Electromagnetic Waves Series 502

Propagation of Radiowaves

2nd Edition

Edited by Les Barclay

IET ELECTROMAGNETIC WAVES SERIES 502

Series Editors: Professor P.J.B. Clarricoats
Professor E.V. Jull

Propagation of Radiowaves

2nd Edition

Other volumes in this series:

Volume 1	**Geometrical theory of diffraction for electromagnetic waves, 3rd edition** G.L. James
Volume 10	**Aperture antennas and diffraction theory** E.V. Jull
Volume 11	**Adaptive array principles** J.E. Hudson
Volume 12	**Microstrip antenna theory and design** J.R. James, P.S. Hall and C. Wood
Volume 15	**The handbook of antenna design, volume 1** A.W. Rudge, K. Milne, A.D. Oliver and P. Knight (Editors)
Volume 16	**The handbook of antenna design, volume 2** A.W. Rudge, K. Milne, A.D. Oliver and P. Knight (Editors)
Volume 18	**Corrugated horns for microwave antennas** P.J.B. Clarricoats and A.D. Oliver
Volume 19	**Microwave antenna theory and design** S. Silver (Editor)
Volume 21	**Waveguide handbook** N. Marcuvitz
Volume 23	**Ferrites at microwave frequencies** A.J. Baden Fuller
Volume 24	**Propagation of short radio waves** D.E. Kerr (Editor)
Volume 25	**Principles of microwave circuits** C.G. Montgomery, R.H. Dicke and E.M. Purcell (Editors)
Volume 26	**Spherical near-field antenna measurements** J.E. Hansen (Editor)
Volume 28	**Handbook of microstrip antennas, 2 volumes** J.R. James and P.S. Hall (Editors)
Volume 31	**Ionospheric radio** K. Davies
Volume 32	**Electromagnetic waveguides: theory and application** S.F. Mahmoud
Volume 33	**Radio direction finding and superresolution, 2nd edition** P.J.D. Gething
Volume 34	**Electrodynamic theory of superconductors** S.A. Zhou
Volume 35	**VHF and UHF antennas** R.A. Burberry
Volume 36	**Propagation, scattering and diffraction of electromagnetic waves** A.S. Ilyinski, G. Ya.Slepyan and A. Ya.Slepyan
Volume 37	**Geometrical theory of diffraction** V.A. Borovikov and B.Ye. Kinber
Volume 38	**Analysis of metallic antenna and scatterers** B.D. Popovic and B.M. Kolundzija
Volume 39	**Microwave horns and feeds** A.D. Olver, P.J.B. Clarricoats, A.A. Kishk and L. Shafai
Volume 41	**Approximate boundary conditions in electromagnetics** T.B.A. Senior and J.L. Volakis
Volume 42	**Spectral theory and excitation of open structures** V.P. Shestopalov and Y. Shestopalov
Volume 43	**Open electromagnetic waveguides** T. Rozzi and M. Mongiardo
Volume 44	**Theory of nonuniform waveguides: the cross-section method** B.Z. Katsenelenbaum, L. Mercader Del Rio, M. Pereyaslavets, M. Sorella Ayza and M.K.A. Thumm
Volume 45	**Parabolic equation methods for electromagnetic wave propagation** M. Levy
Volume 46	**Advanced electromagnetic analysis of passive and active planar structures** T. Rozzi and M. Farinai
Volume 47	**Electromagnetic mixing formulas and applications** A. Sihvola
Volume 48	**Theory and design of microwave filters** I.C. Hunter
Volume 49	**Handbook of ridge waveguides and passive components** J. Helszajn
Volume 50	**Channels, propagation and antennas for mobile communications** R. Vaughan and J. Bach-Anderson
Volume 51	**Asymptotic and hybrid methods in electromagnetics** F. Molinet, I. Andronov and D. Bouche
Volume 52	**Thermal microwave radiation: applications for remote sensing** C. Matzler (Editor)
Volume 53	**Principles of planar near-field antenna measurements** S. Gregson, J. McCormick and C. Parini
Volume 502	**Propagation of radiowaves, 2nd edition** L.W. Barclay (Editor)

Propagation of Radiowaves
2nd Edition

Edited by Les Barclay

The Institution of Engineering and Technology

Published by The Institution of Engineering and Technology, London, United Kingdom

First published 1996 (0 85296 819 1)
Second edition 2003 (0 85296 102 2)

The Institution of Engineering and Technology
Michael Faraday House
Six Hills Way, Stevenage
Herts, SG1 2AY, United Kingdom

www.theiet.org

British Library Cataloguing in Publication Data
Propagation of radiowaves. – 2nd ed
1. Radiowave propagation
I. Barclay, L.W. (Les W) II. Institution of Electrical Engineers
621.3’8411

ISBN (10 digit) 0 85296 102 2
ISBN (13 digit) 978-0-85296-102-5

Typeset in the UK by RefineCatch Ltd, Bungay, Suffolk
First printed in the UK by MPG Books Ltd, Bodmin, Cornwall
Reprinted in the UK by Lightning Source UK Ltd, Milton Keynes

Contents

Preface

This edition of the book 'Propagation of Radiowaves' is an update of the previous edition and includes descriptions of new findings and of new and improved propagation models. It is based on the series of lectures given at the 8th IEE residential course held at Cambridge in 2000.

The preface to the previous edition noted that the topics of main interest in radio communication continue to change rapidly, with the consequent need for new questions to be addressed, arising from the planning of new services. The pressures to provide information for new applications, for improved service quality and performance, and for more effective use of the radio spectrum, require a wider understanding of radiowave propagation and the development of improved prediction methods.

The intention of this book, and of the series of courses on which it is based, is to emphasise propagation engineering, giving sufficient fundamental information, but going on to describe the use of these principles, together with research results, in the formulation of prediction models and planning tools.

The lecture course was designed and organised by a committee comprising D F Bacon, K H Craig, M T Hewitt and L W Barclay and thanks are due to this committee for arranging a comprehensive and timely programme. Thanks are also particularly due to the group of expert lecturers who prepared and delivered the course material contained in this book.

Study Group 3 of the International Telecommunication Union, Radiocommunication Sector is devoted to studies of radiowave propagation. The annual meetings of its working parties aim to include the most recent information into revisions and updates of its Recommendations for propagation models, etc. The course and this book emphasise the applications of these ITU-R Recommendations, since these provide the most up to date, peer-reviewed information available.

ITU-R Recommendations are available in several series with a prefix letter to indicate the series and a suffix number to indicate the version. Thus Recommendation ITU-R P.341–7 refers to the 7th revision of Recommendation number 341 in the propagation series. Before using such Recommendations, it will be desirable to refer to the ITU web site (www.itu.int) (correct at time of publication) to check that the version to be used is in fact the latest available revision.

Chapter 1

Radiowave propagation and spectrum use

Martin Hall and Les Barclay

1.1 Introduction

When Heinrich Hertz undertook his experiments to verify that radiowaves were electromagnetic radiation which behaved as expected from the theory developed by Maxwell, he probably used frequencies between 50 and 500 MHz. He selected the frequency by adjusting the size of the radiating structure, and chose it so that he could observe the propagation effects of reflection and refraction within his laboratory. The first public demonstration of a communication system by Oliver Lodge also probably used a frequency in the VHF range and signals were propagated about 60 m into the lecture hall. Marconi and others took up this idea, and Marconi in particular, by increasing the size of the antennas, reduced the frequency and was able to exploit the better long-distance propagation properties at progressively lower frequencies.

From the beginning it has been the practical use of the propagation of electromagnetic waves over long distances, together with the ability to modulate the waves and thus transfer information, which has provided the opportunity for the development of radio and electronic technologies. This in turn has driven a need to extend knowledge of the propagation environment, and to characterise the transfer function of the radio channel to provide greater communication bandwidths and greater quality of service.

Propagation in free space, or in a uniform dielectric medium, may be described simply. It is the effect of the earth and its surrounding environment which leads to variability and distortion of the radio signal, and which provides the challenge for the propagation engineer. He seeks to provide a detailed description of the signal and a prediction capability for use in the design, planning and operation of radio systems.

1.2 The radio spectrum

Electromagnetic waves may exist with an extremely wide range of wavelengths, and with corresponding frequencies, where the product of frequency and wavelength is the velocity of propagation. In vacuum or air this is very close to 3×10^8 m s^{-1}.

It is convenient to consider that the radio spectrum lies in the range of frequencies between, say, 3 kHz and 3 THz, although significant use is within the range 10 kHz to 275 GHz, the range covered by frequency allocations in the ITU Radio Regulations. The conventional nomenclature for the spectrum is summarised in Table 1.1.

Certain frequency bands are sometimes designated by letter symbols, although these are not recommended for use. The most common of these are listed in Table 1.2.

1.3 Radio services

The various uses of radiocommunication are defined in the ITU Radio Regulations as a number of radio services. From a regulatory viewpoint, each of these has a different requirement for spectrum usage and for protection against interference. A summary of these services is given in Figure 1.1.

Table 1.1 Nomenclature for frequency bands

Band no.	*Symbol*	*Frequency range*	*Wavelength*	*Corresponding metric subdivision of wavebands*	*Symbol*
	ELF	<300 Hz	>1000 km		
3	ULF	300 Hz–3 kHz	1000–100 km	Hectokilometric	B.hkm
4	VLF	3–30 kHz	100–10 km	Myriametric	B.Mam
5	LF	30–300 kHz	10–1 km	Kilometric	B.km
6	MF	300 kHz–3 MHz	1 km–100 m	Hectometric	B.hm
7	HF	3–30 MHz	100–10 m	Decametric	B.dam
8	VHF	30–300 MHz	10–1 m	Metric	B.m
9	UHF	300 MHz–3 GHz	1 m–100 mm	Decimetric	B.dm
10	SHF	3–30 GHz	100–10 mm	Centimetric	B.cm
11	EHF	30–300 GHz	10–1 mm	Millimetric	B.mm
12		300 GHz–3 THz	1 mm–100 μm	Decimillimetric	B.dmm
13		3–30 THz	100–10 μm	Centimillimetric	B.cmm
14		30–300 THz	10–1 μm	Micrometric	B.μm
		300–3000 THz	1–0.1 μm	Decimicrometric	B.dμm

Note 1: band number N (N = band number) extends from 0.3×10^N to 3×10^N Hz
Note 2: prefix: k = kilo (10^3); M = mega (10^6); G = giga (10^9); T = tera (10^{12}); m = milli (10^{-3}); μ = micro (10^{-6})

Table 1.2 Letter designations for some frequency bands (as listed in Recommendation ITU-R V.431)

	Radar (GHz)		Space radiocommunications	
Letter symbol	Spectrum region (GHz)	Examples (GHz)	Nominal designations	Examples (GHz)
L	1–2	1.215–1.4	1.5 GHz band	1.525–1.710
S	2–4	2.3–2.5	2.5 GHz band	2.5–2.690
		2.7–3.4		
C	4–8	5.25–5.85	4/6 GHz band	3.4–4.2
				4.5–4.8
				5.85–7.075
X	8–12	8.5–10.5		
Ku	12–18	13.4–14.0	11/14 GHz band	10.7–13.25
		15.3–17.3	12/14 GHz band	14.0–14.5
K[1]	18–27	24.05–24.25	20 GHz band	17.7–20.2
Ka[1]	27–40	33.4–36.0	30 GHz band	27.5–30.0
V			40 GHz bands	37.5–42.5
				47.2–50.2

[1] For space radiocommunications K and Ka bands are often designated by the single symbol Ka

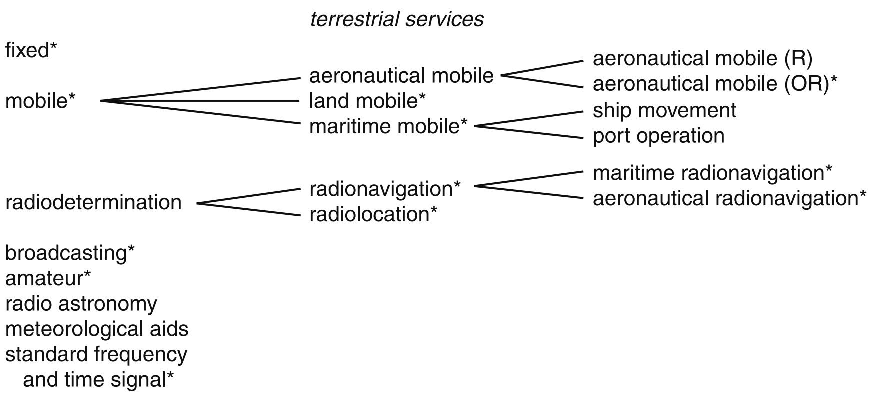


Figure 1.1 ITU-defined radio services (marks those with an equivalent satellite service)*

Each of the services in the figure marked with an asterisk also has an equivalent satellite service (e.g. the fixed-satellite service corresponding to the fixed {i.e. terrestrial} service). In addition there are some services which only have a satellite context. These are:

earth exploration satellite
meteorological satellite
inter-satellite

space operations
space research

Recent developments in the use of radio for a variety of applications means that this division into services may become increasingly less appropriate. It may be better to distinguish uses by the precision in the identification of the location of the radio link terminals, by the ability to exploit directional antennas, by the height of the antennas in relation to the surrounding buildings or ground features, or by the required bandwidth etc.

1.4 The propagation environment

Except for the inter-satellite service, where the propagation path may be entirely in near free-space conditions, propagation for all radio applications may be affected by the earth and its surrounding atmosphere.

The upper atmosphere has a temperature profile as sketched in Figure 1.2. Of particular interest are the troposphere and the variations in atmospheric temperature, pressure and humidity, which are largely confined below the temperature minimum at the tropopause, and the ionosphere which is largely above about 80 km in the thermosphere.

Propagation within the troposphere, which is of most importance owing to

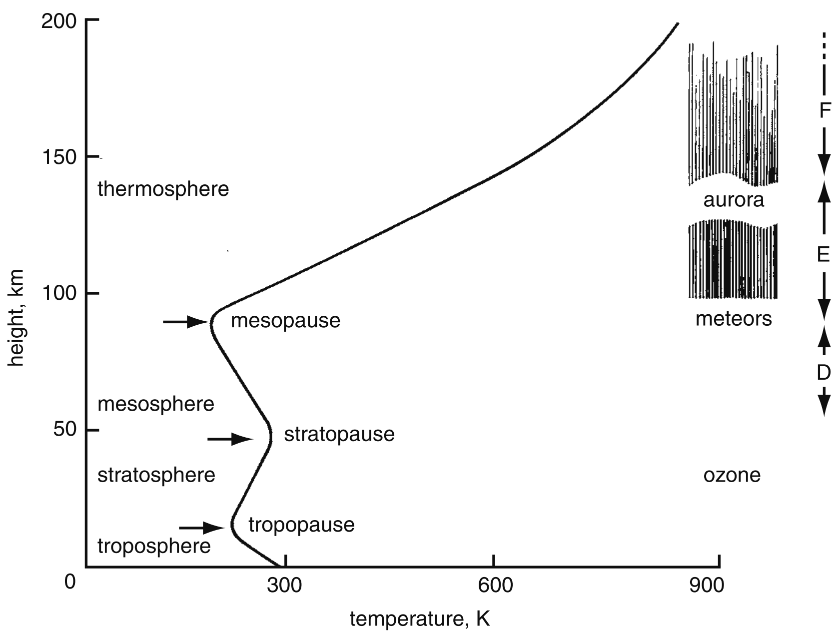


Figure 1.2 Regions of the earth's atmosphere, showing the mean temperature profile and approximate heights of the lettered ionospheric regions and of other features

the wide variety of uses and the very wide available bandwidths at higher frequencies, is complex due to irregularities in the refractive index profile and to the presence of rain and other hydrometeors. The effects are summarised diagrammatically in Figure 1.3.

The profile of electron density in the ionosphere acts as a reflecting layer capable of reflecting signals at HF and lower frequencies back to earth. There are occasional effects which permit some reflection or scatter back to earth at VHF. The ionosphere also has some important effects on earth–space paths up to SHF.

In addition, diffraction, reflection and scatter in relation to the ground and both man-made and natural structures on the surface are of great importance.

All of these aspects will be addressed in subsequent chapters.

1.5 Spectrum use

It is useful to review the way in which each part of the spectrum is used. Although the conventional way of describing the spectrum in decade frequency bands does not match the applications or the propagation characteristics very well, it is used here as a convenient shorthand. None of the frequency boundaries indicated is clear and precise in terms of differing usage.

1.5.1 ELF and ULF (below 3 kHz) and VLF (3–30 kHz)

Typical services:	worldwide telegraphy to ships and submarines; worldwide communication, mine and subterranean communication;
System considerations:	even the largest antennas have a size which is only a small fraction of a wavelength, resulting in low radiation resistance; it is difficult to make transmitter antennas directional; bandwidth is very limited, resulting in only low or very low data rates; there is high atmospheric noise so that inefficient receiving antennas are satisfactory
Propagation:	in earth–ionosphere waveguide, with relatively stable propagation; affected by thick ice masses (e.g. Greenland); propagation east/west and west/east is asymmetric; useful propagation through sea water, which has significant skin depth at these wavelengths
Comment:	there are no international frequency allocations below 9 kHz; there is limited use of frequencies below 9 kHz for military purposes; the successful Omega worldwide navigation system has recently been closed

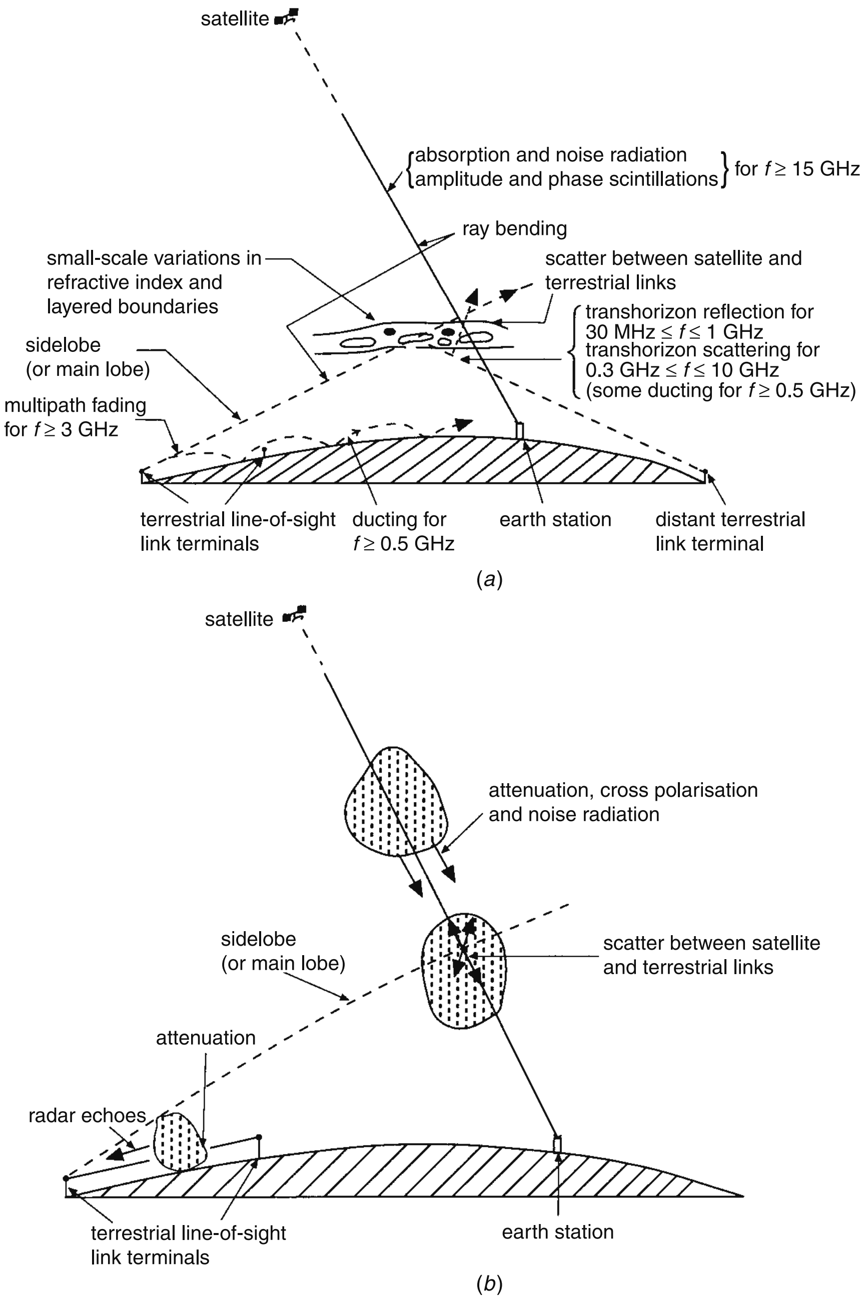


Figure 1.3 Some effects of the troposphere on radiowave propagation
(a) effects of atmospheric gases and the associated changes in refractive index
(b) effects of cloud and precipitation (SHF and above)

1.5.2 LF (30–300 kHz)

Typical services:	long-distance shore-to-ship communication; fixed services over continental distances; broadcasting; time signals
System considerations:	vertical polarisation used (for ground-wave propagation, and for antenna efficiency); efficient but large antennas are possible; directional antennas are very large; high atmospheric noise; limited bandwidth
Propagation:	up to several thousand km; ground wave, strong sky wave at night, slow fading

