimum. Paths are typically 3000-9000km in length, usually with a north/south bias: examples are Europe/Africa, Japan/Australia, North America/South America. The stations in contact are usually symmetrically located with respect to the magnetic dip equator, with the path in between them being perpendicular to it.

Two types of TEP have been recognised. One shows a peak in activity at around 1700-1900 local time. That provides the longest of the contacts (9000km or more) with strong signals and low fading rate. It is thought that the mechanism involves two reflections at the F region without intermediate ground contact. The other tends to peak in the evening at around 2000-2300 local time. Signal strengths are high but there is an accompaniment of deep and rapid flutter. Paths are shorter than for the afternoon type, perhaps no more than 6000km, and opinions are divided as to the mechanism involved. The rapid fading characteristic seems to connect in some way with equa-

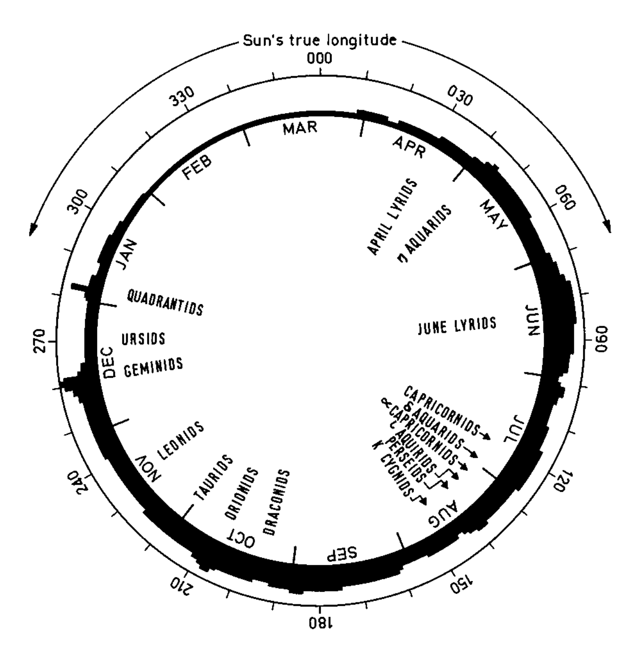


Fig 12.19: Seasonal variation of meteor activity, based on a daily relative index. Prepared from tables of 24h counts made by Dr Peter Millman, National Research Council, Ottawa. The maximum rate corresponds to an average of about 300 echoes per hour corresponding to an equivalent visual magnitude of 6 or greater

torial spread-F, which is a diffuse effect caused by irregularities in the electron density of the F region.

Before television moved from VHF to UHF and beyond, suitably placed radio amateurs in places like Greece were able to entertain visitors from less-fortunate parts of the world with an impressive display of ultra-DX TV, taking advantage of the opportunities offered by TEP.

Operators in southern Europe make use of a form of VHF propagation in which *field-aligned irregularities* (FAI) appear to play an important part. The effect has been observed at mid-latitudes both on the Continent and in North America.

The FAI 'season' runs very closely parallel to that of VHF sporadic E, May to late August, and some instances have been known to follow conventional Es openings.

The mode is characterised by signals arriving away from the expected great-circle bearing. The scattering area responsible is apparently very small and some elevation of perhaps 10-15 degrees has been required (in Italy) to find it. High-power transmitters and low-noise receivers are essential.

Italian amateurs recognise two distinct areas which appear to vary but little from one occasion to the next. One, located in the west of Switzerland, provides 2m contacts with Spain, southern France, Hungary and Yugoslavia. The other is located over Hungary itself and that provides Italian amateurs with openings to the Balkan peninsula.

Although FAI would appear to have very little direct application for UK amateurs at present, a nodding acquaintance with it might lead to fresh discoveries about its capabilities. One theory is that it is associated with anomalies in the Earth's magnetic field. Central Europe is not the only place with those.

The more-conventional *ionospheric scatter mode* is one that is developed commercially to provide communications over some 800-2000km paths, using ionisation irregularities at a height of about 85km, which is in the D-region. The frequencies used were between 30 and 60MHz but, at the top of the range especially, very high powers were necessary.

Received signals were weak, but they had the advantage of being there during periods of severe disruption on the HF bands. The systems that were set up have fallen from favour nowadays and satellite transponders are commonly used instead for much of the traffic. Any amateur involvement would have to be limited to the 50MHz band, but the high power involved suggests that this is now one for the history books.

Meteoric ionisation is caused by the heating to incandescence by friction of small solid particles entering the Earth's atmosphere. This results in the production of a long pencil of ionisation extending over a length of 15km or more, chiefly in the height range 80 to 120km. It expands by diffusion and rapidly distorts due to vertical wind shears. Most trails detected by radio are effective for less than one second, but several last for longer periods, occasionally up to a minute and very occasionally for longer. There is a diurnal variation in activity, most trails occurring between midnight and dawn when the Earth sweeps up the particles whose motion opposes it. There is a minimum around 1800 local time, when only meteors overtaking the Earth are observed. The smaller *sporadic meteors*, most of them about the size of sand grains, are present throughout the year, but the larger *shower* meteors have definite orbits and predictable dates. **Fig 12.19** shows the daily and seasonal variation in meteor activity, based on a 24h continuous watch. Intermittent communication is possible using meteoric ionisation between stations whose antennas have been prealigned to the optimum headings of 5 to 10 degrees to one side of the great-circle path between them. Small bursts of signal, referred to as *pings*, can be received by meteor scatter from distant broadcast (or other) stations situated 1000-1200km away.

Geomagnetism

The Earth's magnetic field is the resultant of two components, a *main field* originating within the Earth, roughly equivalent to the field of a centred magnetic dipole inclined at about 11 degrees to the Earth's axis, and an *external field* produced by changes in the electric currents in the ionosphere.

The main field is strongest near the poles and exhibits slow secular changes of up to about 0.1% a year. It is believed to be due to self-exciting dynamo action in the molten metallic core of the Earth. The field originating outside the Earth is weaker and very variable, but it may amount to more than 5% of the main field in the auroral zones, where it is strongest. It fluctuates regularly in intensity according to annual, lunar and diurnal cycles, and irregularly with a complex pattern of components down to micropulsations of very short duration.