1.5.3 MF (300 kHz–3 MHz)

Typical services:	broadcasting; radionavigation; maritime mobile communications
System considerations:	1/4 wavelength vertical antenna at 1 MHz is 75 m high; directional antennas possible, magnetic receiving antennas
Propagation:	ground wave more pronounced over sea; strong sky wave absorption during the day, but little absorption at night; high atmospheric noise levels

1.5.4 HF (3–30 MHz)

Typical services:	international broadcasting, national broadcasting in tropical regions; long-distance point-to-point communications; aeronautical and maritime mobile communications
System considerations:	arrays of horizontal dipoles; log-periodic antennas (vertical or horizontal), vertical whip antennas; frequency agility essential; crowded spectrum needing good intermodulation performance; external noise environment varies with time and location; bandwidths up to about 6 kHz
Propagation:	propagation up to worldwide distances by ionospheric sky wave, very variable in time; propagation window between MUF and LUF (maximum and lowest usable frequencies) varies from a few MHz to about 20 MHz
Comment:	necessary to change the operating frequency several times during 24 hours; broadcasting uses seasonal schedules of frequencies; fixed and some mobile services

use intelligent frequency adaptive systems; continues to provide the main intercontinental air traffic control system; most modulation bandwidths may exceed the correlation bandwidth

1.5.5 VHF (30–300 MHz)

Typical services: land mobile for civil, military and emergency purposes, maritime and aeronautical mobile; sound (FM and DAB) and (outside UK) television broadcasting (to about 100 km); aeronautical radionavigation and landing systems; analogue cordless telephones; paging; very limited little LEO satellite systems

System considerations: multi-element dipole (yagi) antennas, rod antennas suitable for vehicle mounting, atmospheric noise small but man-made noise significant; some use for meteor burst communications

Propagation: usually by refraction in troposphere; reflections may cause multipath on line-of-sight paths; screening by major hills, but diffraction losses generally small; some anomalous propagation due to refractivity; unwanted ionospheric modes due to sporadic E and meteor scatter; substantial Faraday rotation and ionospheric scintillation on earth–space paths

1.5.6 UHF (300 MHz–3 GHz)

Typical services: television broadcasting; cellular and personal communications; satellite mobile; GPS; important radio astronomy bands; surveillance radars; terrestrial point-to-point service; radio fixed access; telemetry; cordless telephones; tropospheric scatter links

System considerations: small rod antennas; multi-element dipole (yagi) antennas; parabolic dishes for higher frequencies; wide bandwidths available

Propagation: line-of-sight and somewhat beyond; tropospheric scatter for transhorizon paths, screening by hills, buildings and trees; refraction effects; ducting possible; ionospheric scintillation on satellite paths

1.5.7 SHF (3–30 GHz)

Typical services: fixed (terrestrial point-to-point up to 155 Mb s^{-1}); fixed satellite; radar; satellite television; GSO and NGSO

	fixed satellite services; remote sensing from satellites; radio fixed access
System considerations:	high-gain parabolic dishes and horns; waveguides; major inter-service frequency sharing; wide bandwidths
Propagation:	severe screening; refraction and ducting; scintillation; rain attenuation and scatter increasing above about 10 GHz; atmospheric attenuation above about 15 GHz, ionospheric effects becoming small

1.5.8 EHF (30–300 GHz)

Typical services:	line-of sight communications, future satellite applications; remote sensing from satellites; broadband fixed wireless access; fixed service in the future using high-altitude platforms in the stratosphere
System considerations:	small highly directional antennas; equipment costs increase with frequency; little use at present above 60 GHz; very wide bandwidths; short range
Propagation:	severe difficulties: screening; atmospheric absorption; rain; fog; scintillation

1.6 Conclusions

This book, and the lecture course on which it is based, is intended to deal with the practical engineering aspects of radiowave propagation emphasising the propagation concerns and models and the associated prediction procedures which are appropriate for the system applications of current interest.

For many years a Radiocommunication Study Group of the International Telecommunication Union (ITU) has been studying propagation on a world-wide basis and producing the Recommendations which give descriptions and prediction techniques for the propagation of radiowaves. These Recommendations are regularly reviewed and revised by international experts in ITU-R Study Group 3 and probably represent the latest and best tools which the engineer may use. The course deals with a number of these Recommendations.

Chapter 2

Basic principles 1

Les Barclay

2.1 Basic radio system parameters

This Chapter introduces a number of topics which should be useful for the succeeding chapters. Antenna gain, radiated power and transmission loss are commonly used terms when describing systems, but the precision given by the internationally agreed definitions of these terms is necessary if ambiguity is to be avoided. System performance is governed not only by the transmission loss, under some stated conditions, but also by the variability of the signal in time or space, which can then be described in statistical terms, and by the level of background signals – either broadband noise or interfering transmissions. The statistical probability distributions in common use are introduced, and the benefits of diversity reception are outlined. The types of radio noise are described together with the ways in which noise power from a number of sources may be combined for use in performance prediction.

2.2 Propagation in free space

A transmitter with power p_t in free space which radiates isotropically (uniformly in all directions) gives a power flux density s at distance r of:

$$s = \frac{p_t}{4\pi r^2} \tag{2.1}$$

Using logarithmic ratios and practical units,

$$S = -41 + P_t - 20 \log d \tag{2.2}$$

where S is the power flux density in decibels relative to 1 W m^{-2}, P_t is the power in decibels relative to 1 kW and d is the distance in km.

The corresponding field strength e is given by:

$$e = \sqrt{120\pi s} = \frac{\sqrt{30p_t}}{r} \tag{2.3}$$

This relationship applies when the power is radiated isotropically.

A $\lambda/2$ dipole has a gain in its equatorial plane of 1.64 times (see below) and in this case the field strength is:

$$e \approx \frac{7\sqrt{p_t}}{r} \tag{2.4}$$

From the above, for free-space propagation, the signal intensity, or the field strength, decreases by 20 dB for each decade of increasing distance, or by 6 dB for each doubling of distance.

2.2.1 Polarisation

The concept of wave polarisation will be discussed further in Chapters 5 and 6. Suffice it to say here that an electromagnetic wave will have a characteristic polarisation, usually described by the plane of the electric field. For linear antennas in free space the polarisation plane will correspond to the plane containing the electric radiating element of the antenna. For the above expressions for signal intensity it is assumed that the receiving antenna is also oriented in the plane of polarisation.

2.3 Antenna gain

The ITU Radio Regulations formally define the gain of an antenna as: 'The ratio, usually expressed in decibels, of the power required at the input of a loss-free reference antenna to the power supplied to the input of the given antenna to produce, in a given direction, the same field strength or the same power flux density at the same distance'. When not specified otherwise, the gain refers to the direction of maximum radiation. The gain may be considered for a specified polarisation. Gain greater than unity (positive in terms of decibels) will increase the power radiated in a given direction and corresponds to an increase in the effective aperture of a receiving antenna.

Depending on the choice of the reference antenna, a distinction is made between:

(a) absolute or isotropic gain (G_i), when the reference antenna is an isotropic antenna isolated in space (note that isotropic radiation relates to an equal intensity in all directions; the term omnidirectional radiation is often used for an antenna which radiates equally at all azimuths in the horizontal plane, such an antenna may radiate with a different intensity for other elevation angles);

(b) gain relative to a halfwave dipole (G_d), when the reference antenna is a halfwave dipole isolated in space whose equatorial plane contains the given direction;
(c) gain relative to a short vertical antenna conductor (G_v) much shorter than one-quarter of the wavelength, on and normal to the surface of a perfectly conducting plane which contains the given direction.

An isotropic radiator is often adopted as the reference at microwaves and at HF, and a halfwave dipole is often adopted at VHF and UHF, where this type of antenna is convenient for practical implementation. A short vertical antenna over a conducting ground is an appropriate reference at MF and lower frequencies where ground-wave propagation is involved and this usage extends to sky-wave propagation at MF and, in older texts, at HF.

The comparative gains of these reference antennas, and of some other antenna types, are given in Table 2.1.

2.4 Effective radiated power

The Radio Regulations also provide definitions for effective or equivalent radiated power, again in relation to the three reference antennas:

(i) equivalent isotropically radiated power (EIRP): the product of the power supplied to the antenna and the antenna gain in a given direction relative to an isotropic antenna (absolute or isotropic gain); specification of an EIRP in decibels may be made using the symbol dBi;
(ii) effective radiated power (ERP): the product of the power supp1ied to the antenna and its gain in a given direction relative to a halfwave dipole;
(iii) effective monopole radiated power (EMRP): the product of the power

Table 2.1 Gain of typical reference antennas

Reference antenna	g_i	G_i* *(dB)*	*Cymomotive force (for a radiated power of 1 kW) (V)*
isotropic in free space	1	0	173
Hertzian dipole in free space	1.5	1.75	212
halfwave dipole in free space	1.64	2.15	222
Hertzian dipole, or short vertical monopole, on a perfectly conducting ground**	3	4.8	300
quarter-wave monopole on a perfectly conducting ground	3.3	5.2	314

* $G_i = 10 \log g_i$

** for the Hertzian dipole, it is assumed that the antenna is just above a perfectly conducting ground supplied to the antenna and its gain in a given direction relative to a short vertical antenna.

Note that ERP, which is often used as a general term for radiated power, strictly only applies when the reference antenna is a halfwave dipole.

An alternative way of indicating the intensity of radiation, which is sometimes used at the lower frequencies, is in terms of the cymomotive force, expressed in volts. The cymomotive force is given by the product of the field strength and the distance, assuming loss-free radiation. Values of cymomotive force when 1 kW is radiated from the reference antennas are also given in Table 2.1.

2.5 The effect of the ground

The proximity of the imperfectly conducting ground will affect the performance of an antenna. In some cases, where the antenna is located several wavelengths above the ground, it may be convenient to consider signals directly from (or to) the antenna and those which are reflected from the ground or another nearby surface as separate signal ray paths. When the antenna is close to, or on, the ground it is no longer appropriate to consider separate rays and then the effect may be taken into account by assuming a modified directivity pattern for the antenna, including the ground reflection; by modifying the effective aperture of the antenna; or by taking account of the change in radiation resistance etc. A discussion of this for the ground-wave case, where the problem is most difficult, is contained in Annex II of Recommendation ITU-R P.341.

Information concerning the electrical characteristics of the surface of the earth is contained in further ITU-R texts: Recommendations ITU-R P.527 and P.832.

2.6 Transmission loss

The power available, p_r, in a load which is conjugately matched to the impedance of a receiving antenna is:

$$p_r = sa_e \tag{2.5}$$

where a_e is the effective aperture of the antenna, given by $\lambda^2/4\pi$ for an ideal loss-free isotropic antenna.

Thus from Eqn. 2.1 the power received by an ideal isotropic antenna at distance r due to power p_t radiated isotropically is given by:

$$p_r = \frac{p_t}{4\pi r^2} = p_t\left(\frac{\lambda}{4\pi r}\right)^2 \tag{2.6}$$

and the free-space basic transmission loss is the ratio p_t/p_r.

Transmission losses are almost always expressed in logarithmic terms, in decibels, and as a positive value of attenuation, i.e.

$$L_{bf} = 10 \log \left(\frac{p_t}{p_r}\right) = P_t - P_r = 20 \log \left(\frac{4\pi r}{\lambda}\right) \tag{2.7}$$

or

$$L_{bf} = 32.44 + 20 \log f + 20 \log d \tag{2.8}$$

where f is in MHz and d is in km.

The concept of transmission loss may be extended to include the effects of the following mechanisms in the propagation medium, and of the antennas and the radio system actually in use:

Free-space basic transmission loss L_{bf} relates to isotropic antennas and loss-free propagation;

Basic transmission loss L_b includes the effect of the propagation medium, e.g.

(a) absorption loss due to atmospheric gases or in the ionosphere
(b) diffraction loss due to obstructions such as hills or buildings
(c) reflection losses, including focusing or defocusing due to curvature of reflecting layers
(d) scattering due to irregularities in the atmospheric refractive index or in the ionosphere or by hydrometeors
(e) aperture-to-medium coupling loss or antenna gain degradation, which may be due to the presence of substantial scatter phenomena on the path
(f) polarisation coupling loss; this can arise from any polarisation mismatch between the antennas for the particular ray path considered
(g) effect of wave interference between the direct ray and rays reflected from the ground, other obstacles or atmospheric layers

Transmission loss L includes the directivity of the actual transmitting antennas, disregarding antenna circuit losses;

System loss L_s is obtained from the powers at the antenna terminals;

Total loss L_t is the ratio determined at convenient, specified, points within the transmitter and receiver systems.

The relationships between these loss ratios are illustrated in Figure 2.1. It is important to be precise when using the terms, and the full definitions are given in Recommendation ITU-R P.341.

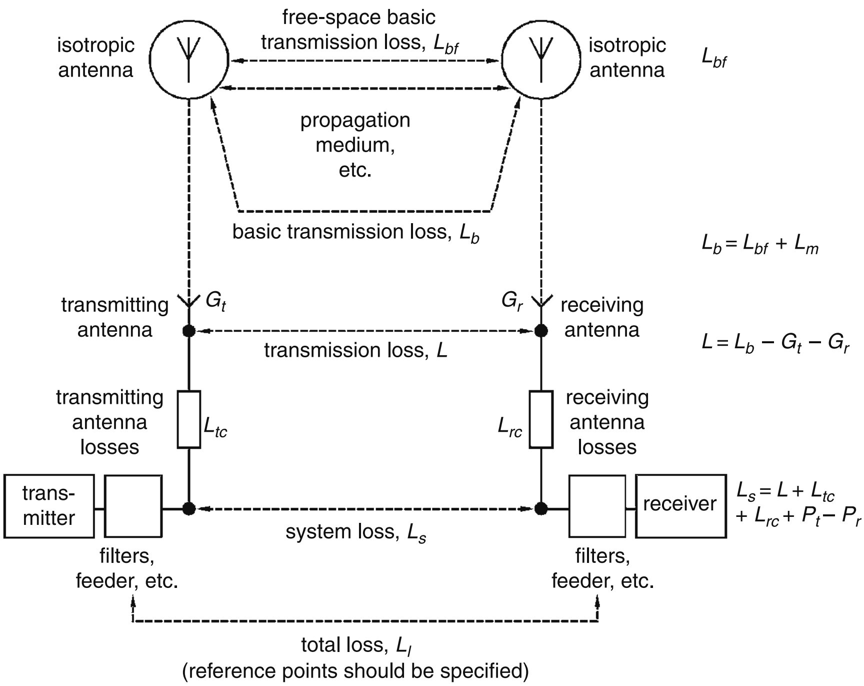


Figure 2.1 Graphical description of terms for transmission loss

2.7 Fading and variability

Both signals and noise are subject to variations in time and with location. These changes in intensity arise from the nature of a random process, from multipath propagation, from changes in refractivity along the path, from movements of the system terminals or the reflecting medium, from changes in transmission loss etc. A knowledge of the statistical characteristics of a received signal may be required in assessing the performance of modulation systems etc.

Statistics of the signal variability are also required for spectrum planning and for predicting the performance of systems. For these purposes it is important to know, for example:

(a) the signal level exceeded for large percentages of time or location (e.g. for the determination of quality of the wanted service or of the service area);
(b) the signal level which occurs for small percentages of time (e.g. to determine the significance of potential interference or the feasibility of frequency reuse).

In some cases signals are subject to rapid or closely spaced variations, superimposed on a slower variability. It may be possible to treat the phenomena separately, say by using a long receiver integration time or by averaging the level of the signal (e.g. with AGC) so that the time interval adopted encompasses many individual short-term or closely spaced fluctuations. In other cases an understanding of the overall variability of the signal may require a consideration of the combined effects of two types of variability.

2.7.1 Occurrence distributions

When the value of a parameter results from the cumulative effect of many processes, each of which has the same central tendency, the probability density $p(x)$ has a bell-shaped distribution. The example in Figure 2.2 is a histogram where individual results are collected into a number of bins or intervals. Representative, typical, values for this distribution are the arithmetic mean, or average, the mode or the median.

For n discrete values of a variable x, measured at regular intervals of time or location etc., the mean value $\overline{x}$ is given by:

$$\overline{x} = \frac{\sum x_n}{n} \tag{2.9}$$

The modal value is that which occurs most frequently and thus is at the peak value of the histogram.

The median is the value which is exceeded by 50 per cent of the values. Other percentiles may also be determined from the distribution: the quartiles, each comprising a quarter of the values, and the deciles, each comprising ten per cent of the values. For an asymmetrical distribution the mean, median and mode will have different values, but they coincide when the distribution is symmetrical.

In radiowave propagation, where many of the parameters are expressed in

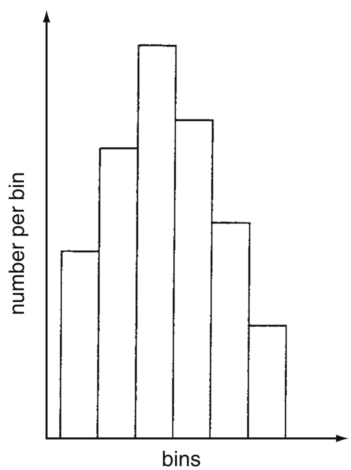


Figure 2.2 Example of a histogram

decibels, an arithmetic mean of a set of logarithmic values in decibels makes little sense, and the median and other percentiles are much more useful.

2.7.2 Normal (Gaussian) distribution

It is often assumed that a symmetrical, centrally peaking, distribution is that of the normal, Gaussian, distribution as shown in Figure 2.3.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right\} \tag{2.10}$$

where σ is a normalising parameter, the standard deviation; σ^2 is also called the variance:

$$\sigma = \sqrt{\frac{\sum(x_n - \overline{x})^2}{n}} \tag{2.11}$$

It may be noted that the probability density given by this relationship is that contained in a bin with a width of one standard deviation.

The cumulative probability density function (CDF), *F(x)*, for this distribution is shown in Figure 2.4 and is given by:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{t-\overline{x}}{\sigma}\right)\right\} dt \tag{2.12}$$

It may be noted that the CDF may be reversed to give the probability that the function is exceeded.

Statistical tables giving values for both *p(x)* and *F(x)* are readily available.

Special graph paper is also available where the abscissa scale is arranged so that a variable which is normally distributed appears as a straight line with a slope proportional to the standard deviation. This may be useful since it is easy to see how far a set of results departs from this distribution.

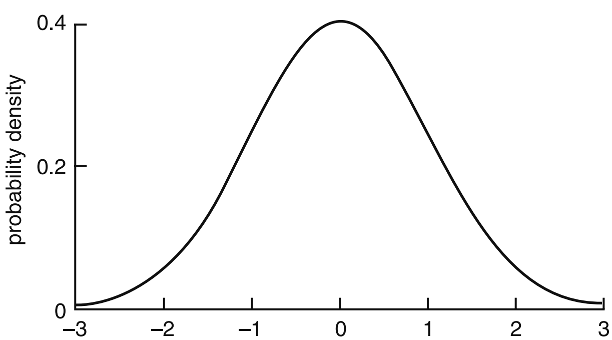


Figure 2.3 The normal probability density function

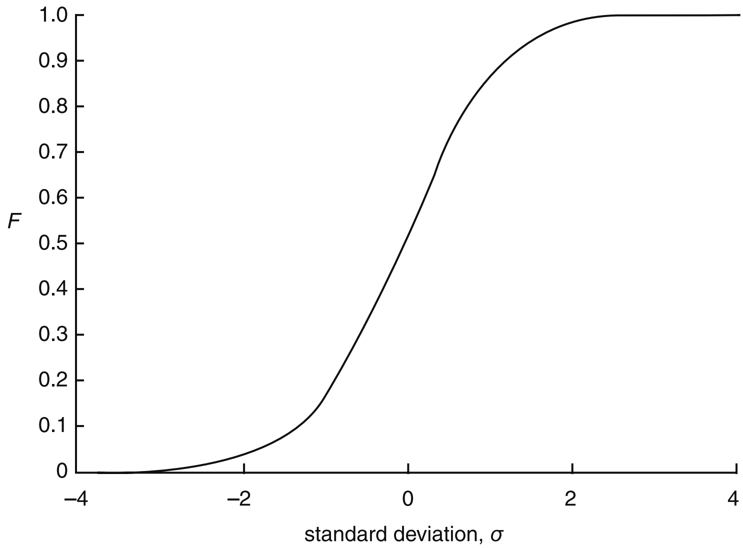


Figure 2.4 The normal cumulative probability function, showing the probability, F, that the function is not exceeded

An approximation for the half of the distribution where $x < \overline{x}$ is given by:

$$F(x) = \frac{\exp(-y^2/2)}{\sqrt{2\pi}\{0.661y + 0.339\sqrt{y^2 + 5.51}\}} \tag{2.13}$$

where

$$y = \frac{\overline{x} - x}{\sigma}$$

The upper half of the distribution may be obtained by using Eqn. 2.13 with $y = (x - \overline{x})/\sigma$, in this case giving $1 - F(x)$.

Table 2.2 gives some examples taken from the distribution.

In fact, in radiowave propagation, a normal distribution of signal power etc. only occurs when there are small fluctuations about a mean level, as might be the case when studying scintillation. Predominantly, it is the normal distribution of

Table 2.2 The normal distribution

Occurrence	
68%	within ± 1σ
95.5%	within ± 2σ
90%	less than +1.28σ
99%	less than +2.33σ
99.9%	less than +3.09σ
99.99%	less than +3.72σ

the logarithms of the variable which gives useful information: the log-normal distribution.

2.7.3 Log-normal distribution

In the case of a log-normal distribution, each parameter (the values of the variable itself, the mean, the standard deviation etc.) is expressed in decibels and the equations in Section 2.7.2 then apply. The log-normal distribution is appropriate for very many of the time series encountered in propagation studies, and in some cases also for the variations with location, for example within a small area of the coverage of a mobile system. Note that, when a function is log-normally distributed, the mean and median of the function itself (for example expressed in watts or volts) are not the same: the median is still defined as the central value of the distribution, whereas the mean of the numerical values upon which the log-normal distribution is based is given by $\overline{x} + \sigma^2/2$.

2.7.4 Rayleigh distribution

The combination of a large number (at least more than three) of component signal vectors with arbitrary phase and similar amplitude leads to the Rayleigh distribution. Thus, this is appropriate for situations where the signal results from the combination of multipath or scatter components. In this case

$$p(x) = \frac{2x}{b}\exp\left(-\frac{x^2}{b^2}\right) \tag{2.14}$$

and

$$F(x) = 1 - \exp\left(-\frac{x^2}{b^2}\right) \tag{2.15}$$

where b is the root mean square value (note that x and b are numerical amplitude values, not decibels).

For this distribution the mean is $0.886b$, the median is $0.833b$, the mode is $0.707b$ and the standard deviation is $0.463b$.

It is useful to note that, for small values of $F(x)$, $F(x) \approx x^2/b^2$, so that when x is a voltage amplitude its power decreases by 10 dB for each decade of probability. However, this is not a sufficient test to determine whether a variable is Rayleigh distributed, since some other distributions have the same property. This property is shown in Table 2.3, which gives some examples from the Rayleigh distribution.

Table 2.3 Rayleigh distribution

$F(x)$	0.999	0.99	0.9	0.5	0.1	0.01	0.001	0.0001
$20\log(x)$	+10 dB	+8.2	+5.2	0	−8.2	−18.4	−28.4	−38.4

Special graph paper is also available on which a Rayleigh distribution is plotted as a straight line: this is the presentation used in Figure 2.6 where the line marked $-\infty$ dB is a Rayleigh distribution. Note, however, that such a presentation greatly overemphasises the appearance of small time percentages, and care should be taken that this does not mislead in the interpretation of plotted results.

2.7.5 Combined log-normal and Rayleigh distribution

In a number of cases the variation of the signal may be represented as having two components: rapid or closely spaced fluctuations, which may be due to multipath or scatter, with a Rayleigh distribution; and the mean of these rapid variations, measured over a longer period of time or a longer distance, with a log-normal distribution.

This distribution is given by Boithias [1]:

$$1 - F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{- x^2 \exp\{-0.23\sigma u\} \frac{u^2}{2}\right\} du \tag{2.16}$$

where σ is the standard deviation (in decibels) of the log-normal distribution. This combined distribution is shown in Figure 2.5. This combination of distributions has also been studied by Suzuki, and his proposed formulation has been evaluated by Lorenz [2].

2.7.6 Rice distribution

The Rice distribution (also called the Nakagami-*n* distribution) applies to the case where there is a steady, nonfading, component together with a random variable component with a Rayleigh distribution. This may occur where there is a direct signal together with a signal reflected from a rough surface, where there is a steady signal together with multipath signals, or at LF and MF where there is a steady ground-wave signal and a signal reflected from the ionosphere.

The probability density for the Rice distribution is given by

$$p(r) = \frac{2r}{b^2} \exp\left(-\frac{r^2 + a^2}{b^2}\right) \cdot I_0\left(\frac{ax}{\sigma^2}\right) \tag{2.17}$$

and

$$1 - F(x) = 2\exp(-a^2/b^2) \int_{x/b}^{\infty} v \exp(-v^2) \cdot I_0\left(\frac{2av}{b}\right) dv \tag{2.18}$$

where

$$I_0(z) = \frac{1}{\pi} \int_0^{\pi} e^{-z \cos\theta} d\theta \tag{2.19}$$

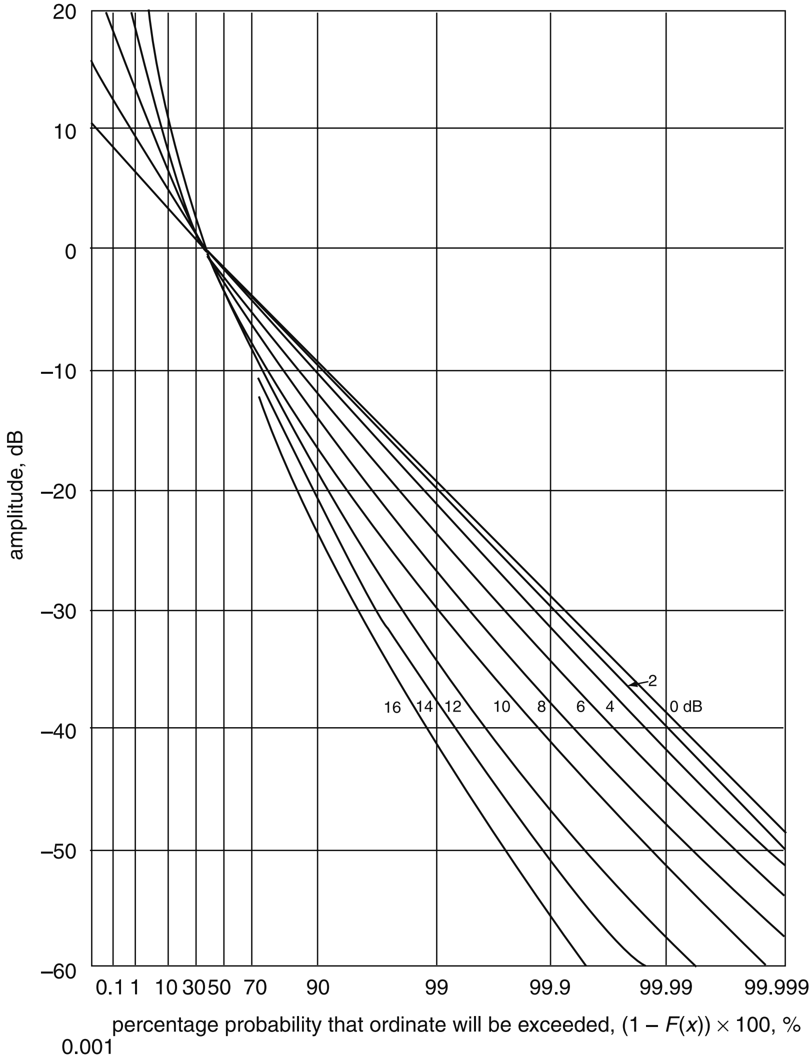


Figure 2.5 Combined log-normal and Rayleigh distribution (with σ of log-normal distribution as parameter)

where the RMS values of the steady and the Rayleigh components are a and b, respectively. A parameter $k = a^2/b^2$, the ratio of the powers of the steady and Rayleigh components, is often used to describe the specific distribution and is often expressed in decibels, K. (Boithias (and the figure in the ITU-R Recommendation) uses the parameter $b^2/(a^2 + b^2)$.) The distribution is shown in Figure 2.6.

In most cases, the power in the fading component will add to the power of the steady signal where, for example, the multipath brings additional signal modes to the receiver. In some other cases, the total power will be constant where the random component originates from the steady signal.

2.7.7 The gamma distribution

For phenomena which mainly occur for small time percentages, for example for rainfall rates, the gamma distribution may be useful. The distribution is given by:

$$p(x) = \frac{a^{v}}{\Gamma(v)} \cdot x^{v-1} \cdot e^{-ax} \tag{2.20}$$

where Γ is the Euler function of the second order. The distribution is shown in Figure 2.7. The useful values of v for propagation purposes are small, of the order of 10^{-2} to 10^{-4}, and in such cases, and where ax is not too small:

$$1 - F(x) \cong v \int_{ax}^{\infty} \frac{e^{-t}}{t} dt \tag{2.21}$$

and this may be approximated as:

$$1 - F(x) \cong v \cdot \frac{e^{-ax}}{0.68 + ax + 0.28 \log ax} \tag{2.22}$$

valid for $v < 0.1$ and $ax > 0.03$.

2.7.8 Other distributions

Many further asymmetrical distributions have been studied and utilised in propagation studies. Griffiths and McGeehan [2] have compared some of these such as the exponential, Wiebull, χ-squared, Stacy and Nakagami-*m* distributions.

It may be appropriate to include such distributions in models of particular propagation behaviour. For example, Lorenz [3] has suggested that the Suzuki distribution is appropriate for VHF and UHF mobile communication in built-up areas and forests, and the Wiebull distribution is appropriate for area-coverage statistics where line-of-sight paths occur frequently.

However, before embarking on the use of an unfamiliar and complicated distribution, the user should be sure that the uncertainty and spread in the observations is sufficiently small so that the use of the distribution will result in a significant improvement in the accuracy of the model. The difference between various distributions for values between, say, 10 and 90 per cent occurrence will often be small, and it is only in the tails of the distribution, where observations may be sparse, that a distinction could be made.

For applications concerned with the quality and performance of a wanted

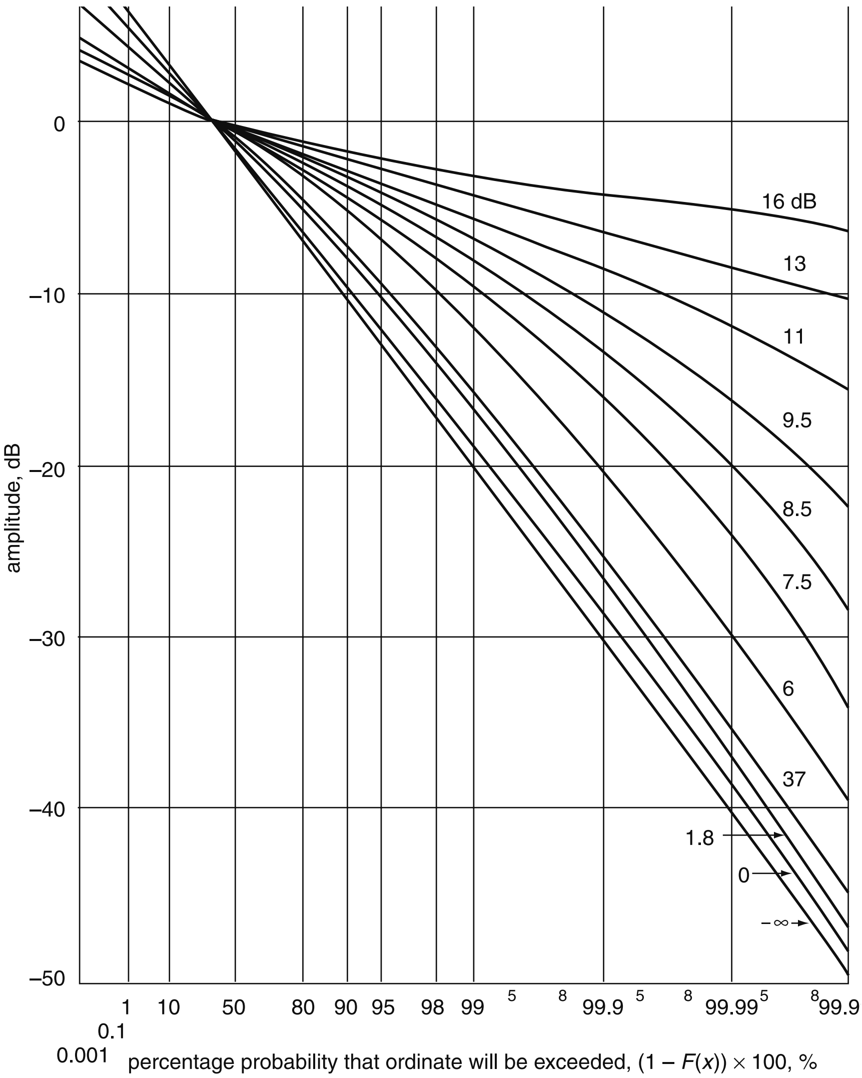


Figure 2.6 Nakagami–Rice distribution function F(r) (values of K (in dB) are shown on the curves)

signal, or with the interference effects of an unwanted signal, it is seldom necessary to consider both tails of a distribution at the same time. In some cases half-log-normal distributions, applying a different value of the standard deviation on each side of the median, will be adequate. The mathematical elegance of a complete distribution should be weighed against the practical convenience of using the appropriate half of a more common distribution for the occurrence percentages of interest.

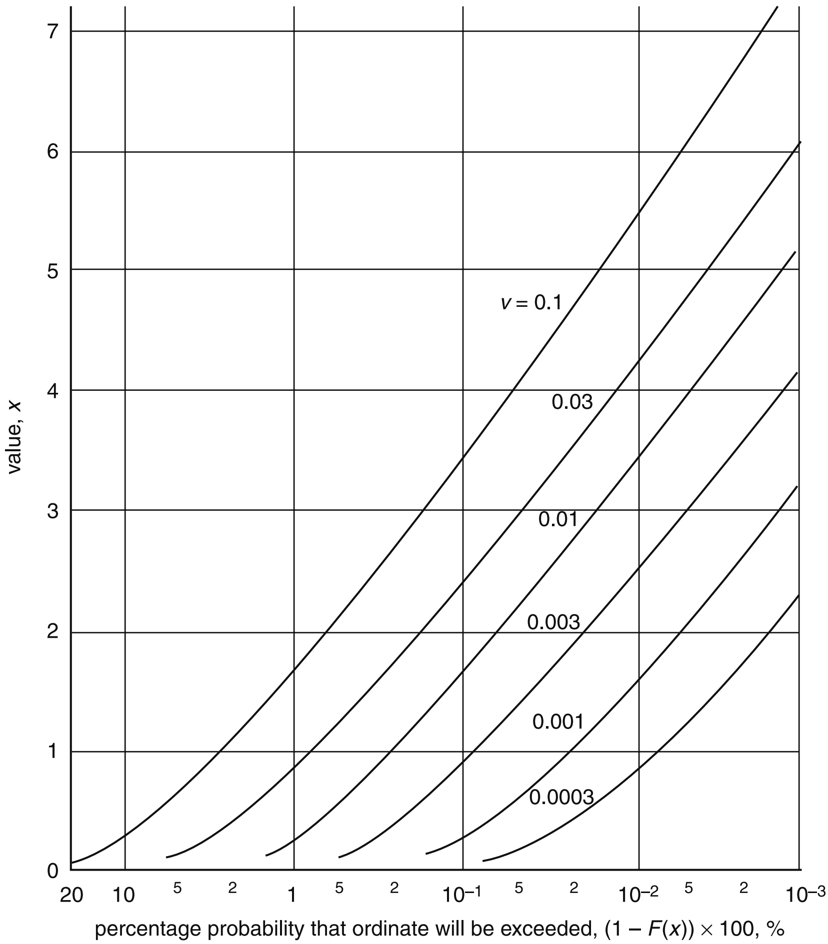


Figure 2.7 The gamma distribution ($\alpha = 1$, $v \leq 0.1$)

2.8 Radio noise

There are a number of types of radio noise that must be considered in any design although, in general, one type will predominate in particular circumstances and that should be taken into account in the system design. Broadly, the noise can be divided into two types: noise internal to the receiving system and noise external to the receiving antenna.

The internal noise is due to losses in the antenna circuits or in the transmission line, or is generated in the receiver itself. It has the characteristics of thermal noise (i.e. white Gaussian noise). External noise arriving at the receiving antenna may be due to:

(i) atmospheric noise generated by lightning discharges, or resulting from absorption by atmospheric gases (sky noise);
(ii) the cosmic background, primarily from the Galaxy, or from the sun;
(iii) broadband man-made noise generated by machinery, power systems etc.

In general, noise is broadband, with an intensity varying only slowly with frequency, but in some cases, e.g. noise emanating from computer systems, the intensity may have considerable frequency variability.

The noise power due to external sources, p_n, can conveniently be expressed as a noise factor, f_a, which is the ratio of the noise power to the corresponding thermal noise, or as a noise temperature, t_a, thus:

$$f_a = \frac{P_n}{kt_0 b} = \frac{t_a}{t_0} \tag{2.23}$$

where k is Boltzmann's constant = 1.38×10^{-23} J K^{-1}, t_0 is the reference temperature, taken as 288 K and b is the noise bandwidth of the receiving system in Hz.

Note there is some confusion in the currently used terminology but here f_a is the numerical noise factor, and the term noise figure, F_a, is used for the logarithmic ratio, so that:

$$F_a = 10 \log f_a \tag{2.24}$$

The available noise power in decibels above 1 W is given by:

$$P_n = F_a + B - 204 \quad \text{dBW} \tag{2.25}$$

where $B = 10 \log b$.

When measured with a halfwave dipole in free space the corresponding value of the RMS field strength is given by:

$$E_n = F_a + 20 \log f_{MHz} + B - 99 \quad \text{dB}(\mu\text{V m}^{-1}) \tag{2.26}$$

and for a short vertical grounded monopole by (taking account of the effect of the ground as mentioned in Section 2.5):

$$E_n = F_a + 20 \log f_{MHz} + B - 95.5 \quad \text{dB}(\mu\text{V m}^{-1}) \tag{2.27}$$

Minimum, and some maximum, values for the external noise figures are shown in Figures 2.8 and 2.9. Generally, one type of noise will predominate, but where the contributions of more than one type of noise are comparable, the noise factors (not the figures in decibels) should be added. Atmospheric noise due to lightning varies with location on the earth, season, time of day and frequency. Maps and frequency correction charts are given in ITU-R Recommendation P.372. Man-made noise varies with the extent of man-made activity and the use of machinery, electrical equipment etc. The relationship for a range of environments is also given in the Recommendation. Note that some care may be needed to ensure that an appropriate curve is selected since the noise generated, for example in a business area, may differ from country to country.

Both atmospheric and man-made noise are impulsive in character, and an

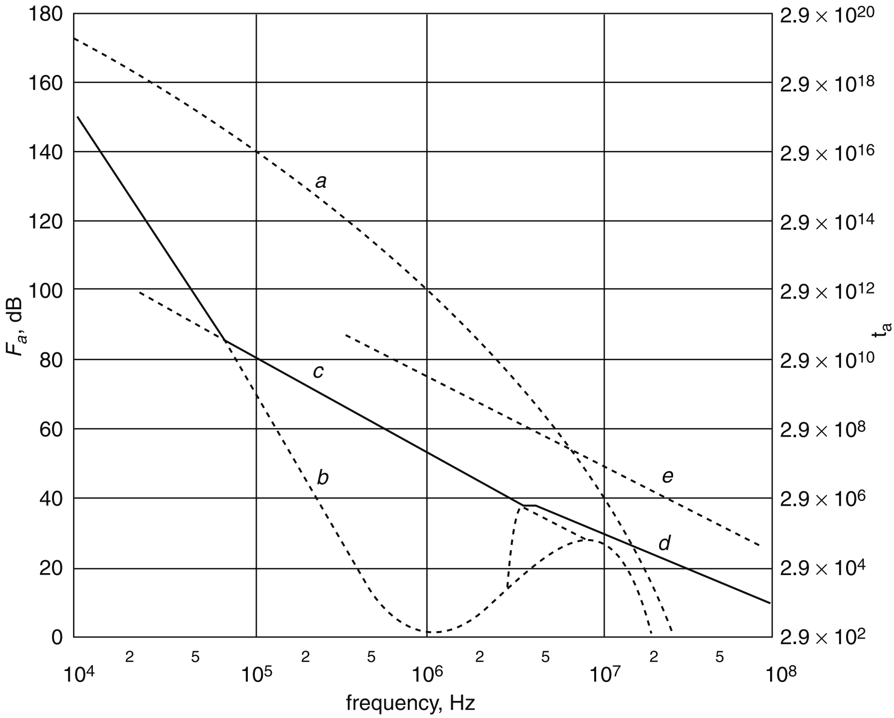


Figure 2.8 Noise figure F_a against frequency; 10 kHz to 100 MHz
(a) atmospheric noise, value exceeded 0.5% of time
(b) atmospheric noise, value exceeded 99.5% of time
(c) man-made noise, quiet receiving site
(d) galactic noise
(e) median business area man-made noise
—— minimum noise level expected

assessment based wholly on the noise power is likely to be inadequate. In some cases the dominant feature which determines system performance will be a parameter derived from the amplitude probability distribution of the noise. The Recommendation includes this information, together with examples of the prediction of system performance. However, in other cases, such as for digital systems, the characteristic duration and repetition rate of noise impulses may be important.