Certain observatories around the world are equipped with sensitive *magnetometers* which record changes in the field on at least three different axes, the total field being a vector quantity having both magnitude and direction. In the aspect of analysis which is of interest to us the daily records, called *magnetograms*, are read-off as eight *K-indices*, which are measures of the highest positive and negative departures from the 'normal' daily curve during successive threehourly periods, using a quasi-logarithmic scale ranging from 0 (quiet) to 9 (very disturbed). The various observatories do not all use the same scale factors in determining *K*-indices; the values are chosen so as to make the frequency distributions similar at all stations. Most large magnetic disturbances are global in nature and appear almost simultaneously all over the world. The more frequently used *planetary K index* is formed by combining the K figures for a dozen selected observatories. It is more finely graded: 0, 1-, 1o, 1+, 2-, 2o, 2+, . . . 9-, 9o, 9+. It is formed by a combination of K-

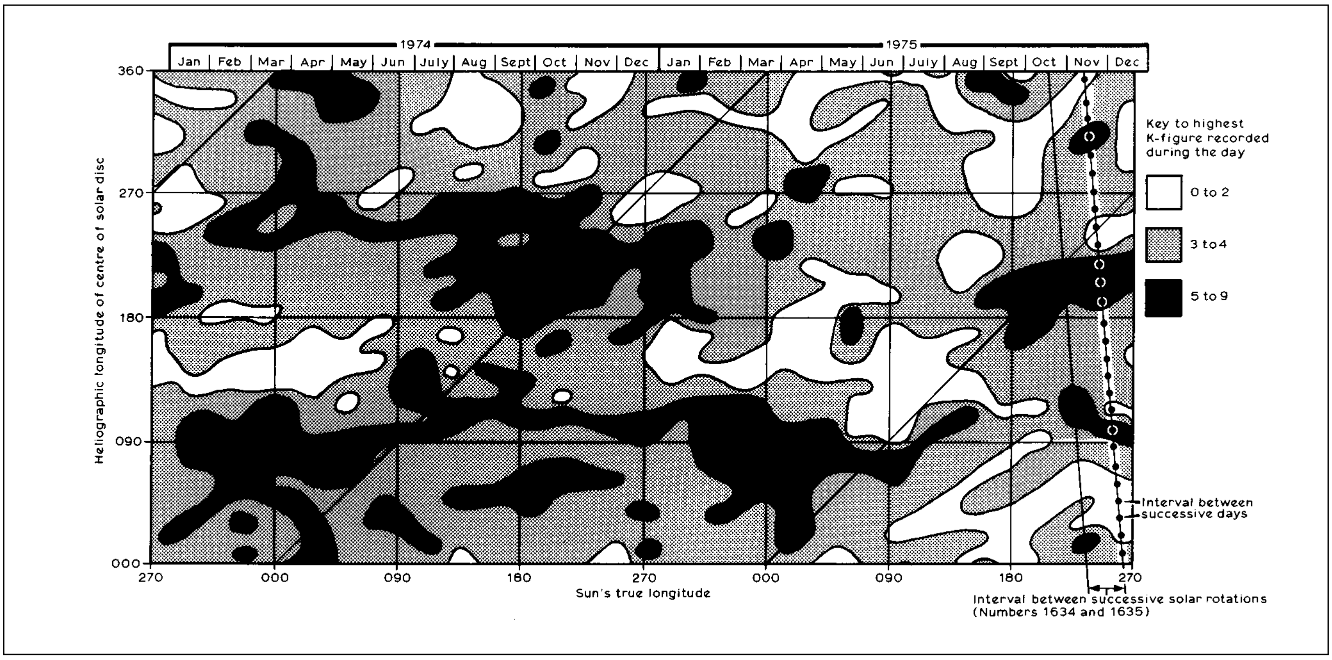


Fig 12.20: Geomagnetic activity diagram. The black areas indicate sequences where a K-figure of 5 or more was recorded at Lerwick Observatory. A horizontal line on this diagram denotes a recurrence period of 27 days, linked to the Sun's rotation period as seen from the Earth; the diagonal lines show the slope associated with a 25-day rotation period, such as the Sun has in relation to a fixed point in space

figures from 12 selected observatories. Indices of 5 or more may be regarded as being indicative of magnetic storm conditions.

K- and K_p-indices are based on quasi-logarithmic scales which place more emphasis on small changes in low activity than high. For some purposes it is more convenient to work with a linear scale, particularly if the values are to be combined to derive averages, as of the day's activity, for example. This is known as the A-index, recording the daily equivalent amplitude on a linear scale running from 0 to 400 - the maximum figure for the most severe storms. As with Kp, the daily Ap figure combines results from a selected group of observatories. Both indices are widely used as a shorthand expression of geomagnetic activity.

There is a tendency for occasions of abnormally high geomagnetic activity (and in consequence auroral activity at VHF and UHF) to recur at intervals of approximately 27 days, linked to the *solar synodic rotation period*. The chart shown in **Fig 12.20** clearly shows some long-persistent activity periods over the two-year interval 1974/75. A blank chart showing these co-ordinates to cover the current year with an overlap (known as a *solar rotation base map*) is published in the information section of the *RSGB Yearbook*.

The original diagram on which it is based was prepared by plotting the highest *K*-figure for each day on the spot determined by the longitude of the Sun's central meridian facing the Earth (thus a measure of the solar rotation), and a parameter called the *Sun's true longitude*, which indicates the position of the Earth around its orbit. Successive rotations build up a raster of daily figures in the way shown by the dots along the sloping right-hand edge, and the resulting chart should really be considered as being cylindrical, with the upper and lower edges brought together. The black areas surround the days when a *K*-figure of five or more was recorded - magnetic storm days - and the unshaded areas enclose relatively quiet days when the *K*-index was two or less. 27-day recurrences are clearly marked on this section of the record but occasionally there are periods when there is a marked tendency for storms to recur after an interval which appears to be linked to the Sun's rotation period relative to the stars - indicated by the slope of the diagonal across the diagram.

Periods of high geomagnetic activity tend to occur somewhat more frequently around the equinoxes, while periods around the summer and winter solstice are more likely to be quieter. The geomagnetic cycle, like the solar cycle, lasts around eleven years, though tending to lag it by roughly a year or eighteen months. However, major disturbances can occur at any time of the year and any stage in the cycle - and quite often do!

Ionospheric Disturbances

Like so much that affects radio propagation ionospheric disturbances have their genesis in the Sun. Recent years have shed much light on the mechanisms involved, though they may not as yet be wholly understood. From time to time flares occur, powerful explosions that hurl vast amounts of highly charged particles out of the Sun.