When noise power is the appropriate parameter, the limit to system performance depends on the overall operating noise factor f from the combination of external and internal (receiver and antenna) sources. When the temperature of the antenna circuits etc. is t_0:

$$f = f_a + f_c f_t f_r - 1 \tag{2.27}$$

where f_c and f_t are the losses (available input power/available output power) for

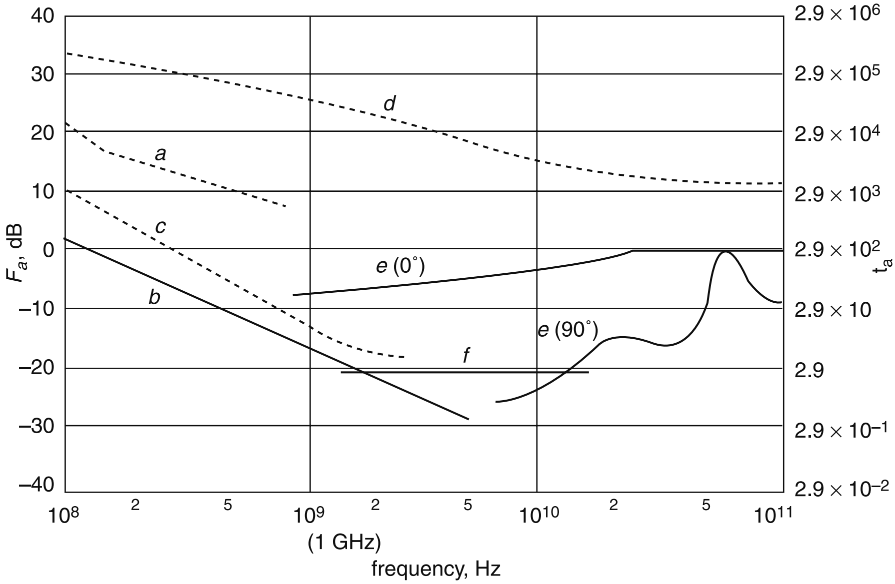


Figure 2.9 *Noise figure F_a against frequency; 100 MHz to 100 GHz*
(a) estimated median business are man-made noise
(b) galactic noise
(c) galactic noise (toward galactic centre with infinitely narrow beamwidth)
(d) quiet sun (½° beamwidth directed toward sun)
(e) sky noise due to oxygen and water vapour (very narrow beam antenna); upper curve, 0° elevation angle; lower curve, 90° elevation angle
(f) black body (cosmic background), 2.7K minimum nose level expected

the antenna circuit and transmission line, respectively; f_r is the noise factor of the receiver and f_a is the noise factor due to the external noise sources.

2.9 Link power budgets

In the design of a communications system it will be necessary to ensure that the level of the received signal is adequate to provide the required quality of service.

The way in which this is done will depend on the details of the system and the requirements for the application.

In some cases, where the performance of a typical, or minimum specification, receiver is assumed and where there may be no control over some of the antenna characteristics, a minimum usable field strength or reference minimum field strength may be specified. This has been done for some broadcast and mobile services. Assuming that the reception approximately corresponds to free-space conditions, e.g. not too close to the ground, which is implicit in this kind of characterisation, Eqns 2.3 and 2.5 may be used to relate the field strength to a minimum available power, $p_{r,min}$, in the typical receiver. Then a simplified link power budget, with the terms expressed in decibels, may be determined:

$$P_t + G_t = L_b + P_{r,\ min} \tag{2.28}$$

In other cases it will be appropriate to consider the performance of individual receiving installations and to specify the required signal/noise ratio, $R_{S/N}$. Where the performance is limited by the effective receiver noise factor, the link power budget becomes:

$$P_t + G_t = L_b + R_{S/N} + F_a + B - 204 \tag{2.29}$$

where P_t is in dBW.

In practical circumstances these skeleton link budgets would have to be refined by adding antenna feeder losses etc. For digital systems other aspects, such as the impulse response of the channel, will also have to be considered, in addition to the transmission loss.

In all cases, however, the basic transmission loss, L_b, is a determining term which will be obtained from propagation prediction or modelling. As described in Section 2.6, the basic transmission loss includes losses due to propagation effects in addition to the free-space spreading loss. These additional losses should be determined for the time percentage used in defining the quality of service. Typically, a service quality will be specified for 95, 99 or 99.99 per cent of the time or location etc., and the channel fading or the incidence of the effect at this percentage should be determined. This will be the subject of later chapters.

The procedure may, if necessary, be extended still further to include the probable error of the prediction of transmission loss which may be due, for example, to the sampling involved in establishing the method. Where no allowance for this is included, the prediction has a confidence level of 50 per cent, since one half of the specific cases encountered are likely in practice to be below the predicted level. An assessment may be made of the probable error and, by applying a normal distribution, an allowance may be made for any other desired confidence level. This has been described by Barclay [4].

2.9.1 Fading allowances

In some cases, for service planning, the specified signal-to-noise ratio for the required grade of service will probably include an allowance for the rapid fading which will affect the intelligibility or the bit error ratio of the system. It may still be necessary to allow for other variations (hour-to-hour, day-to-day, location-to-location) of both signal and noise, which are likely to be log-normally distributed, but uncorrelated. This may be done by determining the basic transmission loss for median conditions and then adding a fading allowance, assuming a log-normal distribution with a variance, σ^2, obtained by adding the variances of each contributing distribution, i.e.

$$\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2 + \ldots \tag{2.30}$$

2.10 Diversity

The allowance for propagation effects necessary to achieve a good grade of service may demand economically prohibitive transmitter powers and antenna gains. In any case, the use of excessive radiated power conflicts with the need for good spectrum utilisation. Techniques for overcoming this problem include coding and diversity. Particularly for circuits with rapid fading, such as those where Rayleigh or Rician fading dominates, copies of the signal with the same characteristics are available displaced in time, position, frequency or, for some types of propagation, with angle of arrival or orthogonal polarisation. For example, a signal message may be repeated later in time if, when first transmitted, a fade had reduced the signal-to-noise ratio. For digital systems this process may be automated by the use of an error-detecting code using a method of automatic repeat requests (ARQ) if errors are detected in the received signal. More modern techniques use sophisticated error-correction and error-detection codes to combat the effects of fading, and these may take account of the expected patterns of error occurrence (e.g. where errors occur in bursts). Spread-spectrum signals, both direct-sequence and frequency-hopping, employ techniques to take advantage of the frequency-selective nature of fading, and this is discussed in later chapters.

Diversity techniques utilise two or more samples of the signal obtained from separated antennas, or sometimes from duplicated transmissions on several frequencies. These signals are then combined in the receiver to produce an output with a smaller fading variability. Signals may be combined by techniques such as:

(i) selection of the stronger or strongest
(ii) combining the output of channels with equal gain
(iii) weighting the combination according to the signal-to-noise ratio of the channel (maximal ratio combining).

Figure 2.10 shows the distributions for two-element diversity [5], where the signals are uncorrelated and each has a Rayleigh distribution, for various methods of combination. The corresponding distribution for four-channel diversity (as might be employed for tropospheric scatter systems) is shown in Figure 2.11. In fact, a substantial advantage is still obtained if the signals are partially correlated. Figure 2.12 shows, for two-element-selection diversity, the effect of varying the correlation coefficient.

2.10.1 Correlation coefficient

The correlation coefficient is obtained as:

$$r = \frac{\sum\{(x-\overline{x})(y-\overline{y})\}}{\left\{\sum(x-\overline{x})^2\sum(y-\overline{y})^2\right\}^{0.5}} = \frac{\sum xy - n\overline{xy}}{n\sigma_x\sigma_y} \tag{2.31}$$

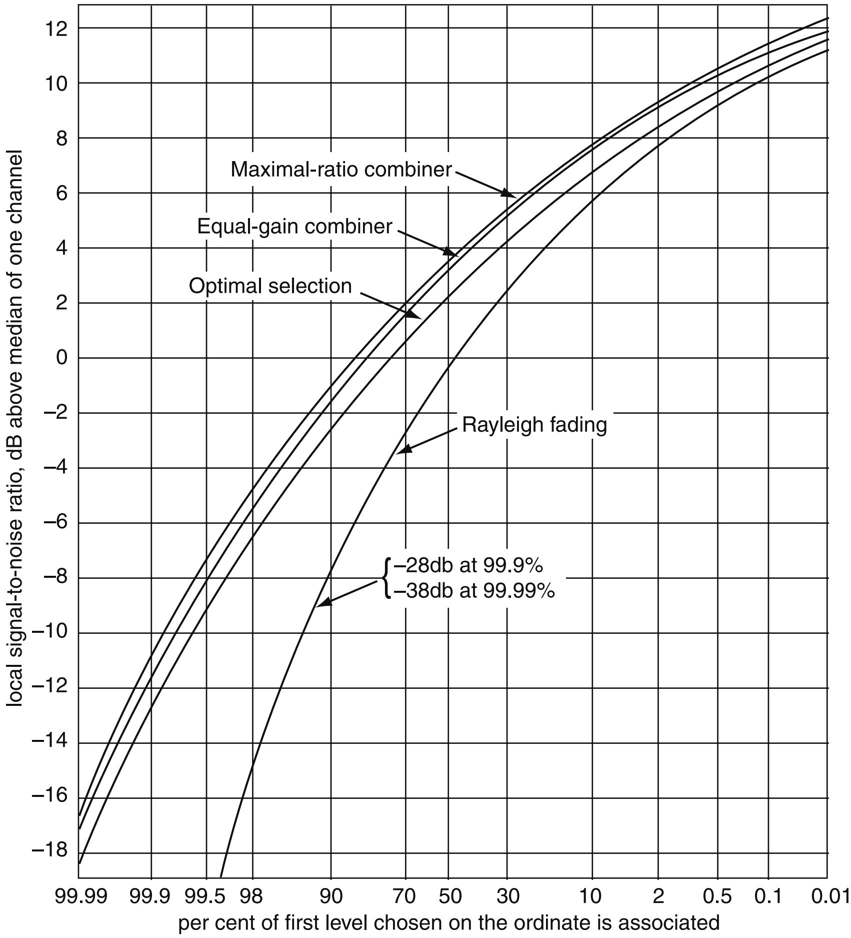


Figure 2.10 Two-element diversity

where x and y refer to simultaneous, or appropriately time-shifted, pairs of values from the two distributions and n is the number of pairs.

2.10.2 Diversity gain and diversity improvement

The advantage given to system performance by the application of diversity may be expressed in one of two ways:

(a) as diversity gain: the increase in the signal level exceeded when using diversity for a specified time percentage;
(b) as a diversity improvement factor: the ratio of the time percentages, without and with diversity, for which the signal fade depth exceeds a specified

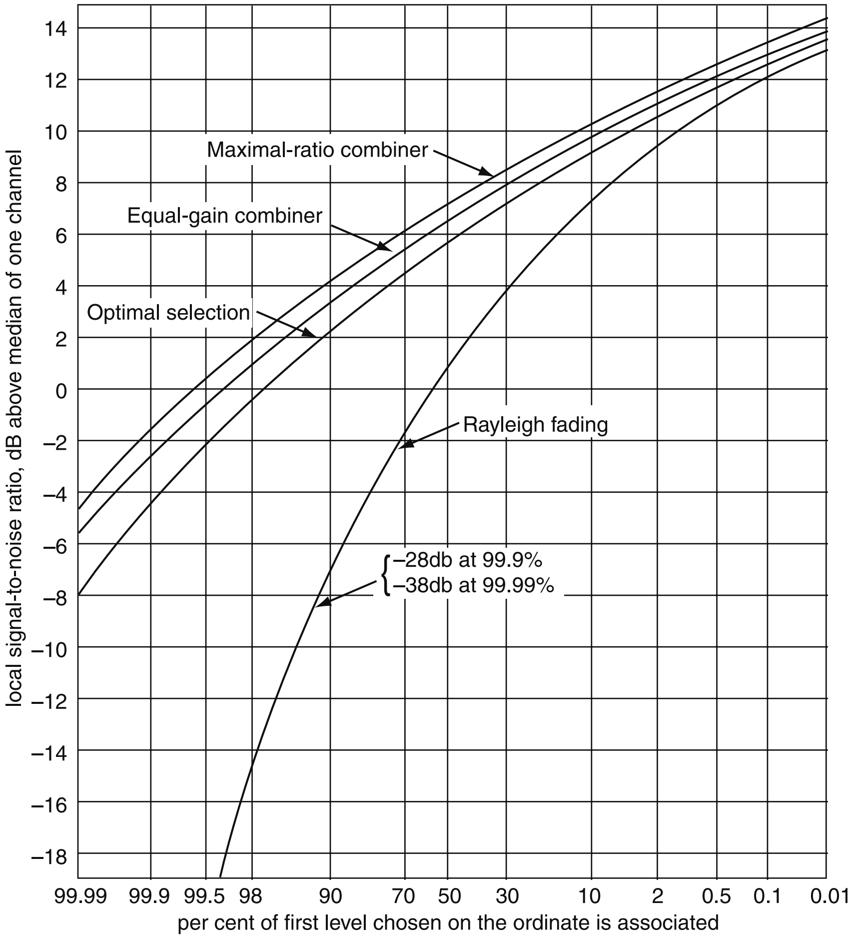


Figure 2.11 Four-element diversity

level (note that Figures 2.10 and 2.11 give the time percentage at which a fade depth is *not* exceeded).

2.11 References

1 BOITHIAS, L.: 'Radio wave propagation' (North Oxford Academic Publishers, 1987)
2 GRIFFITHS, J., and McGEEHAN, J. P.: 'Inter-relationships between some statistical distributions used in radiowave propagation', *IEE Proc F*, 1982, **129**, (6), pp. 411–417
3 LORENZ, R.W.: 'Field strength prediction methods for mobile telephone system using a topographical data bank'. *IEE Conf Pub 188*, 1980, pp. 6–11

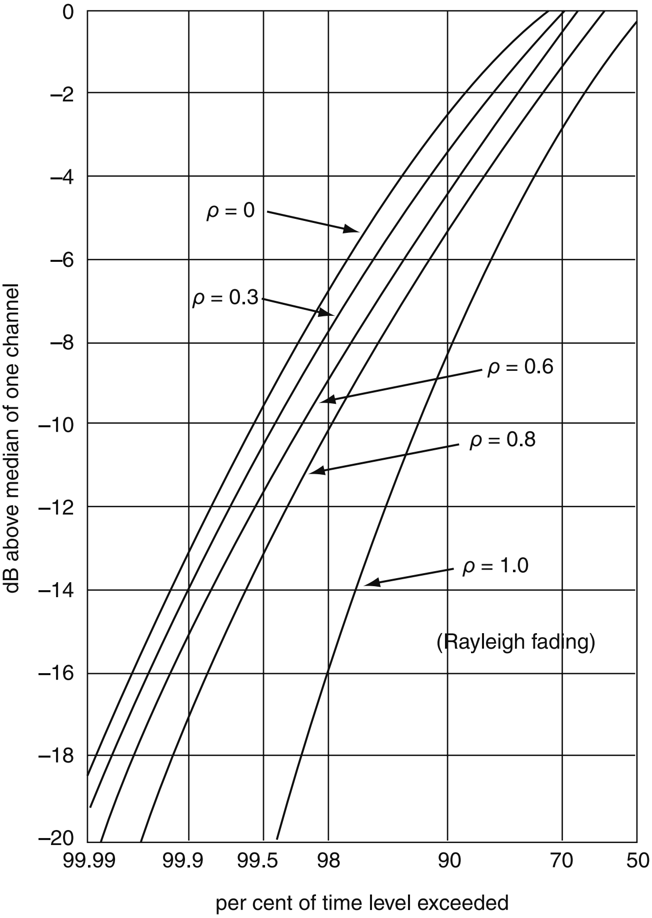


Figure 2.12 Two-element diversity, Rayleigh fading for various degrees of correlation

4 BARCLAY, L. W.: 'Statistical modelling of HF links'. AGARD conference proceedings 238, Ottawa, Canada, 1978
5 BRENNAN, D. G.: 'Linear diversity combining techniques', *Proc. Inst. Radio Eng.*, 1959, **47**, pp. 1075–1102

Chapter 3

Basic principles 2

David Bacon

3.1 Vector nature of radiowaves

Although for many purposes a radiowave can be treated simply as a power flow, there are other situations where it is necessary to take into account the fact that it consists of vector fields.

3.1.1 Qualitative description of the plane wave

Although radiowaves radiate spherically, in a small volume of space far from the transmitting antenna the associated vectors can be treated as straight lines. A visualisation of the resulting plane wave is given in Figure 3.1, with the various directions defined by Cartesian axes. Electric and magnetic fields E and H are drawn parallel to the x and y axes, respectively. The direction of power flow is in the z direction.

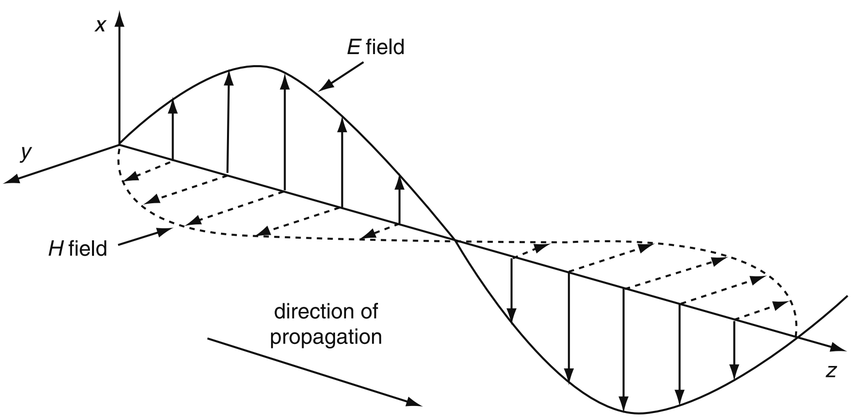


Figure 3.1 Vector fields in an advancing plane wave

The E and H field sinusoids are in phase with each other, and the complete pattern moves in the z direction. Thus the E and H fields vary in both space and time. Although in Figure 3.1 the field vectors are drawn from a single z axis, in fact they fill and are equal over any xy plane.

Two useful concepts in propagation studies are:

Ray: a ray is a mathematically thin line indicating the direction of propagation. In Figure 3.1, any line for which x and y are both constant can be viewed as a ray.

Wavefront: a wavefront is any surface which is everywhere normal to the direction of propagation. In Figure 3.1, any plane for which z is constant can be viewed as a wavefront.

3.1.2 Complex phasor notation

Complex phasor notation is a convenient and efficient method to manipulate radio signals where it is necessary to take account of both amplitude and phase. Figure 3.2 shows an imaginary phasor OE of length A rotating anticlockwise in the complex plane defined by a real axis R, and an imaginary axis I on which quantities are multiplied by $j = \sqrt{-1}$. The phasor OE can represent the amplitude and phase information of a field E.
In general:

$$\theta = \omega t + \phi \quad \text{radians} \tag{3.1}$$

where pulsatance, $\omega = 2\pi f$, is the angular frequency in radians s^{-1}, and ϕ is the offset from an arbitrary phase origin.

The phasor OE contains the amplitude and phase information of a field E. The real value of E at any time is given by:

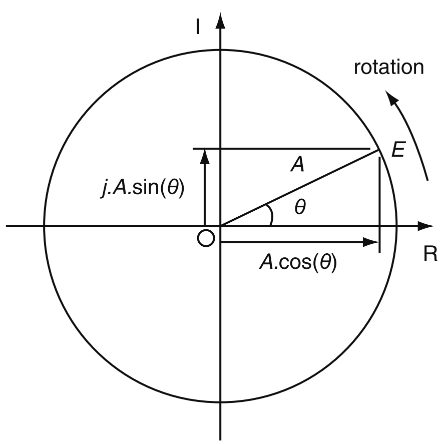


Figure 3.2 Phasor rotation in complex plane

$$\mathrm{Re}(E) = A\cos(\theta) \quad \mathrm{V\,m^{-1}} \tag{3.2}$$

and the so-called imaginary component of E is given by:

$$\mathrm{Im}(E) = j.A\sin(\theta) \quad \mathrm{V\,m^{-1}} \tag{3.3}$$

Complete information on the amplitude and phase of E is contained in the real and imaginary components, and hence E is fully defined by:

$$E = A\,\{\cos(\theta) + j.\sin(\theta)\} \quad \mathrm{V\,m^{-1}} \tag{3.4}$$

Following the identity $\cos(x) + j.\sin(x) = \exp(jx)$, E can alternatively be expressed in the complex exponential form:

$$E = A\exp(j.\theta) \quad \mathrm{V\,m^{-1}} \tag{3.5}$$

3.1.3 The sense of time and space

In Eqn. 3.1 the phase offset angle ϕ will in many cases represent relative phase at different spatial locations. A minor point concerning sign needs to be noted here. Figure 3.3 shows how a selected point on the E sinusoid of Figure 3.1 will change in two different situations.

In Figure 3.3*a* the observer selects point P on the sinusoidal variation of the electric fields strength E, and for a constant value of z notes how E changes as time evolves. For the point P as drawn, E will initially increase as the complete pattern moves in the z direction.

In Figure 3.3*b* the observer selects the same point P but in this case examines how E varies with increasing z, with time frozen. In this case E decreases.