This material may escape the Sun by means of holes in the Sun's outer corona and, if the hole suitably positioned in relation to Earth, the effects will reach Earth. Initially electromagnetic radiation in the form of X-rays, ultraviolet, visible light and radio waves between 3cm and 10m in length will which reach the Earth in about eight minutes. The X-rays and UV light cause immediate increases in the D-layer ionisation, leading to *short-wave* (or *Dellinger*) *fade-outs* which may persist for anything up to two hours. The effects of an SID affect the lower bands first, higher frequencies later, sometimes wiping out even high-powered transmissions for many hours - though a couple of hours is more common. Recovery works in the opposite direction, with the higher frequencies regaining propagation first. Sometimes prolonged bursts of radio noise also occur. Other effects observed are a *sudden enhancement of atmospherics* (SEA), a *sudden absorption of cosmic noise* (SCNA), and *sudden phase anomalies* (SPA) on very low-frequency transmissions.

This is followed after a few hours by the arrival of *cosmic ray particles* and perhaps the onset of *polar-cap absorption* (PCA).

The main stream of particles arrives after an interval of 20-40h and consists of protons and electrons borne by the solar wind. High-speed coronal streams, not necessarily originating in flare activity) can travel at speeds which sometimes exceed 100km/sec. What happens next greatly depends on the strength and orientation of the interplanetary magnetic field - the intensity of which is expressed in nano-Tesla (nT) units, with a southerly orientation favouring coupling with Earth's magnetic field. Where this applies, particles reaching the Earth's magnetosphere manifest themselves in visible displays of aurora and in auroral backscatter propagation at VHF, occasionally extending into the UHF range. Also, a strong polar electrojet may flow into the lower ionosphere.Changes in the make-up of the trapping regions leads to variations in the circulating ring-current which leads to violent alterations in the strength of the geomagnetic field, bringing about the sudden commencement, which is the first indication of a magnetic storm.

Associated with the magnetic storms are ionospheric storms, and both may persist for several days. The most prominent features are the reduction in F2 critical frequencies (f_0F2) and an increase in D-region absorption. During the storm period signal strengths remain very low and are subject to flutter fading. The effects of an ionospheric storm are most pronounced on paths which approach the geomagnetic poles.Conversely, LF signal levels at 136kHz are usually enhanced by up to about 10dB at the peak of a solar flare.

TROPOSPHERIC PROCESSES

Because it is all around us the troposphere is the portion of the atmosphere which we ought to know best. We are dealing here with the Sun's output of electromagnetic radiation which falls in the infrared portion of the spectrum, between 10^{-6} and 10^{-5}m wavelength, is converted to heat (by processes which need not concern us here) and is distributed about the world by radiation, conduction and convection.

At this point our link between solar actions and atmospheric reactions breaks down, because the very variable nature of the medium, and the ease by which it can be modified both by topographical features and the differing thermal conductivities of land and sea, leads to the development of air masses having such widely contrasting properties that it becomes impossible to find a direct correlation between day-to-day climatic features and solar emissions. We must accept the fact that in meteorology 'chance' plays a powerful role and look to functions of the resulting weather pattern for any relationships with signal level, without enquiring too deeply into the way in which they may be connected with events on the Sun.

Pressure Systems and Fronts

The television weatherman provides such a regular insight into the appearance and progressions of surface pressure patterns that it would be wasteful of space to repeat it all here. Suffice it to record that there are two closed systems of isobars involved, known as *anticyclones* and *depressions* (or, less-commonly nowadays, *cyclones*) within or around which appear *ridges* of high pressure, *troughs* of low pressure (whose very names betray their kinship), and *cols*, which are slack regions of even pressure, bounded by two opposing anticyclones and two opposing depressions.

The most important consideration about these pressure systems, in so far as it affects radio propagation at VHF and above, is the direction of the vertical motion associated with them.

Depressions are closed systems with low pressure at the centre. They vary considerably in size, and so also in mobility, and frequently follow one another in quick succession across the North Atlantic. They are accompanied by circulating winds which tend to blow towards the centre of the system in an anti-clockwise direction. The air so brought in has to find an outlet, so it rises, whereby its pressure falls, the air cools and in doing so causes the relative humidity to increase. When saturation is reached, cloud forms and further rising may cause water droplets to condense out and fall as rain. Point one: depressions are associated with rising air.

Anticyclones are generally large closed systems which have high pressure in the centre. Once established they tend to persist for a relatively long time, moving but slowly and effectively blocking the path of approaching depressions which are forced to go round them. Winds circulate clockwise, spreading outwards from the centre as they do, and to replace air lost from the system in this way there is a slow downflow called *subsidence* which brings air down from aloft over a very wide area. As the subsiding air descends its pressure increases, and this produces dynamical warming by the same process which makes a bicycle pump warm when the air inside it is compressed. The amount of water vapour which can be contained in a sample of air without saturating it is a function of temperature, and in this particular case if the air was originally near saturation to begin with, by the time the subsiding air has descended from, say, 5km to 2km, it arrives considerably warmer than its surroundings and by then contains much less than a saturating charge of moisture at the new, higher, temperature. In other words, it has become warm and dry compared to the air normally found at that level. Point two: anticyclones are associated with descending air.

In addition to pressure systems the weather map is complicated by the inclusion of *fronts*, which are the boundaries between two air masses having different characteristics. They generally arrive accompanied by some form of precipitation, and they come in three varieties: *warm*, *cold* and *occluded*.

Warm fronts (indicated on a chart by a line edged with rounded 'bumps' on the forward side) are regions where warm air is meeting cold air and being forced to rise above it, precipitating on the way.

Cold fronts (indicated by triangular 'spikes' on the forward side of a line) are regions where cold air is undercutting warm. The front itself is often accompanied by towering clouds and heavy rain (sometimes thundery), followed by the sort of weather described as 'showers and bright intervals'.

Occluded fronts (shown by alternate 'bumps' and 'spikes') are really the boundary between three air masses being, in effect, a cold front which has overtaken a warm front and one or the other has been lifted up above the ground.

It is perhaps unnecessary to add that there is rather more to meteorology than it has been possible to include in this brief survey.

Vertical Motion

It is a simple matter of observation that there is some correlation between VHF signal levels and surface pressure readings, but it is generally found to be only a coarse indicator, sometimes showing little more than the fact that high signal levels accompany high pressure and low signal levels accompany low pressure. The reason that it correlates at all is due to the fact that high pressure generally indicates the presence of an anticyclone which, in turn, heralds the likelihood of descending air.

The reason that subsidence is so important stems from the fact that it causes dry air to be brought down to lower levels where it is likely to meet up with cool moist air which has been

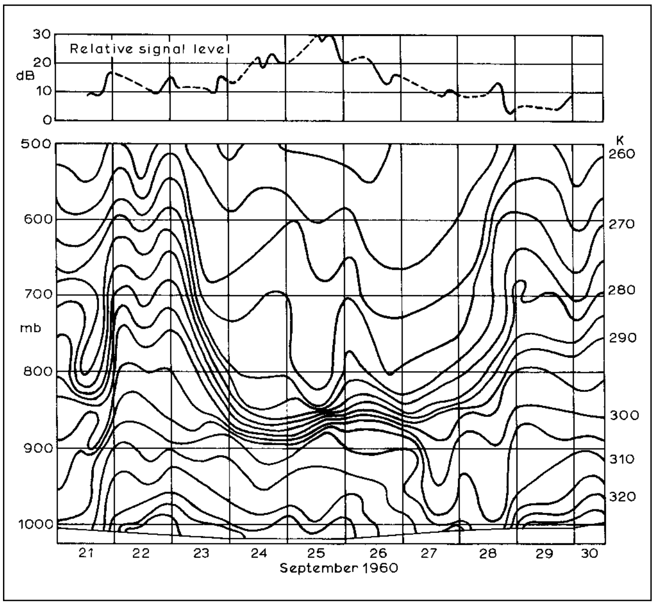


Fig 12.21: The relationship between variations in potential refractive index in the atmosphere and signal strengths over a long-distance VHF tropospheric path. (With acknowledgements to *J Atmos Terr Phys*, Pergamon Press.)

stirred up from the surface by turbulence. The result then is the appearance of a narrow boundary region in which refractive index falls off very rapidly with increasing height - the conditions needed to bring about the sharp bending of high-angle radiation which causes it to return to the ground many miles beyond the normal radio horizon. Whether you regard this in the light of being a benefit or a misfortune depends on whether you are more interested in long-range communication or in wanting to watch an interference-free television screen.

The essential part of the process is that the descending air must meet turbulent moist air before it can become effective as a boundary. If the degree of turbulence declines, the boundary descends along with the subsiding air above it, and when it reaches the ground all the abnormal conditions rapidly become subnormal - a sudden drop-out occurs. Occasionally this means that operators on a hill suffer the disappointment of hearing others below them still working DX they can no longer hear themselves. Note, however, that anticyclones are not uniformly distributed with descending air, nor is the necessary moist air always available lower down, but a situation such as a damp foggy night in the middle of an anticyclonic period is almost certain to be accompanied by a strong boundary layer. Ascending air on its own never leads to spectacular conditions. Depressions therefore result in situations in which the amount of ray bending is controlled by a fairly regular fall-off of refractive index. The passage of warm fronts is usually accompanied by declining signal strength, but occasionally cold fronts and some occlusions are preceded by a short period of enhancement.

To sum up, there is very little of value about propagation conditions which can be deduced from surface observations of atmospheric pressure. The only reliable indicator is a knowledge of the vertical refractive index structure in the neighbourhood of the transmission path.

Radio Meteorological Analysis

It remains now to consider how the vertical distribution of refractive index can be displayed in a way which gives emphasis to those features which are important in tropospheric propagation studies. Obviously the first choice would be the construction of atmospheric cross-sections along paths of interest, at times when anomalous conditions were present, using values calculated by the normal refractive index formula. The results are often disappointing, however, because the general decrease of refractive index with height is so great compared to the magnitude of the anomalies looked for that, although they are undoubtedly there, they do not strike the eye without a search.

A closely-related function of refractive index overcomes this difficulty, with the added attraction that it can be computed graphically and easily, directly from published data obtained from upper-air meteorological soundings. It is called *potential refractive index K* and may be defined as being the refractive index which a sample of air at any level would have if brought *adiabatically* (ie without gain or loss of heat or moisture) to a standard pressure of 1000mb; see references [4] and [5].

This adiabatic process is the one which governs (among other things) the increase of temperature in air which is descending in an anticyclone, so that, besides the benefits of the normalising process (which acts in a way similar to that whereby it is easy to compare different-sized samples of statistics when they have all been converted to percentages) there is the added attraction that the subsiding air tends to retain its original value of potential refractive index all the time it is progressing on its downward journey. This means that low values of *K* are carried down with the subsiding air, in sharp contrast to the values normally found there. A cross-section of the atmosphere during an anticyclonic period, drawn up using potential refractive index, gives an easily-recognisable impression of this.

Fig 12.21 is not a cross-section, but a time-section, showing the way in which the vertical potential refractive index distribution over Crawley, Sussex, varied during a 10-day period in September 1960. There is no mistaking the downcoming air from the anticyclone and the establishment of the boundary layer around 850mb (about 1.5km). Note how the signal strength of the Lille television transmission on 174MHz varied on a pen recording made near Reading, Berkshire, during the period, with peak amplitudes occurring around the time when the layering was low and well-defined, and observe also the marked decline which coincided with the end of the anticyclonic period. Time-sections such as these also show very clearly the ascending air in depressions (although the *K* value begins to alter when saturation is reached) and the passage of any fronts which happen to be in the vicinity of the radiosonde station at ascent time.

For anyone interested in carrying out radiometeorological analysis at home - and, be assured, it is a very rewarding exercise in understanding the processes involved - the propagation chapter in the RSGB *VHF/UHF Handbook* will provide full details (see the bibliography at the end of this chapter).

PRACTICAL CONSIDERATIONS

Map Projections

Maps are very much a part of the life of a radio amateur, yet how few of us ever pause to wonder if we are using the right map for our particular purpose, or take the trouble to find out the reason why there are so many different forms of projection.

The cartographer is faced with a basic problem, namely that a piece of paper is flat and the Earth is not. For that reason his map, whatever the form it may take, can never succeed in being faithful in all respects - only a globe achieves that. The amount by which it departs from the truth depends not only on how big a portion of the globe has been displayed at one viewing, but on what quality the mapmaker has wanted to keep correct at the expense of all others.