Thus E and H fields vary with the opposite sense in time and space. As a result the phasor angle can be written:

$$\theta = \omega t - kz \quad \text{radians} \tag{3.6}$$

where pulsatance ω is given by:

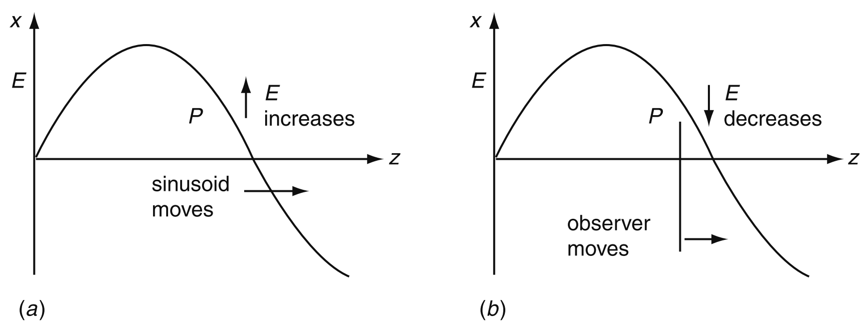


Figure 3.3 Evolution of field strength in time and space
(a) constant z: evolution with time
(b) constant time: evolution for increasing z

$$\omega = 2\pi f \quad \text{radians s}^{-1} \tag{3.6a}$$

and wavenumber k is given by:

$$k = 2\pi/\lambda \quad \text{radians m}^{-1} \tag{3.6b}$$

with f = frequency in Hz and λ = wavelength in metres.

Thus the temporal and spatial variation of field strength E for the coordinate system used in Figure 3.1 can be expressed efficiently in complex form by:

$$E = \exp\,(j.[\omega t - kz]) \quad \text{V m}^{-1} \tag{3.7}$$

noting that for a plane wave E is constant with x and y.

If only the spatial variation of E at a given instance of time is of interest, then t can be set to zero, and E is given by:

$$E = \exp\,(-j.kz) \quad \text{V m}^{-1} \tag{3.8}$$

This form for the complex representation of a signal is widely used in radiowave propagation.

3.1.4 Linear, circular and elliptical polarisation

The plane wave shown in Figure 3.1 is linearly polarised. The direction of linear polarisation is, by convention, defined by the direction of the electric field. Thus in figure 3.1, if the x axis is vertical, the wave would be described as being vertically polarised.

Circular polarisation can be viewed as two linearly-polarised waves, equal in amplitude, mutually orthogonal and in quadrature. If the x and y electric fields are unequal in amplitude the wave will be elliptically polarised. A right-hand circularly-polarised wave is illustrated in Figure 3.4. Reversing one of the fields will change the polarisation to left hand.

3.2 Multipath propagation

Multiple radio paths (or rays) between a transmitter and a receiver can occur for various reasons, such as the combination of ground and sky waves, multimode sky-wave propagation, atmospheric refraction and layer reflection, and reflections from the ground and objects such as buildings. It is called multipath propagation, and can give rise to several effects on the received signal.

3.2.1 Reinforcement and cancellation

Multiple rays can combine constructively or destructively according to relative phase. The result is a standing wave pattern in both space and frequency. Cancellation between two rays occurs when the difference in path length is given by:

$$\Delta d = (n + 0.5)\lambda \quad n = 0, 1, 2\ldots \tag{3.9}$$

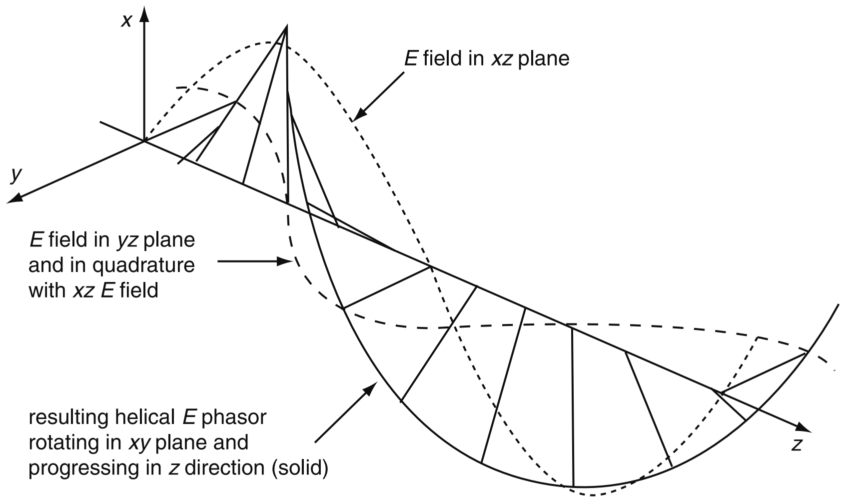


Figure 3.4 Orthogonal E fields in a circularly-polarised wave

Thus for a constant path length difference Δd, cancellation occurs when the frequency is given by:

$$f = (n + 0.5)\,\frac{c}{\Delta d} \quad n = 0, 1, 2 \ldots \tag{3.10}$$

where c = velocity of propagation.

In practical units, nulls due to cancellations occur at intervals of a wavelength in path difference, and of $300/\Delta d$ MHz in frequency where path difference, Δd, is in metres.

Figure 3.5 shows the variation of field strength with frequency for two rays

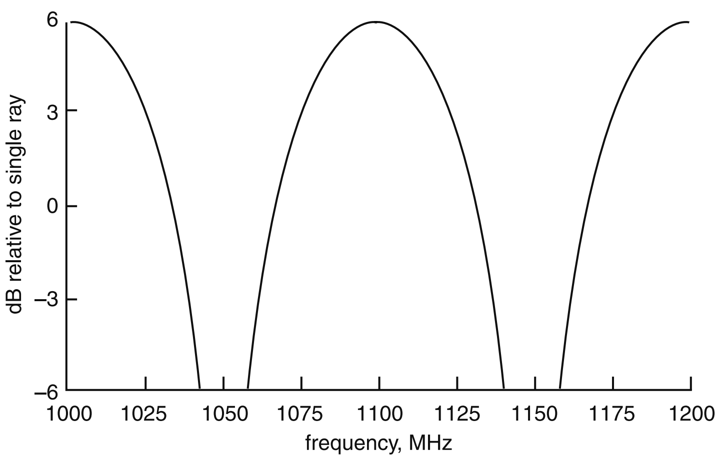


Figure 3.5 Field strength variation for a path difference of three metres

having a path length difference of three metres. The pattern of cancellation and reinforcement repeats every 100 MHz. The field strength is in dB relative to a single ray. At perfect reinforcement the enhancement is 6 dB, consistent with the E phasors summing to produce twice the single-ray value. The nulls will, in theory, be at minus infinity for equal amplitude rays. Although this is vanishingly improbable in practice, it is characteristic of multiple-ray standing-wave patterns plotted in dB that the nulls are narrow and deep compared to the reinforcements.

The scales in Figure 3.5 have been chosen to demonstrate that the median field strength is 3 dB higher than the single ray case, consistent with twice the single-ray power flux. The same type of field strength distribution will exist when Δd is varied for a given frequency.

3.2.2 Flat and selective fading

Many propagation attenuation mechanisms, such as shadowing by obstructions and attenuation by atmospheric gasses and rain, have frequency dependence, but the variation of loss is usually small over a radio system bandwidth. These mechanisms are referred to as flat fading.

Multipath propagation, on the other hand, can produce large signal variations over a system bandwidth which is referred to as frequency-selective fading, often abbreviated to selective fading.

This can result in the type of situation illustrated in Figure 3.6, which shows a somewhat idealised selective-fading situation with a well-defined multipath minimum within the bandwidth of a radio system. Because the minimum covers a much smaller range of frequencies than the system bandwidth, it is intuitive to assume that it will be less susceptible to the fade. This is indeed the case. Measures are described in later chapters for not only countering selective fading but also exploiting it.

3.2.3 The broadband channel and correlation bandwidth

The relationship between selective fading and a radio system bandwidth leads to the concept of the broadband channel. In communication terms the concept of

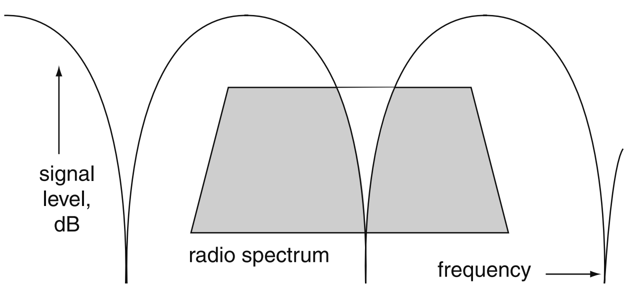


Figure 3.6 Frequency-selective fading

bandwidth can be interpreted on the basis of information capacity, e.g. a high-bandwidth internet connection. In propagation terms it refers to the relationship of the radio system bandwidth to a property of the radio paths known as correlation or coherence bandwidth. Figure 3.7 shows a hypothetical graph of the correlation coefficient over time between the levels of two signals propagating between the same pair of antennas, plotted against their frequency difference. The correlation bandwidth of the radio channel is the frequency offset at which the correlation coefficient has fallen to a stated level.

Figure 3.7 has been discussed in terms of multipath conditions which change over time. This would be appropriate for assessing multipath effects due to atmospheric layering for a line-of-sight link. For mobile systems the correlation of signal levels against frequency offset would typically be assessed as one terminal is moved. In propagation terms a radio system is considered broadband if its bandwidth is greater than the correlation bandwidth of the channel over which it operates.

3.2.4 Fast and slow fading

Multipath propagation due to reflections from the ground, buildings etc., produce an irregular pattern of reinforcement and cancellation due to multiple reflections with dimensions of the order of a wavelength. These are superimposed on larger-scale fluctuations in field strength due to attenuation by obstructions such as building or hills.

Figure 3.8 shows an example of this type of signal variation. Where a large number of approximately equal-amplitude reflected rays exist the small-structure signal variation will be approximately Rayleigh distributed. The large-structure variation due to obstructions will typically be log-normally distributed.

For nonline-of-sight services, such as radio broadcasting and mobile radio, it is impracticable for a propagation model to predict accurately the fine structure

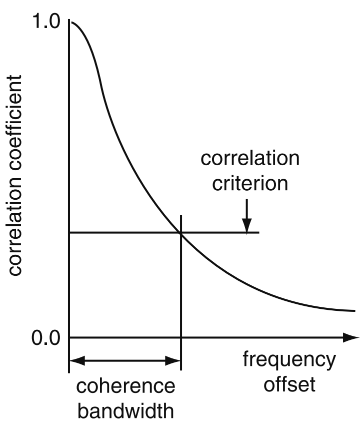


Figure 3.7 Correlation bandwidth

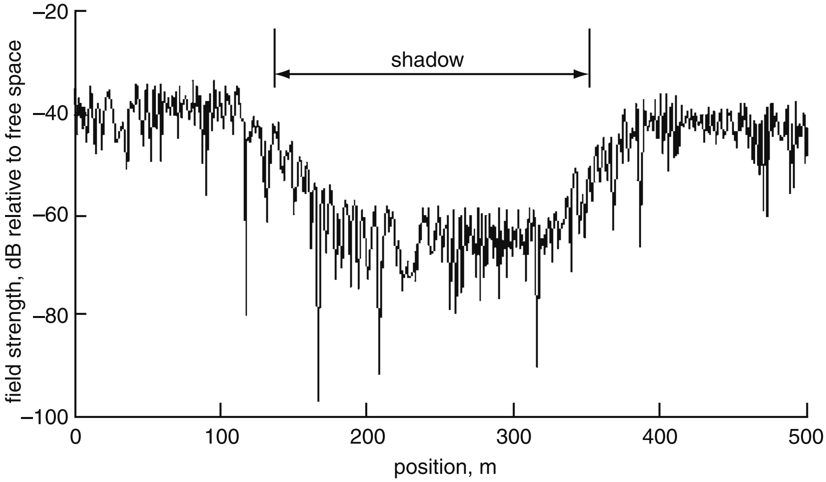


Figure 3.8 Fast and slow fading

of fast fading, particularly for outdoor systems. If topographic data showing the location of individual buildings, etc., is available then it is practicable to predict shadowing effects with the fast fading averaged. Otherwise the most common approach is to predict the statistics of shadowing, particularly the signal exceeded at 50 per cent locations within a limited area.

3.2.5 Delay spread

Another consequence of multipath propagation is that the contributions made by each ray are spread in time of arrival at the receiver due to differences in path length. This causes the well known ghosts in analogue television pictures. In digital systems it can cause inter-symbol interference, which is not correctable by increasing the received power.

The effect is often generally referred to as delay spread, although delay spread as such is only one of several parameters with exact definitions used to characterise time dispersion in a radio channel. These are illustrated in Figure 3.9, which is based on Recommendation ITU-R P.1145 Figure 1. The most widely used parameters are:

(a) *Average delay* is the power-weighted average of the excess delays for the power received above a stated cut-off level from t_{LOS} to t_3, t_{LOS} being the line-of-sight delay time.
(b) *Delay spread* is the power-weighted standard deviation of the excess delays.
(c) *Delay window* is the time from t_1 to t_2 containing a stated percentage of the total energy above the cut-off level.
(d) *Delay interval* is the time from t_4 when the signal level first exceeds a threshold P_{th} to t_5 when it drops below it for the last time. In this example, P_{th} is 15 dB below the maximum power density.

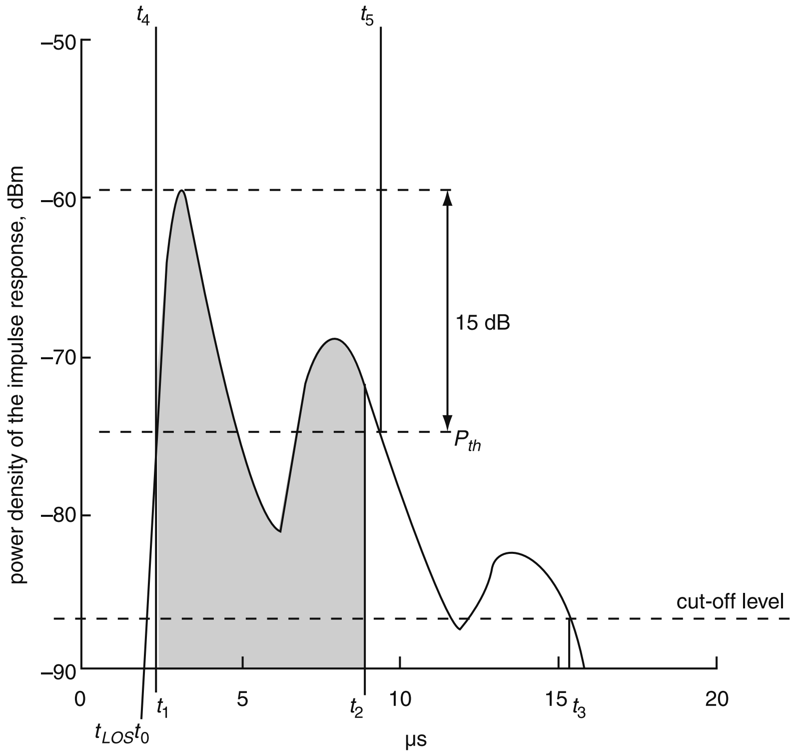


Figure 3.9 *Example of averaged delay power profile showing delay spread parameters (the parameters concerned are 't' with various suffixes, as defined in the text of §3.2.5)*

3.2.6 Frequency spread, or Doppler

If either or both the transmitter or receiver are moving, or if objects causing reflections, such as road traffic, are moving, then reflected signals will be shifted in frequency by the Doppler effect. Although in many cases such frequency shifts are small, Doppler can be a major design consideration, for instance in mobile systems used in high-speed vehicles.

The Doppler shift for each ray is proportional to the rate of change of path length. For stationary reflection points this will be a maximum in parallel with the mobile's motion. A reflection from in front will cause a positive Doppler shift, and a negative shift from behind. Figure 3.10 illustrates a resulting effect which is characteristic of mobile radio systems in an urban environment.

Mobile M will tend to receive reflections from along the street, which in most cases is parallel to the mobile's motion. The spectrum of Doppler shifts will thus show maxima close to the maximum possible Doppler shifts determined by the mobile's speed.

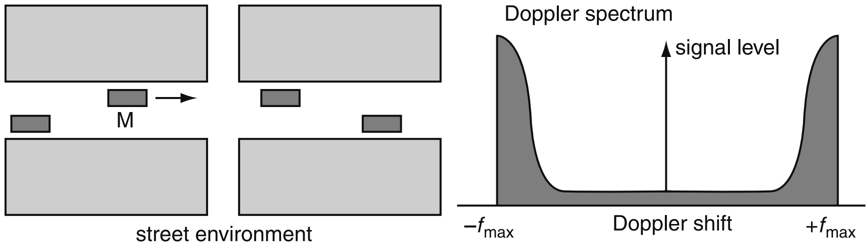


Figure 3.10 Doppler spectrum in an urban mobile environment

3.2.7 Direction of arrival

Yet another result of multipath propagation is that different versions of the received signal may arrive from different directions. This is particularly true for mobile radio systems in an urban environment, where there is considerable scope for improving system capacity by adaptive antennas used to exploit the effect.

3.2.8 Conclusions

The combination of the vector nature of radiowaves and the effects of multipath propagation introduces a wide range of issues in addition to the relatively simple question of signal strength.

When performing measurements, for instance, it is essential to be clear as to what range of the spatial or frequency distribution is to be captured. It is often impracticable to measure the local standing-wave pattern in a reproducible manner, and in most cases this will be highly time variant. Assuming that it is intended to measure larger-scale signal characteristics, it is important to distinguish between relatively small-scale variations due to, say, building shadowing from larger-scale changes determined by the complete propagation path.

There is also an increasing demand for propagation models to provide many more parameters, or the statistics of these parameters, than simple field strength or basic transmission loss. Digital high-capacity radio systems can be designed only with reliable knowledge of the characteristics in frequency, space, time and direction-of-arrival for the channels over which they will be used.

3.3 Footnote: a useful mathematical approximation

There is a simple mathematical approximation which is widely used in radiowave propagation. It concerns situations where a right-angled triangle is long and thin, that is, the hypotenuse is only slightly longer than one of the other sides. This type of geometry occurs in connection with small differences in radio path lengths, and in calculations involving the curvature of the earth. The approxima-

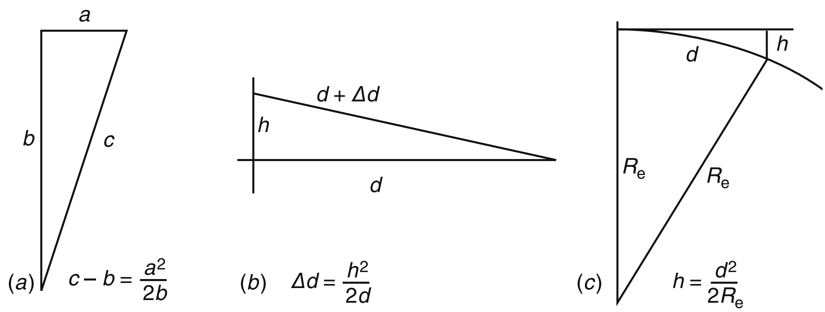


Figure 3.11 The parabolic approximation for small differences
(a) basic geometry
(b) path length difference
(c) effect of earth curvature

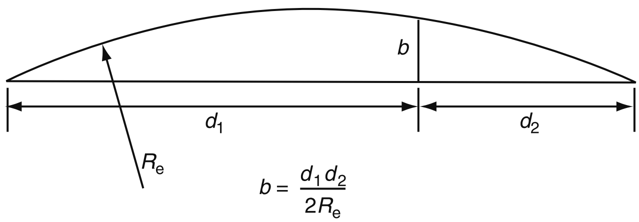


Figure 3.12 The parabolic approximation for earth bulge

tion is sufficiently accurate for most practical purposes, and is much simpler than the alternative full calculations.

The basic geometry is shown in Figure 3.11*a*. If $a \ll b$, then the difference between c and b is given to a good approximation by:

$$c - b = a^2/2b \tag{3.11}$$

In Figure 3.11*b* this approximation is used to calculate the small difference in radio path length Δd for a ray starting at height h compared with the horizontal path of length d.

In Figure 3.11*c* it is used to calculate the height difference h between a local horizontal and a smooth-earth path at distance d. For many radio paths the approximation $d \ll R_e$, where R_e is the radius of the earth, holds sufficiently well that it is immaterial whether d is measured horizontally or along the curved surface of the earth, and it is acceptable for h to be viewed as being normal to the local horizontal.

An extension of this approximation is illustrated in Figure 3.12. This is useful in calculating the bulge due to earth curvature along a radio path above a Cartesian baseline drawn between the ends of the path.

Chapter 4

Propagation data requirements for radiocommunications system planning

Tim Hewitt

4.1 Introduction

It is neither economically realistic nor spectrally efficient to engineer a radio-communications system to the extent that it can be guaranteed to work for 100 per cent of the time, and realistic allowances have to be made for equipment failures, propagation problems, and interference. When planning a system it is therefore necessary to have design targets for the maximum percentages of time for which the circuit may be interrupted, or its error performance degraded.

The amount of time for which a degraded level of service can be accepted is dependent on the particular application. High-grade telecommunications systems are very tightly specified in this respect, typically needing to be fully operational for all but few hundredths of a per cent of the time. Even local access radio systems generally need to work to 99.99 per cent of the time, or better. Broadcasting is more tolerant, with perhaps 0.1 per cent of time being allowed for interruptions. However, mobile networks, which live in an interference-dominated environment in which performance is statistically less certain, are normally designed for median conditions and are not so concerned about the smaller time percentages.