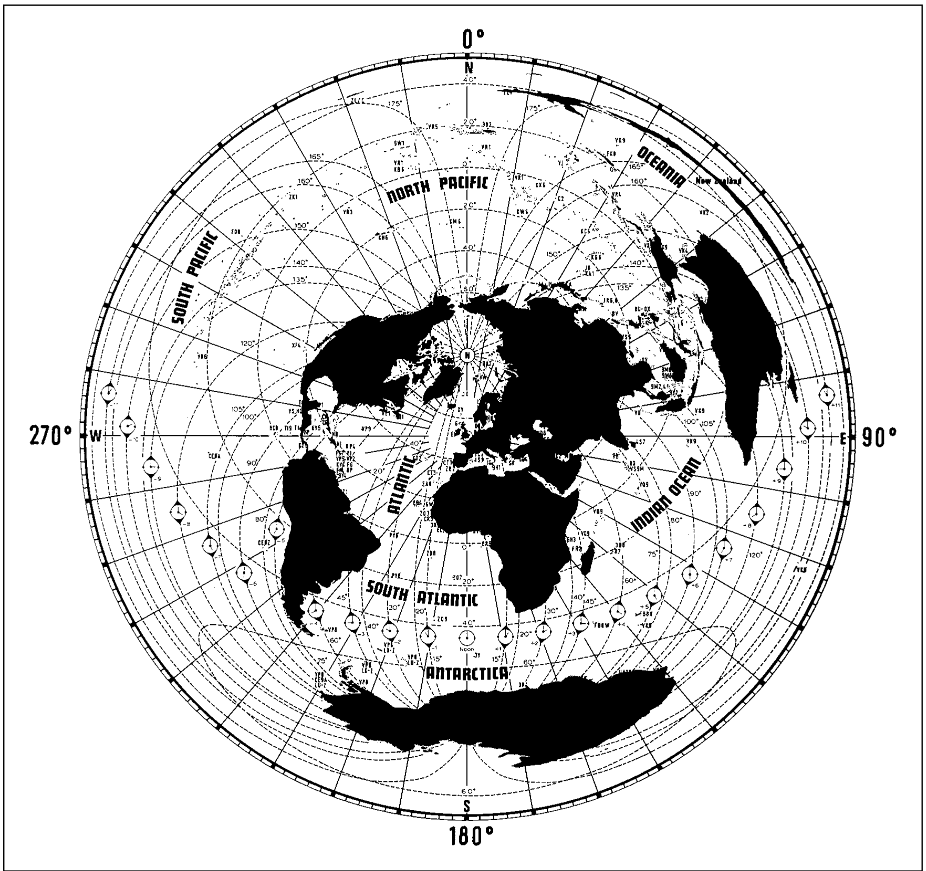


Fig 12.22: Example of an azimuthal equidistant (or great-circle) map. This map, available in a large size suitable for wall-mounting, shows the true bearing and distance from London of any place elsewhere in the world. (For magnetic bearings add 6° to the true bearing)

Projections can be divided into three groups: those which show areas correctly, described, logically enough, as *equal-area projections*; those which show the shapes of small areas correctly, known as *orthomorphic* or *conformal projections*; and those which represent neither shape nor area correctly, but which have some other property which meets a particular need.

The conformal group, useful for atlas maps generally, weather charts, satellite tracks etc includes the following:

Stereographic, where latitudes and longitudes are all either straight lines or arcs of circles, formed by projection on to a plane surface tangent either to one of the poles (*polar*), the equator (*equatorial*), or somewhere intermediate (*oblique*). Small circles on the globe remain circles on the map but the scale increases with increasing radius from the centre of projection.

Lambert's conformal conic, where all meridians are straight lines and all parallels are circles. It is formed by projection on to a cone whose axis passes through the Earth's poles.

Mercator's, where meridians and parallels are straight lines intersecting at right-angles. The meridians are equidistant, but the parallels are spaced at intervals which rapidly increase with latitude. It is formed by projection on to a cylinder which touches the globe at the equator. Any straight line is a line of constant bearing (a *rhumb line*, not the same thing as a *great circle*, which is the path a radio wave takes between two given points on the Earth's surface). There is a scale distortion which gets progressively more severe away from the equator, to such an extent that it becomes impossible to show the poles, but most people accept these distortions as being normal, because this is the best-known of all the projections.

Transverse Mercator is a modification of the 'classical' system, and is formed by projection on to a cylinder which touches the globe along selected opposing meridians. It therefore corresponds to an ordinary Mercator turned through 90 degrees, and is of value for displaying an area which is extensive in latitude but limited in longitude. A variant is the *universal transverse Mercator*, which forms the basis of a number of reference grids, including the one used on British Ordnance Survey maps.

The equal-area group is used when it is necessary to display the relative distribution of something, generally on a worldwide scale. It includes the following.

Azimuthal equal-area projection, having radial symmetry about the centre, which may be at either pole (*polar*), at the equator (*equatorial*), or intermediate (*oblique*). With this system the entire globe can be shown in a circular map, but there is severe distortion towards its periphery.

Mollweide's homolographic projection, where the central meridian is straight and the others elliptical.

Sinusoidal projection, where the central meridian is straight and the others parts of sine curves.

Homolosine projection, which is a combination of the previous two, with an irregular outline because of interruptions which are generally arranged to occur over ocean areas.

The final group includes the two following, which are of particular interest in propagation studies.

Azimuthal equidistant, centred on a particular place, from whence all straight lines are great circles at their true azimuths. The scale is constant and linear along any radius. Well known as a *great-circle map*, **Fig 12.22**.

Gnomonic, constructed by projection from a point at the Earth's centre on to a tangent plane touching the globe. Any straight line on the map is a great circle. Because the size of the map expands very rapidly with increasing distance from the centre they do not normally cover a large area. Often produced as a skeleton map on which a great circle can be drawn and used to provide a series of latitudes and longitudes by which the path can be replotted on a more detailed map based on a different projection.

Beam Heading and Locators

The shortest distance between two points on the surface of the Earth lies over the great circle that passes through them. This is easily done if you have a globe. All you need is to join the two locations with a tightly stretched thread. The shortest length of thread that can do this will be the great circle route and indicate the requisite beam heading to direct your signal the most effectively. (Irregularities in the distribution of ionization may on occasion result in signals being diverted off the great circle, but this should be understood as an exception to the general rule.). Alternatively, a great circle map along the lines of Fig 12.22 indicates optimum beam headings. For most working purposes the map works well for most UK operators. Operators far removed from London, or who simply wish to be more exact, can readily find freeware programs on the Internet that will supply beam headings and distances for any location.

Great-circle Calculations

It is sometimes useful to be able to calculate the great-circle bearing and the distance of one point from another, and the expressions which follow enable this to be done.

First label the two points A and B.

Then let L_a = latitude of point A

L_b = latitude of point B

L_0 = the difference in longitude between A and B

C = the direction of B from A, in degrees east or west from north in the northern hemisphere, or from south in the southern hemisphere.

D = the angle of arc between A and B.