Once total allowable time percentages for degraded performance and outage have been decided, they generally need to be apportioned (not necessarily equally) between equipment failures, propagation effects and interference. Furthermore, for example, the propagation allowances may need to be further apportioned between several tandem hops in a radio network. In some cases this can result in very small time percentages indeed being allowed for propagation and interference effects on each individual radio hop. The system designer is then working with the tails of the wanted signal and interference propagation distributions, and with all the uncertainties that these represent.

In practice, the quality of a digital radio circuit is defined and/or measured in two ways, i.e. by its availability for the transport of information, and by its error performance in terms of the number of errors that occur while it is available. These two quality objectives are discussed in greater detail below.

Designing radio systems to recommended performance and availability targets raises a number of issues relating to the types of propagation information required. Unfortunately, as things currently stand, the expectations of the design engineer cannot always be matched by the available propagation data and prediction models. It is therefore necessary for the designer to appreciate the limitations of current propagation knowledge, and to devise pragmatic solutions that will allow the design work to move forward within the constraints of the data that is available.

The notes below explore the general issues associated with performance and availability requirements for modern digital radiocommunications systems, as set out within the relevant ITU-R Recommendations. The emphasis is on examples that illustrate the various propagation issues that can arise, and the notes are not intended to be a tutorial on radio link design.

4.2 Quality parameters for radiocommunications circuits

4.2.1 Key terms and definitions

When discussing radio link design it is necessary to get a little involved in the language of that art. In the context of this present discussion there are four terms that are of particular relevance to the planning of digital radiocommunications systems:

(i) *Errored seconds* (ES) are seconds during which errors occur in the information bits being carried. The errors will normally be in the context of the baseband (information) signal rather than the RF channel. Modern digital systems with adaptive demodulators and error correction can do much to correct the basic transmission errors. Trade-offs between system sophistication and fading and interference margins are therefore available, and are nowadays used to excellent effect.

(ii) *Severely errored seconds* (SES) are seconds during which the error rate (typically at the 64 kbit s^{-1} level) within the 1 s period exceeds a pre-defined threshold, normally one error in 10^3 bits.

(iii) *Error performance* is the measure of the transmission quality that a circuit should, or does, achieve while it is available for the transport of information. For digital circuits, performance is described in terms of the errored seconds ratio (ESR) and the severely errored seconds ratio. These are the ratios of errored and severely errored seconds, respectively, to the total number of available seconds, usually in any month.

(iv) *Availability* is the percentage of the total time during which the circuit should be, or is, delivering the required performance, i.e. the total time that it is not out of service due to equipment, propagation or interference problems. For digital circuits, a period of unavailability begins at the start of a sequence of ten consecutive SES, and ceases at the start of the next ten second without any SES.

Recommendation ITU-R F.557-4 states '. . . that in the estimation of unavailability, one must include all causes which are statistically predictable, unintentional and resulting from the radio equipment, power supplies, propagation, interference, and from auxiliary equipment and human activity, and that the estimate of unavailability should include consideration of the mean time to restore'.

This indicates that the overall recommended unavailability allowance must be apportioned between these different causes. This might imply that both the fading of the wanted signal and the statistics of unwanted interference must be analysed together to ensure that the quality objectives are met. How the allowances are apportioned between the different causes is a fundamental decision for the design engineer.

It is therefore necessary to have propagation models to predict the propagation conditions on the wanted path (e.g. Recommendation ITU-R P.530 for terrestrial paths or Recommendation ITU-R P.618 for earth–space paths), as well as those to predict the levels of interference between the wanted system and other cochannel systems (e.g. Recommendation ITU-R P.452). These important propagation recommendations are described in other chapters.

4.2.2 Related propagation issues

Ideally, the design engineer who is planning the system should have available propagation data that will enable him to demonstrate that the design will meet both the performance and the availability objectives.

This provides the propagation engineer with many significant challenges. For a digital radio circuit the concept of availability, as defined above, poses special problems in the context of propagation data.

Figure 4.1 shows a short period of transmission on a digital radio link. During this period fading is occurring, sometimes to a significant extent. Within the first 24 seconds the fading is significant, but the event durations are short and there are no continuous periods of SES of ten seconds or longer that would count as unavailability, as defined above. All of the errors in this period will therefore count towards the allowance for error performance degradation, i.e. they will contribute to the ESR and/or the SESR. From the 24th second there is a continuous block of 12 SES, and the start of this block constitutes the onset of availability. At the 36th second there begins a period of short events with the SES events being <10 s. However, as there is no clear ten second period without SES until after the 48th second, the unavailability continues until this latter point. Beyond this there are no further SES and the circuit again becomes available. The ES in the 49th second will count towards error performance degradation.

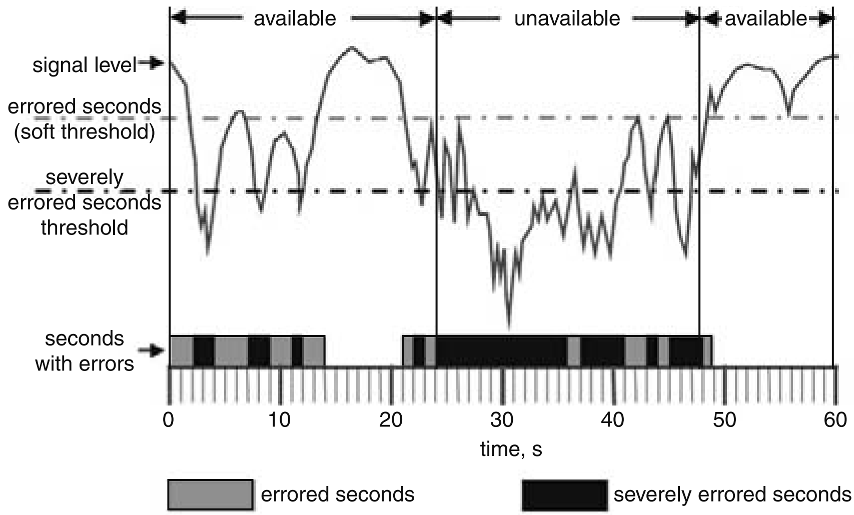


Figure 4.1 Fading and consequential errors on a digital radio system

What this means in propagation terms is that:

- all fade events that cause ten or more consecutive SES will add to the unavailability;
- fade events causing periods of SES <10 s will either count towards the error performance degradation, or they will add to the unavailability, depending on how they are positioned in the general sequence of SES fading events.

To allow the design engineer to undertake a performance/availability analysis strictly in accordance with ITU-R performance and availability objectives, for both wanted signal and interference scenarios, propagation measurements would have to be available to better than one second resolution. Such data would need to be analysed in terms of the second-by-second characteristics of the propagation medium, with events not only being classified by duration greater or less than ten seconds, but also in terms of the positioning in time of the shorter events compared to that of the longer events. Such data are not generally available, but some new data have recently been provided and some examples are given in Section 4.3.

4.2.3 A note of caution

In Figure 4.1 reference is made to a soft threshold for errored seconds. This is soft because there is no specific level at which errors begin. For link design, however, some sort of threshold is needed. For discussion purposes we might consider that a significant ESR begins at an error rate of 1 in 10^7, but it will depend on the particular application. The margin between this soft threshold and the hard SES threshold is worth exploring a little further as it has some propagation data implications.

Figure 4.2 shows some error rate versus signal-to-noise (S/N) ratio curves for a number of different coding methods that might be used for a modern digital radio link. Curve *e* is for uncoded data, but curves *d*, *c*, *b* and *a* are for increasingly sophisticated coding techniques ranging from block coding in curve *d* through trellis coding in *c* to complex multilevel coding in curves *b* and *a*. The

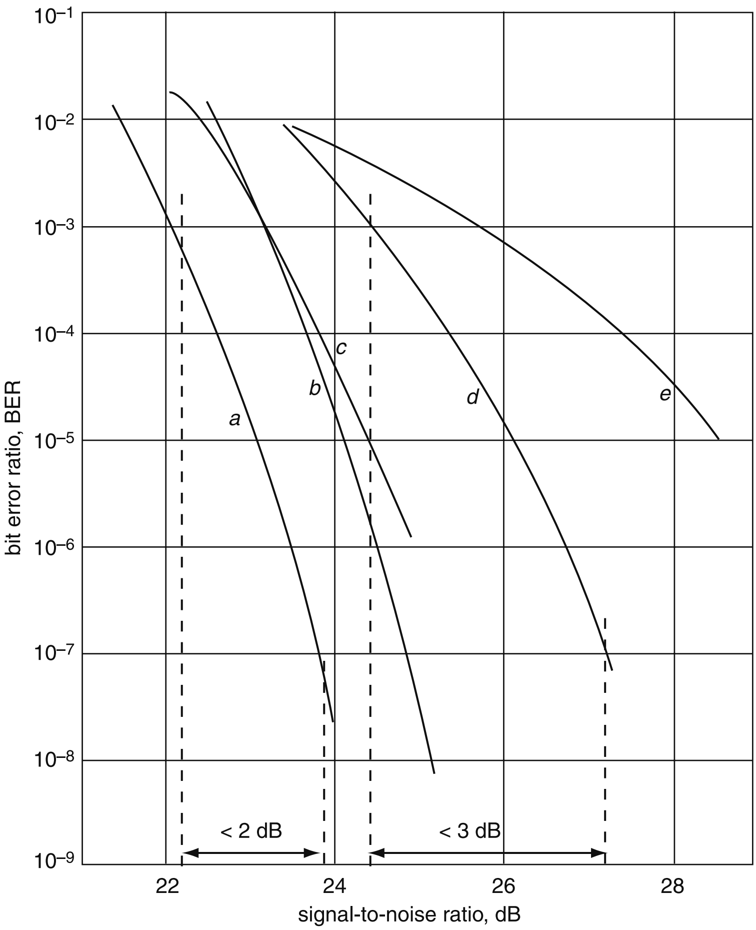


Figure 4.2 Error ratios for different coding methods
(a) MLC (6.29/7)
(b) MLC (6.5/7)
(c) TCM-4D (6.5/7)
(d) BCH (511,484)
(e) uncoded

actual coding schemes are not important here, but further detail can be found, if required, from Recommendation ITU-R F.1011.

Of consequence here is the S/N ratio margin between the SES threshold of 1 in 10^3 error ratio, and what we have chosen to represent as the onset of significant numbers of errored seconds at 1 in 10^7 error ratio. For the simple block coding (curve *d*) this margin is less than 3 dB, for the most advanced multilevel coding (curve *a*) it is less than 2 dB. These are very small differences in terms of the accuracy of propagation models, and great care must be taken not to claim a greater accuracy than the propagation data can deliver. If the SES error ratio is critical in a design, a margin might need to be added to accommodate the standard deviation of the propagation model, otherwise the predicted errored seconds could in reality easily end up as severely errored seconds!

4.3 Event durations

It is clear from the discussion of unavailability that the durations of fading events are all important. There has been very limited data available on this topic until quite recently. However, there is currently a campaign in ITU-R Study Group 3 to collect such information, and some very high quality data is at last becoming available.

By way of an example, Table 4.1 shows some statistics of the numbers of (rain fading) events of >0 s and ≥10 s duration, collected from an 18 GHz propagation experiment in Sweden. The data were usefully analysed to give the total unavailable time (i.e. made up of events >10 s) as a percentage of the total fading time at each attenuation threshold.

The unavailable time is clearly by far the greater portion of the total time, i.e. greater than 95 per cent for all thresholds. However, it should be noted that there

Table 4.1 Unavailable time as a function of total fading time

Attenuation (dB)	10	15	20	25	30	35	40
Number of fade events >1 s	888	85	36	22	10	38	2
Mean duration of fade events >1 s	65.3	122.7	160.5	153.2	187.0	25.6	226.0
Total outage time (s)	57986	10430	5778	3370	1870	973	452
Number of fade events ≥10 s	259	47	26	14	4	4	1
Mean duration of fade events ≥10 s	219.1	219.4	220.7	238.4	461.0	231.0	452.0
Total unavailable time (s)	56747	10312	5738	3338	1844	924	452
Unavailable time in % of total outage time	97.9	98.9	99.3	99.0	98.6	95.0	100

Source: ITU-R Doc. 3M/88, February 1999 (Telia AB, Swden) is a possible error here because, as Figure 4.1 suggests, some of the 0 to < 10 s events (and indeed the short gaps between them) could be contributing to the unavailability. No more detailed information exists as to the timing of the short events in relation to the >10 s events, so we might assume that 50 per cent of the <10 s events contribute towards the SES time, and 50 per cent towards unavailability. Figure 4.3 illustrates the situation.

4.4 Digital radio networks

4.4.1 General considerations

Later in this chapter we will be examining some of the performance and availability objectives provided within the ITU-R Recommendations for the fixed service. The early fixed service digital radio networks came into being to support, *inter-alia*, the roll out of the Integrated Services Digital Network, or ISDN. International agreement on quality objectives was crucial, and several important ITU-R fixed service recommendations address the use of digital radio links within the high-grade (i.e. core or backbone), medium-grade (or regional) and local-grade (local access) parts of the ISDN. These different elements of the network have different performance and availability criteria, as will be seen later.

Figure 4.4 shows a schematic representation of part of an ISDN based on radio transmission. The high-grade portion comprises the long-distance backbone route carrying several 140 Mbit s^{-1} bearers, and it must therefore offer an overall high degree of availability.

The first digital radio systems were derived from the earlier analogue radio technologies. A route therefore comprised some terminal stations with multiplex

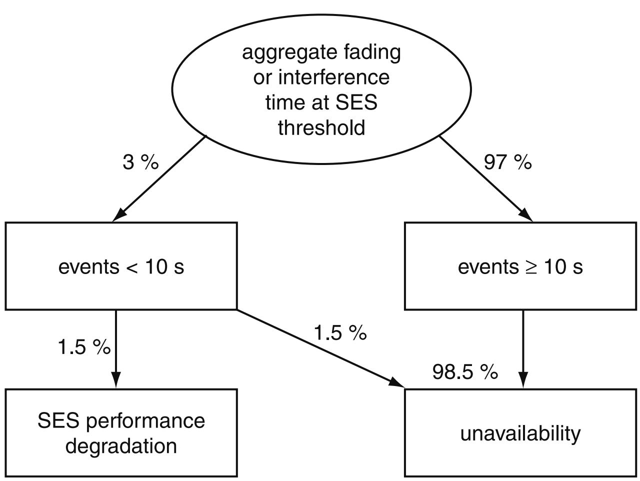


Figure 4.3 Propagation events contributing to performance degradation and unavailability

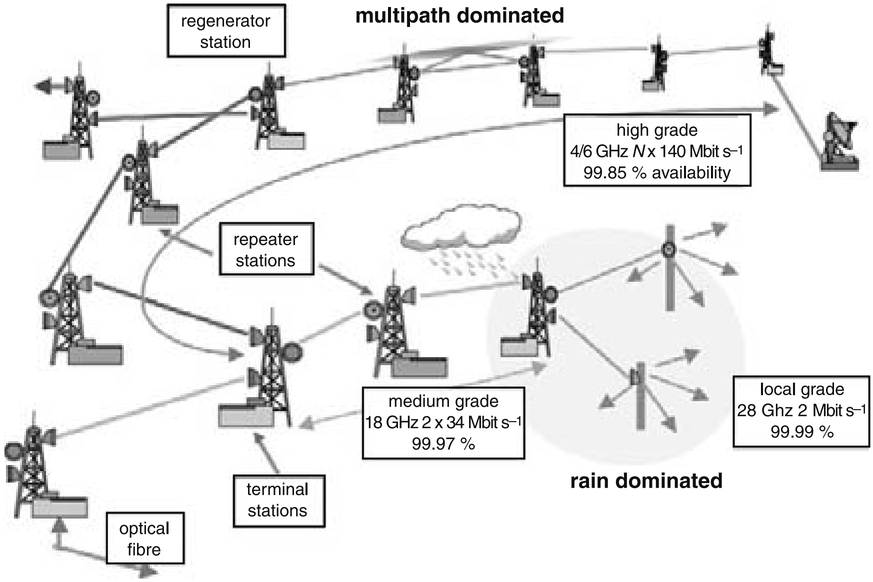


Figure 4.4 Radio network with high, medium and local grade sections

equipment to get from the digital baseband to the RF channels, and a number of repeater stations which relied on down conversion, IF amplification and up-conversion to achieve the repeater function. These repeaters did not therefore demodulate the digital signal down to the baseband level.

However, the accumulation of noise and jitter through such a chain of linear repeaters limited the number of such repeaters that could be used without regeneration of the baseband signal. Thus, for long routes, some repeaters were replaced by regenerator stations provided with modulators/demodulators and, where required, multiplex/demultiplex equipment. One such station is shown in Figure 4.4. These regenerator stations were often conveniently sited where there was a node in the radio network, because the same baseband and multiplex equipment was needed to split/combine baseband traffic from different routes.

The high-grade portion was needed to cover long distances, and thus low frequencies were used (typically 2 GHz, 4 GHz, 6 GHz and 11 GHz). Each hop carried typically eight both-way 140 Mbit s^{-1} channels, and thus a considerable amount of traffic was involved. The length of the hops (30–45 km in the UK) meant that multipath fading was the key problem to overcome. Techniques such as $N + 1$ protection channels (giving a measure of frequency diversity) and antenna height diversity were used from the outset to combat multipath fading and hence ensure a high availability for this traffic.

At the ends of the core routes the traffic was split between a number of medium-grade radio, coax or fibre routes, each carrying less traffic. These routes could sometimes be allowed slightly less demanding performance and availability objectives.

Shorter distances might encourage the use of higher frequency bands (18 GHz, 22 GHz etc.) which would be rain dominated. The shorter hop lengths needed to combat the rain fading meant that multipath was a secondary issue on these links.

Although most local access is *via* the copper network, radio does play a part. High frequency links (18 GHz, 22 GHz, 28 GHz, etc.) are used in point-to-point and point-to-multipoint configurations to deliver ISDN to customers. Radio is particularly important when the ISDN service is provided to customers at the $N \times 64$ kbit s^{-1} levels.

4.4.2 Modern digital radio systems

In recent years digital radio technology has moved forward in leaps and bounds. Many new digital links are built for networks using the synchronous digital hierarchy (SDH) which allows individual traffic circuits to be added and dropped, even at the 64 kbit s^{-1} level, at each station on the network. This requires access to the baseband signal at all such points, and the older-type linear repeater stations are no longer appropriate.

Modern digital equipment using advanced modulation and coding techniques allows for adaptive equalisation of the channel and a high degree of error correction on every hop. Slipless protection switching and advanced diversity combiners allow an essentially no-break use of mitigation techniques. More of the unavailability allowance therefore becomes available to deal with propagation problems. However, the basic challenges of designing for good performance and a high availability remain as before.

4.4.3 Related propagation issues

To design complex radio networks to meet the demanding performance and availability objectives of an ISDN, the planning engineer needs propagation data to predict multipath and/or rain fading, and information that will help to design the mitigation techniques such as frequency diversity and antenna height diversity. Information is also needed on interference propagation between the cofrequency stations in the network, and possibly between entirely separate cofrequency networks.

4.4.4 Simultaneous fading

The longer routes will involve a number of high-grade and/or medium-grade hops. The ITU-R performance objectives are specified for the whole length of a high or medium-grade route, and it is up to the design engineer to decide how many hops are needed and how the unavailability and performance degradation allowances are to be shared between the individual hops.

A key issue is whether the fading experienced by tandem hops is simultaneous, partly simultaneous or wholly independent. If the outage on the various hops is not correlated, then the overall allowances must be subdivided between all the hops (not necessarily equally). This can lead to very demanding unavailability and performance degradation allowances for an individual hop.

There are few propagation data relating to the correlation of fading on separate hops, but the topic is amongst the study questions of ITU-R Study Group 3. By way of an example, Figure 4.5 shows a recent experiment in Japan to collect simultaneous (worst month) fading statistics from long 4/6 GHz links that might form part of a high-grade circuit.

The source document from which this information was extracted provides data for many combinations of these hops, and only a sample of the data is shown in the table. This was a truly excellent study. Table 4.2 shows a small

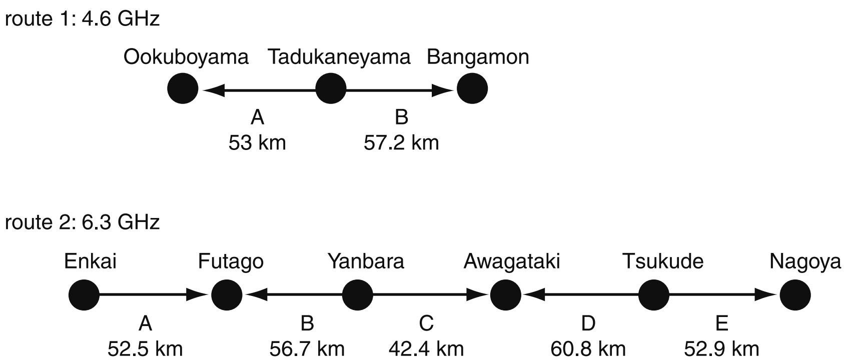


Figure 4.5 Simultaneous propagation measurements in Japan
Source: ITU-R Doc. 3M/87, February 1999 (Japan)

Table 4.2 Statistics of simultaneous fading on different hops

Route 2: 6.3 GHz, hops B and C					
Percentage of worst month	0.001	0.002	0.003	0.005	0.010
Fade depth, hop B (dB)	25.8	24.0	22.9	20.8	18.5
Fade depth, hop C (dB)	29.5	24.3	21.0	18.0	13.5
Simultaneous fade depth (dB)	8.4	7.8	7.0	6.4	4.8

Route 2: 6.3 GHz, hops B and D					
Percentage of worst month	0.001	0.002	0.003	0.005	0.010
Fade depth, hop B (dB)	25.8	24.0	22.9	20.8	18.5
Fade depth, hop D (dB)	27.8	25.7	24.8	23.5	21
Simultaneous fade depth (dB)	4.6	3.3	2.7	1.7	1.5

Source: ITU-R Doc. 3M/87, February 1999 (Japan) sample of the results from Route 2 of Figure 4.5. The upper part of the table addresses the simultaneous fading on hops B and C of Route 2, and the second extract is for hops B and D. The data is shown for the high fading end of the distributions, i.e. for time percentages from 0.001 per cent to 0.01 per cent. However, the data submitted to the ITU-R by Japan covers a much greater time percentage range.