It follows that:

$$\cos D = \sin L_a . \sin L_b + \cos L_a . \cos L_b . \cos L_0$$

D can be converted to distance, knowing that:

1 degree of arc = 111.2km or 69.06 miles

1 minute of arc = 1.853km or 1.151 miles

Once D is known (in angle of arc), then:

$$\cos C = \frac{\sin L_b - \sin L_a . \cos D}{\cos L_a . \sin D}$$

Note:

1. For stations in the northern hemisphere call latitudes positive.
2. For stations in the southern hemisphere call latitudes negative.
3. Cos L_a and cos L_b are always positive.
4. Cos L_0 is positive between 0 and 90°, negative between 90° and 180°.
5. Sin L_a and sin L_b are negative in the southern hemisphere.
6. The bearing for the reverse path can be found by transposing the letters on the two locations.

It is advisable to make estimates of the bearings on a globe, wherever possible, to ensure that they have been placed in the correct quadrant.

The IARU Locator

It is always nice to know where the other fellow lives and operators have been asking that question since the earliest days of operating. But all too often the answer has been something like '16k south' of some place your atlas does not deign to mention. That is particularly unsatisfactory if points in a contest or a personal distance record is at stake. Happily, the International Amateur Radio Union developed a system, originally called the Maidenhead Locator but now termed just Locator, that could be used anywhere in the world, in which information about the location is expressed in a group of six characters. The result defines the location to within 0.04 degrees of latitude and 0.08 degrees of longitude. This is sufficiently precise for most purposes; on the relatively rare occasions where greater exactitude is required an eight-character version is available.

The *RSGB Yearbook* contains a map of grid squares in Europe. Several programs for converting latitude and longitude data into 'grid locators' (and conversely) anywhere in the world, as well as the distance between any two points, are readily available on the Internet.

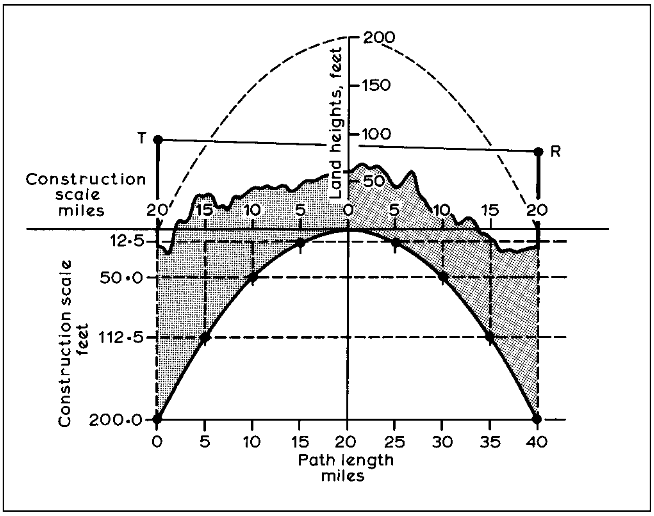


Fig 12.23: Construction of a 'four-thirds-earth' profile from data in Table 12.9. Land heights are measured upward from the lower curve. On this diagram, rays subjected to 'normal' variations of refractive index with height may be represented by straight lines

Plotting Path Profiles for VHF/UHF Working

For a tropospheric propagation study, it is standard practice to construct a path profile showing the curvature of the Earth as though its radius was four-thirds of its true value. This is so that, under standard conditions of refraction, the ray path may be shown as a straight line relative to the ups and downs of the intervening terrain.

It will be found convenient to construct the baseline of the chart using feet for height and miles for distance, for the reason that, by a happy coincidence, the various factors then cancel, leaving a very simple relationship that demands little more than mental arithmetic to handle.

Suppose that a profile is required for a given path 40 miles in length. First, decide on suitable scales for your type of graph paper (Fig 12.23 was drawn originally with 1in on paper representing 100ft, height, and 10 miles, distance). Then construct a sea-level datum curve, taking the centre of the path as zero and working downwards and outwards for half the overall distance in each direction, using the expression $h = D^2/2$ to calculate points on the curve (Table 12.9). Remember that this works only if h is in feet and D is in miles.

Then prepare a height scale. This will be the construction scale reversed if the contours of your map are in feet, or its metric equivalent if it is one of the more recent surveys. Usually it will be sufficient to limit this scale to just the contour values encountered along the path. Plot the heights, corresponding to the contours, vertically above the datum curve (a pair of dividers may be found helpful here) and add the antenna heights at the

D (miles)	Horizontal scale (inches)	D^2	$D^2/2 = h$	Vertical scale (inches)
5	0.5	25	12.5	0.125
10	1.0	100	50.0	0.500
15	1.5	225	112.5	1.125
20	2.0	400	200.0	2.000

Table 12.9: Points on the curve drawn in Fig 12.23

two terminal points. Draw a line through all the points and you are ready for business.

If you can draw a straight line from transmitting antenna to receiving antenna without meeting any obstacles along the way, then your path should be clear under normal conditions of refraction.

Ionospheric Predictions

The quality of a radio circuit is highest when it is operated at a frequency just below the maximum usable frequency (MUF) for the path. Three regions, E, F1 and F2, are considered in the determination of MUF.

For the F2-layer, which is responsible for most HF long-distance contacts, the MUF for paths less than 4000km is taken as being the MUF that applies to the mid-point. For paths longer than 4000km, the MUF for the path is the lower of the MUFs at the two ends along the path direction. For this purpose the end point locations are not those of the terminals, but of their associated *control points* where low-angle radiation from a transmitter would reach the F2-layer. Those points are taken to be 2000km from each terminal, along the great circle joining them.

To put this into perspective, a control point for a station located in the midlands of England would lie somewhere above a circle passing close by Narvik, Leningrad, Minsk, Budapest, Sicily, Algiers, Tangier, mid-Atlantic and NW Iceland, the place depending on the direction of take-off. It means that (for example) the steep rise in MUF associated with UK-end sunrise will occur something like four hours earlier on a path to the east than on a path to the west.

Taken over a month, the day-by-day path MUF at a given time can vary over a considerable range. A ratio of 2:1, as between maximum and minimum, could be considered as typical. Monthly predictions are based on median values, that is to say, on those values which have as many cases above them as below. Therefore the median MUFs in predictions represent 50% probabilities because, by definition, for half of the days of the month the operating frequency would be too high to be returned by the ionosphere. Note, though, that however certain you might be that a given circuit at a given frequency at a given time might be open 15 days in a particular month (the meaning of 50% probability in this context), there is no way of knowing which 15 days they might be. Nor can you be sure that 'good days' on one circuit will be equally rewarding on another.