What is apparent from both these extracts is that, although all the hops suffer fading of up to about 30 dB for small time percentages, the levels of simultaneous fading are quite small (8.4 dB maximum, even at 0.001 per cent of time). From these data one must conclude that all the significant fading (i.e. that likely to cause errored seconds or severely errored seconds) is uncorrelated. This is bad news for the radio planner.

These data are for multipath fading. The limited work that has been done on rain fading on horizontally spaced or tandem paths shows broadly similar trends.

4.5 Long-term and short-term criteria

4.5.1 The long-term criteria

Customers for high-grade telecommunications services do not like errors. Indeed when the error rate is important, customers often keep their own regular check on the performance and availability of the circuits for which they are paying. The challenge for radio networks has always been that coaxial cable and, more recently, optical fibre systems provide very high quality circuits, and for radio to be a useful option for telecommunications networks, the same demanding quality objectives have to be achieved.

Very short periods of errors due to propagation etc. are generally accepted, but regular errors under normal conditions are not. A margin must therefore be established above the noise level of the receiver such that the normal error rate is very small indeed. This margin, together with the associated time percentage, is generally known as the long-term criterion.

The long-term performance requirements of modern digital telecommunications circuits are very challenging indeed. Typically, for all but 10 or 20 per cent of time, the expectation for fixed service links is an error rate of 1 in 10^{11} under normal conditions. This represents one errored second in about 434 hours at the 64 kbit s^{-1} level. This leaves little scope for accommodating adverse propagation effects or interference problems.

Knowledge of the performance of the system hardware (e.g. the kind of data shown in Figure 4.2) allows the error performance and availability requirements to be translated into equivalent carrier-to-noise or carrier-to-interference ratio requirements.

For systems to achieve the required background error-rate performance, the design must provide a signal level that offers, for all but the stipulated time percentage, an adequate margin to offset propagation fading and an assured margin over predictable interference levels from other stations in the network and other spectrum users in the band. At around 10–20 per cent of time the propagation and interference conditions are reasonably predictable, and a reliable design can normally be achieved in this context.

4.5.2 The short-term criteria

For small percentages of time the propagation conditions and interference problems will worsen. When the required signal fades too deeply (due, for instance, to rain attenuation or multipath fading) and/or the incoming interference increases too much (perhaps due to ducting or hydrometeor scatter), the required system will eventually exceed a maximum permissible error rate (typically 1 in 10^3). When this occurs for periods longer than ten seconds, the system will be deemed to have become unavailable. Shorter periods of SES will affect the error performance. These situations must not be allowed to occur for more than a stipulated percentage of time. The performance and availability objectives must be translated into short-term fade and interference margins which must be achieved by the system design for all but the acceptable percentages of time.

4.5.3 A simple design exercise

It is useful here to reflect on how the performance and availability requirements can be matched to the available propagation data. This is a much-simplified example, and it should not be taken as being the proper way to design a radio link. The first step is to convert the relevant error rates mentioned in the objectives to the signal level margins which must exist above the noise and/or interference levels.

As noted, to meet the errored seconds requirements it is necessary to keep the nominal signal at such a level above the background noise level that virtually no errors occur for most of the time. This leaves the error allowances to be used up during the short periods when the propagation conditions become more difficult. The nominal (design) signal level should therefore equal or exceed the level required to achieve the necessary background error rate. The long-term threshold A is shown associated with 20 per cent of time in Figure 4.6.

However, the signal will undoubtedly fade for some percentage of the time, as shown by the dashed line in Figure 4.6. Point B on this figure represents the margin, relative to the system noise floor (assuming a noise-limited situation), at which the system will produce an error rate of 1 in 10^3. If the signal falls beyond this point severely errored seconds will occur and the hop could become unavailable. The design received signal level must therefore be such that the predicted fading distribution passes through this level at, or below, the time percentage associated with the short-term objective (in this case 0.01 per cent, purely chosen for the purposes of illustration).

Furthermore, there may well be short periods of enhanced levels of interference, as illustrated by the predicted cumulative distribution for interference

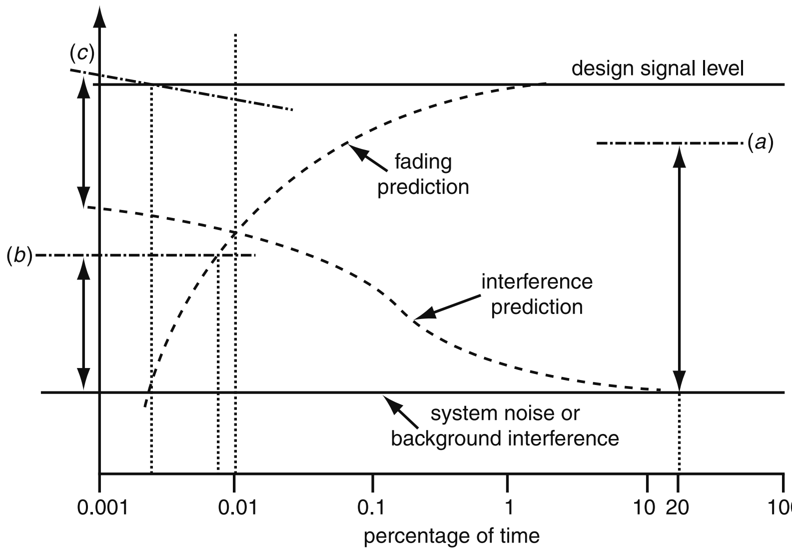


Figure 4.6 Long-term and short-term planning criteria
(a) long-term criterion: ≤ 1 × 10^{-11} error rate for all but 20 per cent of the time, showing required C/I (or C/N) margin
(b) short-term criterion: ≤ 1 × 10^{-3} error rate for all but 0.01 per cent of the time, showing required carrier to noise (C/N) margin
(c) carrier to interference (C/I) margin for 1 × 10^{-3} error rate

shown in the diagram. This interference will also cause errored seconds unless the received signal level has a sufficient margin over this distribution for at least the required time percentage. Such a margin is shown at C in Figure 4.6. To keep the total number of severely errored seconds within the overall 0.01 per cent allowance, it is necessary to ensure that the time, t_f, for which the fading exceeds the allowable margin B, plus the time, t_i, for which the signal level is below the required margin, C, over the enhanced interference, does not exceed the 0.01 per cent allowed. Juggling the fading and interference predictions can be quite difficult.

In this example it is the interference margin that is determining the design signal level. In practice it can be this, or the fading margin, or (less commonly) the long-term margin that determines the received signal requirement. It is therefore important to check all three, and not to assume that it will necessarily be the fading that determines the necessary signal level.

Finally, it will be seen in the next section that although availability requirements are normally referenced to as a percentage of an average year, the error performance objectives are usually described in terms of any month. In any design analysis of the type just discussed, care must be taken to convert all objectives and propagation predictions to the same time basis.

4.6 ITU-R performance and availability objectives

4.6.1 General

To identify the propagation issues associated with quality objectives it is useful to explore some of the relevant ITU-R Recommendations on performance and availability. These recommendations are an extremely valuable source of information, and for those involved in the design of radiocommunications systems it is well worth taking time to gain a good understanding of their contents.

Extracts from a number of ITU-R Recommendations are used in the discussion below. In general the emphasis is on digital systems as these are the subject of most new design activities.

Three types of quality objective are discussed:

- performance and availability for high-grade ISDN applications in the fixed service
- performance and availability for ISDN circuits in the fixed satellite service
- fixed service and fixed satellite service protection from interference in bands shared by these two services.

These particular examples have been selected to bring to the foreground some of the more important propagation issues, and they are not meant to be a complete treatment of the subject.

4.6.2 Fixed service performance objectives

4.6.2.1 ITU-R objectives

A good starting point for the fixed service is Recommendation ITU-R F.758–2. Tables 4.3 and 4.4, derived from this Recommendation, provide a set of references within which can be found the performance and availability objectives of various applications in the fixed service.

Table 4.3 Digital performance and availability objectives

Objective	*Hypothetical reference digital path (HRDP)*	*High-grade ISDN (core network)*	*Medium-grade ISDN (regional, or feeder network)*	*Low-grade ISDN (local access)*
Error performance	Rec. ITU-R F.594–4	Rec. ITU-R F.634–4	Rec. ITU-R F.696–2	Rec. ITU-R F.697–2
Availability	Rec. ITU-R F.557–4	Rec. ITU-R F.695	Rec. ITU-R F.696–2	Rec. ITU-R F.697–2

Table 4.4 Analogue performance and availability objectives

Objective	Telephony	Television	Transhorizon (troposcatter)
Noise in hypothetical reference circuit (HRC)	Rec. ITU-R F.393–4	Rec. ITU-R F.555–1	Rec. ITU-R F.397–3
Noise in real links	Rec. ITU-R F.395–2	—	Rec. ITU-R F.593
Availability	Rec. ITU-R F.557–4	—	—

To simplify the discussion, just the digital applications in Table 4.3 will be considered in the notes below.

Reference to Recommendations ITU-R F.594, F.634, F.696 and F.697 confirms that the performance requirements for ISDN digital circuits running over the fixed service are expressed in terms of errored seconds ratio (ESR) and severely errored seconds ratio (SESR) in any month.

As an example of digital performance requirements, for a high-grade ISDN link (here meaning a multihop route) of length L km (where $280 \leq L \leq 2500$):

- the errored seconds ratio for an $N \times 64$ kbit s^{-1} circuit ($N <= 32$) should not exceed $0.0032 \times L/2500$ in any month
- the SESR should not exceed $0.00054 \times L/2500$ in any month.

Note that these demanding targets relate to errors at the $N \times 64$ kbit s^{-1} level, and not to the raw digital radio signal.

4.6.2.2 Propagation in any month

The next important propagation issue, identified above, is that the required performance objectives must be achieved in any month.

Nevertheless, when talking about radio systems which may be subjected to the full range of propagation conditions that mother nature can deliver, this requirement is frighteningly open ended. Over the lifetime of a trunk radio system (20+ years) the propagation environment can be affected by some extreme weather conditions. The very heavy rain and floods in the UK in the Autumn of the year 2000 are a good example. To provide a design that would keep working to its full performance requirements through rare extreme fading and interference events would be expensive and inefficient in terms of the use of the spectrum.

The solution adopted by the radiocommunications community is the concept known as the worst month. A particular form of the worst month, known as the average worst month, was devised to provide a month-based propagation prediction that took account of the fact that the most significant events normally occurred in only a few calendar months, and the lower level events were more widely distributed through the year. The worst month is an important concept and it is discussed more fully in Section 4.7.

4.6.2.3 Errored seconds ratio

The error performance objectives for the high-grade systems set out in Recommendation ITU-R F.634 (and indeed those for the medium-grade circuits described in Recommendation ITU-R F.696) specify both ESR and SESR targets.

The SESR target is easiest to deal with as, *via* knowledge of the system hardware, it is possible to convert this to a defined attenuation threshold and a defined percentage of time.

The ESR target is more difficult, as it does not represent a fixed error rate, but rather a range of error rates from one per second up to 1 in 10^3. This translates into a range of attenuation values, with error rates increasing as the attenuation increases, bounded at its upper limit by the SESR attenuation threshold. To determine the ESR performance, the predicted propagation distribution (for events <10 s) would need to be convolved with the system fade/error characteristics.

As the right kind of propagation information is not available, and as the analysis would be complex, emphasis is generally placed on meeting the SESR objective.

The errored seconds ratio is protected in another way. As just one error can create an ES, system designers try to keep the normal signal level well above the background noise and interference, i.e. the general design goal is virtually no errors at all for most of the time. This leaves all of the ESR for the periods of short-term fading or interference. Having no errors is unrealistic, so a low error rate (typically 1 in 10^{11} for terrestrial links and 1 in 10^7 for satellite links) is set as a long-term objective, i.e. for all but 10 or 20 per cent of time.

4.6.3 Fixed service availability objectives

4.6.3.1 ITU-R objectives

For digital systems, Recommendation 3 of Recommendation ITU-R F.557–4 defines unavailability (of a both-way circuit) as '. . . at least 10 consecutive severely errored seconds (SES) in at least one direction of transmission'. The unavailable time for the both-way circuit is therefore, theoretically, the combination of the unavailable time for each direction of transmission, concurrent outage in the two directions only counting once.

Reference to Recommendation ITU-R F.695 shows the availability objective *A* per cent for a high-grade ISDN link of length L km to be:

$$A = 100 - (0.3 \times L/2500)\ \%$$

This is 99.7 per cent of the time for the 2500 km route.

However, what constitutes the time in this context remains, rather surprisingly, open to debate. Recommendation 1 of Recommendation 557 states '. . . the percentage being considered over a period of time sufficiently long to be statistically valid, this period is probably greater than 1 year; the period of time is under study . . .'.

On the assumption that the time in this case can be interpreted as an average year, then for a 300 km route, the value of A will be 99.976 per cent (an unavailability of only ~7568 seconds).

Recommendation ITU-R F.696 shows that for medium-grade ISDN circuits there are four classes of system and that the unavailability in these cases is not distance dependent and ranges from 0.033 to 0.1 per cent of time. Again, these represent time percentages of '. . . a period probably greater than one year . . .'.

Recommendation ITU-R F.697 shows that for local-grade ISDN circuits 'the unavailability might range from 0.01% averaged over 1 year to up to 1.0% averaged over one or more years for a bi-directional radio system'.

There is clearly no single consistent definition of the base-line time for fixed service availability, even after they have been in use for more than 30 years!

4.6.3.2 Propagation issues

The performance objectives discussed above were specified in the context of any month. Here we have a different requirement for propagation data relating to an average year.

The lack of clarity over the base-line time period to which the availability objectives must be referenced is surprising. However, propagation prediction models are generally based, one way or another, on measured data, and the most reliable data is that collected over a number of years in order to get a representative average year. The pragmatic solution offered by the propagation community is therefore the use of average year (otherwise known as average annual) statistics.

The ISDN performance and availability objectives generally apply to a whole high-grade (core network) route or a whole medium-grade (regional) route. These error performance and availability objectives cannot therefore be directly applied to a single hop on a route. The allowances for a single hop are generally smaller.

Given that the unavailability also has to be apportioned between a number of causes, it becomes quite a complex task to determine the propagation fading allowances for a particular hop, as they depend on the mitigation techniques available (e.g. antenna height diversity, protection channels, error correction) on the system to combat the effects of propagation fading and interference.

4.6.4 Fixed satellite service performance objectives

4.6.4.1 ITU-R objectives

It is possible to look at the comparable situation for the fixed satellite service, which generally provides high quality circuits over earth–space propagation paths.

The ITU-R S. Series of Recommendations provides a set of texts dealing with the performance and availability objectives for the fixed satellite service (FSS). Because of the particular nature of satellite circuits, and their heavy dependence on coding, the FSS quality-of-service requirements tend to be more complex than their fixed service counterparts.

With respect to the requirements for protecting FSS performance objectives, we can look to Recommendations ITU-R S.614-3 and S.1062-1.

For example, Recommendation ITU-R S.614 provides the allowable error performance for an HRDP in the fixed satellite service operating below 15 GHz when forming part of an international ISDN connection.

Recommendation 1 of Recommendation 614 states that: 'The bit error ratio (BER) at the output (i.e. at either end of a two-way connection) of a satellite HRDP operating below 15 GHz and forming part of a 64 kbit s^{-1} ISDN connection should not exceed during the available time the values given below:

- 1×10^{-7} for more than 10 per cent of any month,
- 1×10^{-6} for more than 2 per cent of any month,
- 1×10^{-3} for more than 0.03 per cent of any month'.

In the case of this FSS ISDN application, the long-term error rate is specified as 1 in 10^7 at ten per cent of any month. In the FSS the link power budget has to be tightly managed, and it is not possible to have a more generous background error rate. Furthermore, Note 11 of this Recommendation ventures into propagation advice when it states: 'It is desirable that systems be planned on the basis of propagation data covering a period of at least 4 years. The performance recommended to be met for any month should be based on the propagation data corresponding to the median worst month of the year taken from the monthly statistics of all the years for which reliable data are available'.

4.6.4.2 Propagation issues

As for the fixed service, the performance requirements of this example are expressed in terms of a percentage of any month. What is different in this case is that, instead of just long-term and short-term requirements, the performance objectives are described by a three-point distribution, with error-rate values for ten, two, and 0.03 per cent of any month. In this case the propagation models would need to be able to provide fading and/or interference data for these various time percentages. It is therefore important that, to be of maximum use to planning engineers now and in the future, propagation models can predict a propagation loss value for any time percentage likely to be needed. However, the realistic lower limit for average year statistics is about 0.0001 per cent (5.6 minutes a year), although confidence falls much below 0.001 per cent unless data from an exceptionally long measurement period is available.

4.6.5 Fixed satellite service availability objectives

4.6.5.1 ITU-R objectives

Recommendation ITU-R S.579-4 discusses the availability objectives for a hypothetical reference circuit (HRC) or hypothetical reference digital path (HRDP) when used in the FSS for PCM telephony, or as part of an ISDN.

Within this recommendation, Recommendation 3 states that: 'the unavailability due to propagation should not be more than:

- 0.2 per cent of any month for one direction of an HRDP in the FSS
- 0.1 per cent of any year . . . for one direction of an HRC in the FSS'.

4.6.5.2 Propagation issues

In this case the availability requirement for digital circuits (the HRDP) is expressed as a percentage of any month, and for the analogue case (the HRC) the requirement is based on any year. This is yet another descriptor of the baseline time which, if taken literally, would require the design engineer to address the most extreme years likely to be encountered over the planned lifetime of the system. However, pragmatism again comes into play, and for practical purposes it is taken to mean an average year.

4.6.6 Sharing between the fixed service and the fixed satellite service

As well as looking at the wanted signal performance and availability, it is useful to look at the requirements for interference between fixed service and fixed satellite service systems sharing the same frequency bands. Many of the most important telecommunications bands are shared in this way.

When the performance and availability objectives for individual services are translated into sharing criteria the situation becomes more complex. ITU-R Recommendations in the SF Series apply here. For example, Recommendation ITU-R SF.558-2 recommends maximum allowable values of interference from terrestrial radio links to satellite systems employing 8-bit PCM encoded telephony. To complement this, Recommendation ITU-R SF.615-1 gives maximum allowable values of interference from the fixed satellite service into terrestrial radio relay systems which may form part of an ISDN.

4.6.6.1 Fixed satellite service protection

For the protection of the fixed satellite service, Recommendation ITU-R SF.558 makes the following recommendations:

- the interfering power, averaged over any ten minutes, should not exceed, for more than 20 per cent of any month, 10 per cent of the total noise power at the input of the demodulator that would give rise to an error ratio of 1 in 10^6;

- the interfering RF power should not cause an increase of more than 0.03 per cent of any month during which the bit error ratio exceeds 1×10^{-4} averaged over one minute;
- the interfering RF power should not cause an increase of more than 0.005 per cent of any month during which the bit error ratio exceeds 1×10^{-3} averaged over one second.

These three protection criteria for the fixed satellite service protect both long-term and short-term error rate objectives, and all are clearly addressed in terms of any month.

4.6.6.2 Fixed service protection

For the protection of the fixed service, Recommendation ITU-R SF.615 provides three equivalent criteria:

- the interfering emissions should not degrade the performance by causing an increase of more than 0.0054 per cent of the period of time in any month during which the bit error ratio exceeds 1×10^{-3} (integration time one second);
- the interference emissions should not degrade the availability by causing an increase in the period of unavailability, as defined in Recommendation ITU-R F.557, of more than 0.03 per cent of any year;
- the interference emissions should not degrade the performance by causing an increase in the number of errored seconds measured at the 64 kbit s^{-1} interface by more than 0.032 per cent in any month.

In this case there are both worst-month performance and worst-month availability criteria to be met, together with an average year unavailability requirement.

There is an apparent conflict between these recommendations and those for the specific services discussed in Sections 4.6.3 and 4.6.4. The service-specific recommendations (e.g. Recommendations F.557 and F.695) imply that the availability objective should cover '. . . all causes that are statistically predictable . . .', normally deemed to include propagation, interference and hardware failures (including restoration time). However, the recommendations discussed in this section talk about not increasing the unavailability '. . . as defined in Recommendation ITU-R F.557 . . .' by more than a given amount due to interference arising from the sharing situation. This inconsistency is not helpful as it significantly affects the time percentages to be used for the interference prediction.