If you want to ensure the most reliable communication over a particular path, say to make a daily contact at a given time, you will need to operate below the median MUF. Conventionally this optimum working frequency (OWF), at which contact should be possible on 90 per cent of days in the month, is found by multiplying the MUF figure by 0.85. Thus, for a monthly MUF of 20MHz the OWF will be around 17MHz. At the other end of the scale, there will be days when the monthly MUF figure is exceeded. This highest path frequency (HPF) is found by dividing MUF by 0.85 - in the example given earlier this would give a figure around 23.5MHz. Always remember that signals will propagate best if you are operating on the band closest to the operational MUF on any particular day.

All this may seem to suggest that the lower the operating frequency, the greater would be the chance of success, but that is not true. The reason is that the further down from the median MUF that one operates, the greater become the losses due to absorption, and eventually these become the dominant factor. For any given path there is a *lowest useful frequency* (LUF) at which those losses become intolerable.

Unfortunately it is beyond the scope of this book to be able to provide detailed information about the preparation of monthly predictions and about how to relate the values to the amateur bands. However, each issue of *RadCom* contains a table giving current data in a form that will satisfy the needs of most operators based in the UK.

Those predictions are unique in that the information is given in the form of percentage probability for each of the HF amateur bands, taking both the HPF and the LUF into account. A figure of 1 represents 10% (or three days a month), a figure of 5 represents 50% (or 15 days a month), and so on. Multiply by three the figure given in the appropriate column and row and you will have the expected number of days in the month on which communication should be possible at a given time on a given band.

Many operators believe that if it is known that solar activity is higher (or lower) than was expected when the predictions were prepared, then they can raise (or lower) all the probabilities by a fixed amount to compensate for the changed circumstances. That is not so. The highest probability will always appear against the band closest to the OWF (as defined in the predictions sense). If the OWF is altered by fresh information, then the probabilities on one side of it will rise but on the other side will fall. The only sure way of establishing amended figures would be to enter the revised particulars into the computer program and to run off a complete set of new predictions. It is beyond the scope of this book to provide detailed descriptions of how to prepare predictions and relate them to the amateur bands. However, each issue of *RadCom* contains a table giving predictions for the month ahead for a wide range of paths, and these should suffice for most people's purposes. For those wishing to prepare your own predictions, a number of useful freeware or shareware programs are available on the Internet. These prediction programs customarily assume a quiet geomagnetic field.

Monthly prediction programs usually require an index of solar activity comparable to the 12-month smoothed relative sunspot number R_{12}, or to IF2, which is derived from ionospheric data. There is nothing to be gained by substituting the latest unsmoothed monthly sunspot figure. Also, the use of raw daily figures, such as those from Boulder or Meudon, in the expectation that they will yield meaningful daily ionospheric predictions, is a misunderstanding of the highest degree. The ionosphere does not respond to fluctuations of daily sunspot numbers. On a daily time scale signal performance is much more responsive to changes in the level of geomagnetic activity. Information about current levels of geomagnetic activity and short-term forecasts of the 'radio weather" is available from several Internet sites.

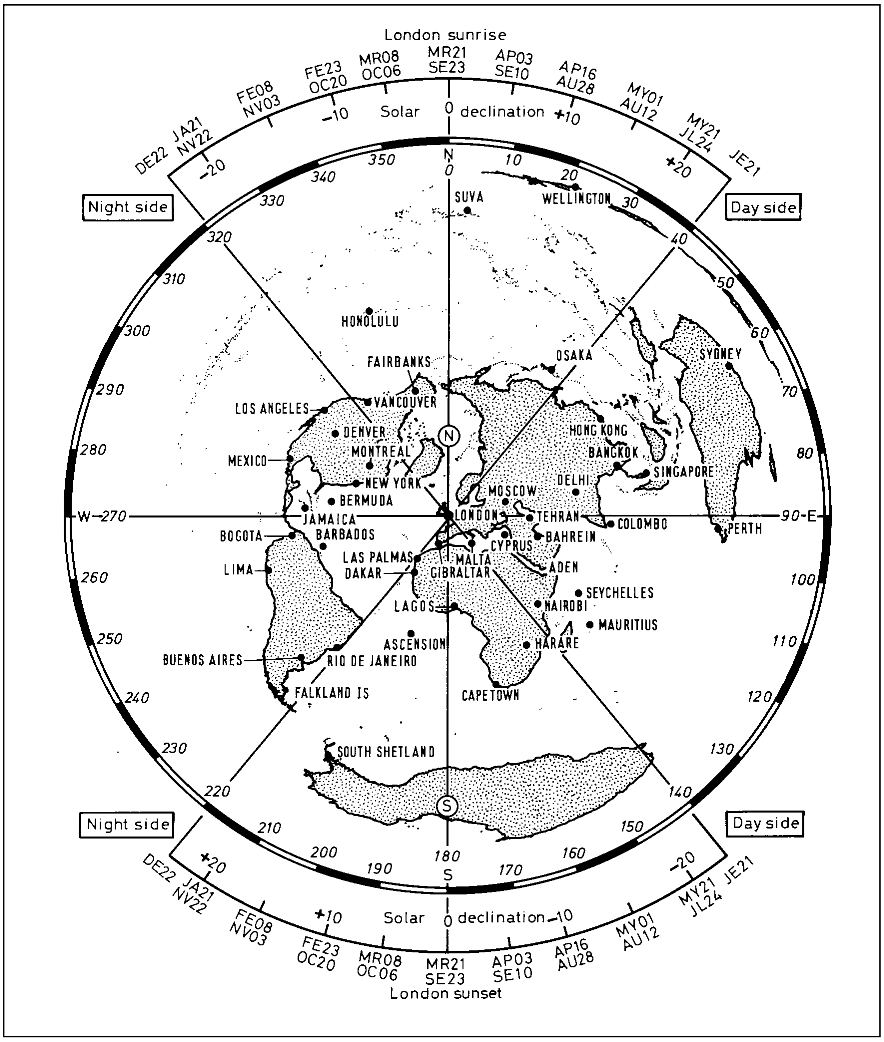
The locations of places for which predictions are given every month in *RadCom* may be seen against their relative beam headings in the great-circle outline map of Fig 12.22.

Grey-line Propagation

The grey line is the ground-based boundary around the world that separates day from night, sunlight from shadow.

Many operators believe, and with some justification, that signals beamed along the grey line near sunrise or sunset will

Fig 12.24: Grey-line propagation map



reach distant locations that are also experiencing sunrise or sunset for relatively short periods of time when conventional predictions may appear to be pessimistic.

Noon is a north-south phenomenon; all places having the same longitude encounter it at the same instant of time. But the grey line runs north and south only at the equinoxes, in March and September. At all other times it cuts across the entire range of meridians and time zones.

Sunrise and sunset are the periods when MUFs may be expected to rise or fall through their greatest range of the day.

Couple that with the fact that the one grey line represents sunrise (rising MUF) on one side of the world and sunset (falling MUF) on the other, spring or summer on one side of the equator and autumn or winter elsewhere, and you have a very strong prospect that favourable circumstances for propagation will occur somewhere along the line.

This mode is for the lower frequency bands, 1.8, 3.5 and 7MHz, because the normal house rules apply: the operating frequency has to be below the MUF for the path or the signals will pass through the ionosphere somewhere instead of being reflected.

On the ground the grey line may be considered to be a great circle, but a rather badly defined one because the Sun is not a point source of light - hence twilight time, of course.

To work the grey line you need to know three things: your latitude (atlas), sunrise/sunset times for your area (daily newspaper), and the declination of the Sun for the day in question (*Whittaker's Almanack* or the *British Astronomical Association's Handbook*).

The ground azimuth at sunrise is given by the expression

$$\sin \text{azimuth} = \frac{\sin \text{declination:}}{\cos \text{latitude}}$$

and the ground azimuth at sunset by:

$$\sin \text{azimuth} = \frac{-\sin \text{declination}}{\cos \text{latitude}}$$

and their reciprocals. Those equations are the basis for the sunrise and sunset scales above and below the skeleton great-circle map, Fig 12.24. Interpolate for the appropriate date and lay a straight-edge right across the map so that it passes through the centre (London, in this case). That is your grey line and it should work in either direction, if it is to work at all. Your line will show which areas ought to be accessible if fortune smiles upon you.

That is the good news. The bad news is that it is not the ground-based shadow that determines the state of the ionosphere because the Earth's shadow is shaped like a cone; the area of darkness at F2 heights is appreciably less than has been considered in the preceding paragraph. Sunrise comes earlier than on the ground, sunset later. In fact, in mid-summer the F2-layer is in sunlight for all 24 hours of the day over the whole of the UK.

So, the ionospheric grey line (as opposed to the ground-based grey line) cannot be considered as a great circle and, therefore, cannot be represented by the straight line on the great-circle map.

That should not stop you from trying your luck, however. But there is no point in trying to calculate the true outline of the Earth's shadow, because your signals are going to take the great-circle route no matter what you come up with on your computer. You may as well make all your plans using your ground-based data, because it is easy to come by, and make up for its likely deficiencies by being generous with your timing. At the sort of frequency you will be using, the beamwidth of the antenna will be wide enough to take care of direction.

The Beacon Network

One of the most useful aids to understanding propagation and exploiting whatever possibilities there may be effectively is offered by beacon transmissions. These also offer a useful basis for personal propagation research projects.

Beacons are found mainly on the higher HF bands upwards. For the most part they operate continuously with a simple message in Morse.

This will always include their callsign and a long dash to facilitate signal strength measurements; many also include their grid locator, town and power. Because, unlike other amateur stations, they are always there they offer a useful indication of the state of a particular path.

This is particularly true at HF of the network of beacons created by the North California DX Foundation. This consists of 18 beacons, strategically dispersed around the world, transmitting in sequence on 14.100, 18.110, 21.150, 24.930 and 28,200MHz in turn in the course of a three-minute cycle. The power is stepped down from 100W to 100mW in four stages, giving a further indication of the state of the circuit.

There are literally hundreds more beacons worldwide at HF and the VHF and UHF bands. They are far too numerous to list here but details of those likely to be heard in the UK are listed in the *RSGB Yearbook* and changes between editions can be found on the Internet.

GB2RS News

Finally, every week the weekly *GB2RS* news bulletin, posted every Friday on the RSGB website and broadcast every Sunday, contains items of propagation interest, including a summary of solar-geophysical events over the preceding week. Times and frequencies of these broadcasts vary according to location; details are printed in the *RSGB Yearbook*.

BIBLIOGRAPHY

Books

Radio Auroras, C Newton, G2FKZ, RSGB, 1991

Radio Propagation - Principles & Practice, Ian Poole, G3YWX, RSGB, 2004

Your Guide to Propagation, by Ian Poole, G3YWX, RSGB, 1998

LF Today, Mike Dennison, G3XDV, RSGB, 2004

The VHF/UHF Handbook, edited by Dick Biddulph, M0CGN, RSGB

Internet

The Radio Propagation Page: http://www.keele.ac.uk/depts/por/psc.htm. Run by the RSGB Propagation Studies Committee this carries links to a wide range of sources, ranging from the explanation of basic terms in radio propagation through the various forms of propagation to sites carrying data affecting propagation to HF prediction programs.

Amateur Radio Propagation Studies: http://www.df5ai.net

Basics of Radio Wave Propagation: http://ecjones.org/physics.html

Bouncing Radio Waves off the Sky: http://www.geocities.com/rf-man/skyrange.html

Glossary on Solar-Terrestrial Terms: www.ips.gov.au/Main.php?CatID=8

Introduction to the Ionosphere: http://www.ngdc.noaa.gov/stp/IONO/ionointro.html

Near Real-Time Global MUF Map: http://www.spacew.com/www/realtime.php

Real-Time Space Weather: http://www.spacew.com/plots.htm

SEC Radio Users' Page: http://www.sel.noaa.gov/radio.html

Solar Terrestrial Activity Report: http://www.dxlc.com/solar

Space Environment Center: http://www.sec.noaa.gov

Spaceweather.com: http://www.spaceweather.com

Search engines

A multitude of information on all aspects of radio propagation can be gleaned from a simple search request to any of the more popular Internet search engines.

REFERENCES

[1] 'Flare spot', P W Sollom, *Radio Communication* December 1970, p820; January 1971, p20; February 1971, p92

[2] 'The ground beneath us', R C Hills, *RSGB Bulletin* June 1966, p375

[3] Report 239-6 in *Propagation in non-ionised media*, Study Group 5, CCIR, XVIth Plenary Assembly, Dubrovnik, 1986, Geneva, ITU, 1986

[4] *VHF/UHF Handbook*, Dick Biddulph (ed), RSGB, 1997, Chapter 3 (Propagation)

[5] 'Patterns in propagation', R G Flavell, *Journal of the IERE*, Vol 56 No 6 (Supplement), pp175-184