4.6.6.3 Propagation issues

For the protection of the fixed satellite service, Recommendation ITU-R SF.558 introduces an additional requirement to those encountered in the examples above, i.e. that of specifying an integration time of longer than one second (e.g. ten minutes and one minute) for the counting of errors due to interference.

To facilitate the evaluation of the interference environment for the satellite service the design engineer would ideally like to have available propagation data tailored to these integration times. Although interference propagation mechanisms like ducting tend to have a relatively slow rate of change, there are many others (e.g. layer reflections, hydrometeor scatter and tropospheric scatter) that can be far more peaky in their behaviour. In some cases averaging over a minute or ten minutes would make a significant difference compared to a one second integration time.

Unfortunately propagation models that can deal with different integration time requirements are not generally available, and it is difficult to see how such complex performance requirements can be realistically fulfilled.

4.7 Worst month

4.7.1 The concept of the worst month

The need to design a radiocommunications system to meet performance and availability objectives in any month is a tall order. Propagation conditions vary considerably from month to month, and the monthly variability can change significantly from year to year. Furthermore, it is not possible to be certain about what might happen in any monthly period for the next 20 years or more, as it could be more extreme than anything so far observed. In order to provide propagation data to help the designer who is faced with quality objectives based on any month, it was necessary to get away from this totally open-ended concept. To provide a solution the propagation community came up with the concept of the worst month, as defined in Recommendation ITU-R P.581-2.

The worst-month concept is only relevant to the higher-grade telecommunications services. It does not generally find applications in, for example, broadcasting, radio astronomy, space research, radio navigation or radio determination services, and arguably does not apply to terrestrial mobile or satellite mobile telecommunications services which generally have more relaxed performance and availability objectives than their fixed counterparts. In the context of interference assessment, worst month might only apply to some sharing scenarios and even then might only apply in one direction of interference propagation when only one of the sharing services required worst-month protection.

Worst-month propagation data has to be determined from extended measurement periods to ensure that a representative monthly variability is observed, particularly for the tails of the propagation distributions.

The propagation worst month concentrates on calendar months. Once several years of data are available, the data for the individual calendar months from each year (e.g. all the individual January data sets) can be averaged. The long-term averaged data for each of the 12 calendar months can then be plotted together and the envelope of this set of monthly distributions provides the worst-month

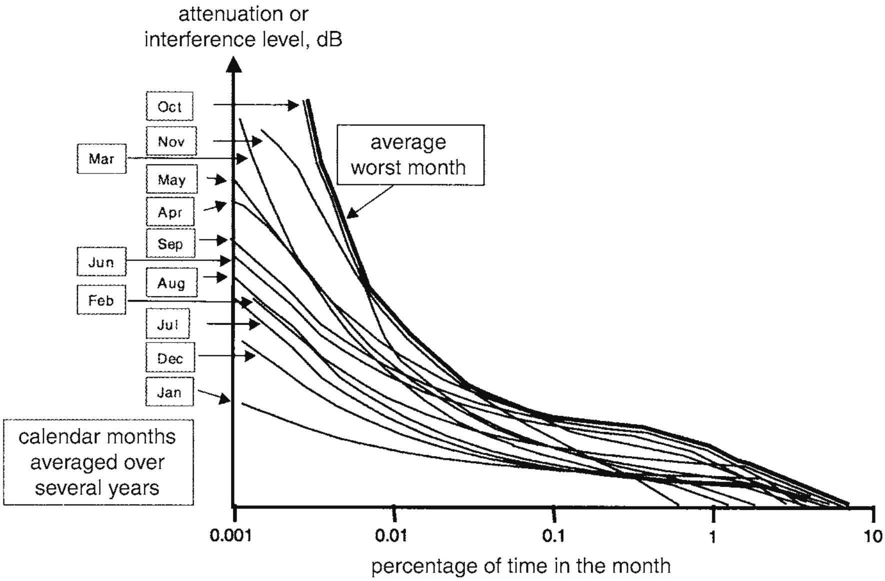


Figure 4.7 The determination of the average worst month from measured propagation statistics

distribution. Different points on the worst month curve can, and do, derive from different months. Figure 4.7 illustrates the approach.

What is found from this exercise is that the low-level events (shallow fades or low levels of interference) are distributed across several months, and for low thresholds the ratio of the percentage of time in the worst month to that in the average year is around 2 or 2.5. However, at the more extreme ends of the propagation distributions (the deepest fades or the highest levels of interference) the most significant events (perhaps only one or two a year) are concentrated into one or two months. The ratio worst month to annual time percentage then becomes 5 or 6. This ratio, generally known as Q, is important as it provides a way of predicting worst-month data from average annual data.

However, it should be noted that the values of Q vary from climate to climate and between the different propagation mechanisms. The ability to provide a reliable prediction for worst month therefore depends on having some experience of long term observations for the relevant propagation mechanism (rain fading).

4.7.2 Average worst-month statistics

Using the averaged calendar months from long-term measurements gives a generally reliable relationship between monthly and annual characteristics, but avoids the most extreme situations that would be technically unrealistic and uneconomic as design parameters. In order to gain an appreciation of the value

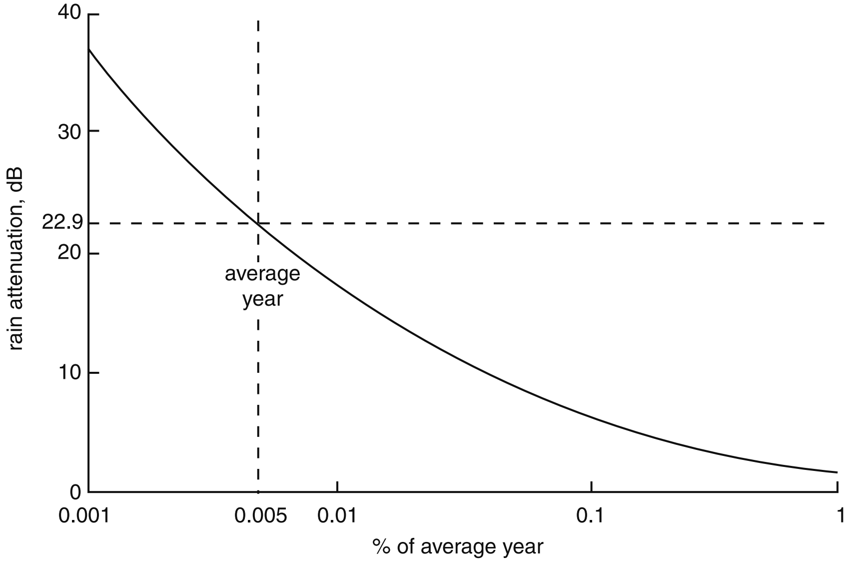


Figure 4.8 Average year rain attenuation prediction derived using Recommendation ITU-R P.530 (18 GHz, 20 km, vertical polarisation, rain climate E)

of the average worst month as a realistic solution to the any month requirement, it is useful to examine the relationship between annual and worst-month propagation statistics.

To illustrate the issues, rain attenuation data for a 20 km fixed service radio link operating in the UK at 18 GHz will be used. The 18 GHz band is heavily used in the UK and in several other countries of Europe for both high-capacity and local access applications in the fixed service. A wide range of performance and availability requirements are therefore encountered by engineers planning systems in this band.

Although rain attenuation is used as the example here, similar considerations would apply for the longer multipath-dominated links in the lower frequency bands below 10 GHz.

The average annual rain attenuation on the example 18 GHz link can be determined using the well established method set out in Recommendation ITU-R P.530-8. This method is covered elsewhere so it is not necessary to consider the detail here. However, Figure 4.8 shows a predicted cumulative distribution for the 18 GHz link for rain rate climate E (as defined in Recommendation ITU-R P.837-2) which applies to parts of the UK.

Recommendation ITU-R P.841 provides a method for the conversion of such average annual cumulative distributions to those that would be expected in the average worst month. For any average annual time percentage in the range 0.001 to 100 per cent, the method provides the ratio of the worst-month time percentage to the average annual time percentage. This ratio is given the identity Q, where $1 \leq Q \leq 12$. For rain fading in NW Europe, the method gives the values of Q shown in Figure 4.9.

These values of Q can then be used to derive worst-month equivalents of the average annual propagation predictions. Figure 4.10 shows worst month prediction derived from the data in Figure 4.8 using the method of Recommendation ITU-R P.841.

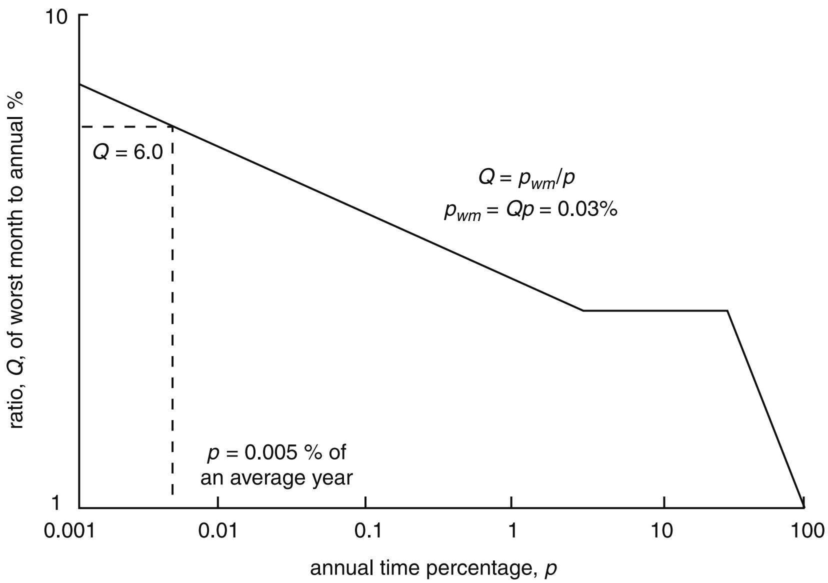


Figure 4.9 Dependence of Q on the average annual time percentage (NW Europe)

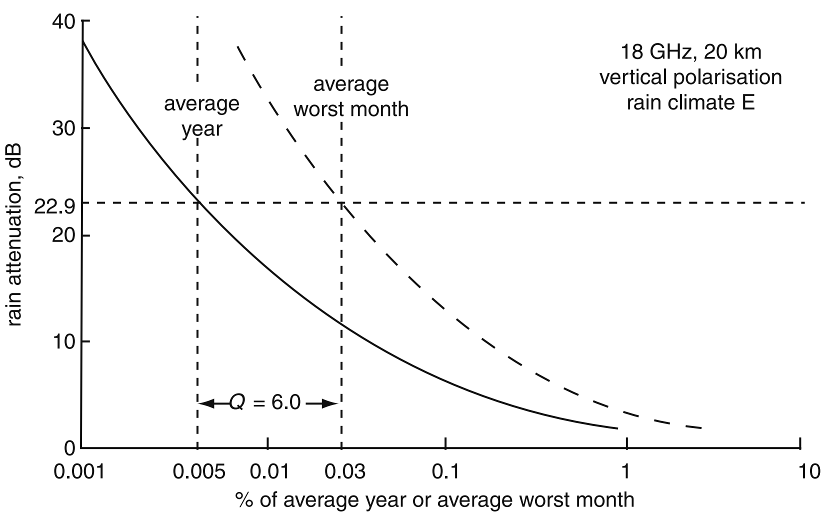


Figure 4.10 Worst-month prediction derived using the method of Recommendation 841

4.7.3 Worst months for other return periods

The worst-month data in Figure 4.10 represents an average worst month as defined by Recommendation ITU-R P.581–2. However, it is possible to derive statistics for the worst month within longer return periods. Recommendation ITU-R P.678–1 provides a method of extending these predictions to the worst months to be expected within 1.25, 2, 5, 10, 20 and 50 years.

Recommendation 678 provides, for a given value of Q, the ratio, W_R, of the worst-month time percentage for the specified return period to that of the average annual worst-month time percentage. From the recommendation, the values of this additional multiplier W_R can be found for the example link. These are shown in Figure 4.11.

If these additional corrections are then applied to the attenuation predictions for the 18 GHz link, it is possible to get a set of worst-month curves for various return periods (2, 5, 10 and 20 years are shown). Figure 4.12 shows the resulting curves.

As might be expected, the worst-month fading becomes more severe as the degree of averaging is reduced and the effects of the more extreme rainfall periods become evident. At, say, 0.03 per cent of the worst month, the difference in attenuation between the average worst month and that for the worst month for the 20-year return period is some 10 dB. It is hard enough just trying to cater for the average worst month, and to design a system to cater for these rare-event

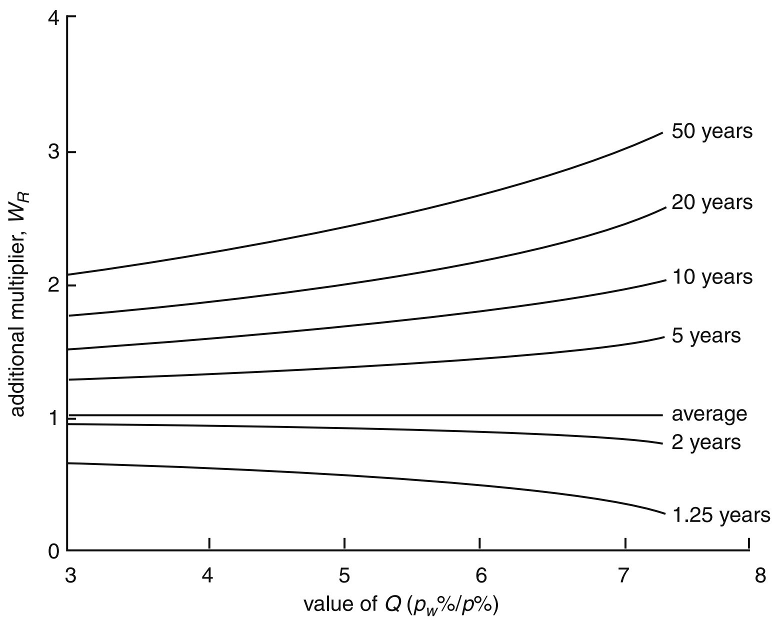


Figure 4.11 The additional correction necessary to predict worst-month attenuation statistics within a given return period

statistics is just not practical. It is also nearly impossible to collect statistically valid data for the longer return periods for all the propagation mechanisms and climates that are encountered.

It is for these reasons that propagation engineers adopted the compromise of the average worst month as the only viable solution to performance and availability criteria specified in terms of a percentage of time in any month.

4.8 Simple checks on the design

Using the average year and worst-month predictions of Figure 4.10, and the knowledge relating to the percentage of the total fading time that will probably be counted as unavailability, it is possible to make some simple checks on the progress of a radio hop design. Figure 4.12 shows one approach that could be used.

The starting point for the analysis is the availability allowance for the hop. Let it be assumed here that we are dealing with a medium grade 18 GHz route with four hops. Once the necessary apportionment between the causes of unavailability has been done, and a separate apportionment between the hops has been determined, we could envisage a working availability allowance for one hop of 0.005 per cent of an average year. This is equivalent to ~1580 seconds per year.

The fade depth for 0.005 per cent is 22.5 dB, as indicated by point A on Figure 4.12. Table 4.1 showed that the mean duration of such events at this fading threshold might be about ~230 seconds. Only about seven such events could therefore be accommodated per year.

Reference to Figure 4.9 shows that, for an average year time percentage of

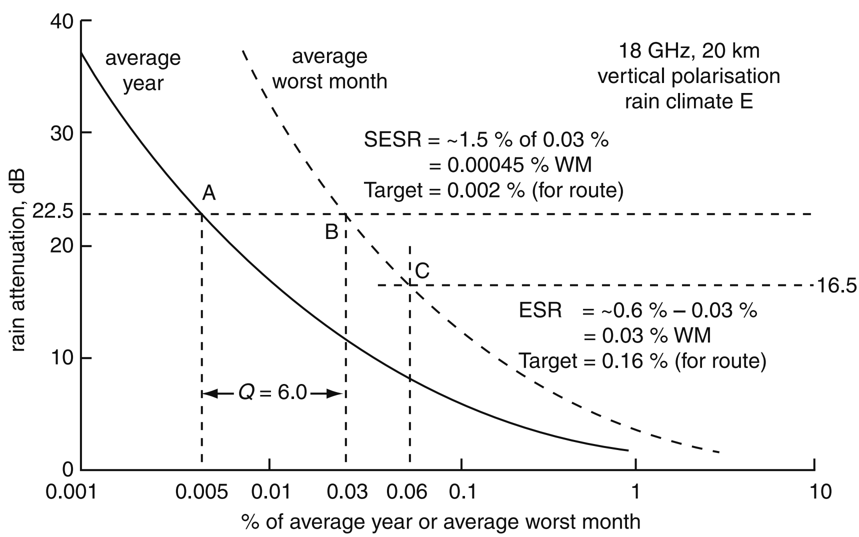


Figure 4.12 Simple checks on the design

0.005 per cent of time, we can expect a value of Q of 6, and hence an equivalent 0.03 per cent unavailability in the worst month. This is indicated by point B on Figure 4.12.

If a cautious approach to the ratio of SES error performance time to unavailable time (Figure 4.3) is taken, we might assume 97 per cent of the SES time is for events >10 s. This 97 per cent value for unavailable time is a conservative value, as most values in Table 4.1 are higher and would therefore give fewer SES to affect the performance. If we assume 50 per cent of the remaining SES time for SES within the error performance allowance, this gives an SESR of $1.5 \times 0.03 = 0.00045$ per cent of the worst month. For the whole medium-grade route the target might be 0.002 per cent, and this might need to include an allowance for interference. The per-hop figure of 0.00045 per cent SESR might therefore be a little too high for a four-hop route. All possible trade-offs would need to be examined, but if no other solution were available it might be necessary to design to a tighter unavailability figure than 0.005 per cent (say 0.003 per cent) in order to have more confidence that the SES element of the error performance would be achieved. The design would then need to be repeated based on this new availability target. Shorter hops, larger antennas or a lower frequency band would be some of the options available.

In addition to the SES element of the error performance, there is often an ES target to be achieved as well. In Figure 4.2 we saw that the ES threshold might only be about 2–3 dB below the SES threshold. For a conservative design a figure of 6 dB or more might be used. A 6 dB ES threshold is shown in Figure 4.12. The ES ratio is also a worst-month parameter, and the 16.5 dB threshold intersects the worst month at 0.06 per cent of the worst month (point C). Errored seconds would therefore occur between 0.03 and 0.06 per cent of the month, a total of 0.03 per cent. A typical target for ES might be 0.16 per cent of the worst month, so there is probably ample scope to accommodate four-hops in this context, even with some provision for interference.

The critical case is clearly the SES contribution to the error performance during available time. It was useful for this example that the data given in Table 4.1 was available. Without such data it would be very hard to build confidence in the design. It is for these reasons that nearly all major operators of radio links invest money in locally-derived propagation data. Such data gives the all important confidence that the radio network will deliver the performance promised to its customers. It is also fortunate that operators are generally willing to share their data and propagation models (through the ITU or regional or national organisations) to the general benefit of all players.

So much is still needed by way of propagation data to support digital radio network design. It is only through the co-operative environment of ITU-R that we can hope to secure the necessary data and propagation models within realistic timescales.

Chapter 5

Electromagnetic wave propagation

M.J. Mehler

5.1 Introduction

The objective of this Chapter is to introduce the principles of electromagnetic theory which are essential to an understanding of radiowave propagation. Where possible a practical engineering approach has been adopted. However, the material has been structured to provide those with a more mathematical background with a satisfactory account of the subject, together with references to further reading material.

Initially, propagation between two antennas in free space is considered by means of power flux concepts. This enables important ideas such as free-space path loss, antenna gain and effective aperture to be introduced. Maxwell's equations are presented. and plane-wave solutions are derived as a means of introducing polarisation, wave impedance and electromagnetic power flow. Finally, radiation from a current distribution is examined and illustrated by deriving the fields of a dipole antenna.

5.2 Power budget

Radio propagation engineers are often concerned with determining the power received by a distant antenna when the transmitted power is given. Such a calculation is the key to the design of radio systems and involves the determination of the power budget taking account of each element of the radio link. In practice, when both transmitter and receiver are in free space, this calculation can easily be made using a simple conservation-of-energy argument. Moreover, even when the antennas are not in free space the concept of a link budget, based on energy conservation, can be extended to take account of effects such as diffraction loss.

5.2.1 Transmitting antennas

Initially, consider a point source in free space radiating energy equally in all directions. If such an isotropic source radiates a total of P_t (W), then the power density F_0 at a distance d will be given by:

$$F_0 = \frac{P_t}{4\pi d^2} \tag{5.1}$$

Note that such a source could not be physically realised since the presence of polarisation cannot be neglected. However, this concept is useful for specifying a reference against which antenna gain may be defined. In practice an antenna will possess a radiation pattern, as illustrated in Figure 5.1.

If an antenna is fed with a power P_t, the power density $F_d(\theta,\phi)$ may be specified at a distance d in a given direction θ, ϕ. The gain of the antenna in the θ, ϕ direction may be defined with respect to an isotropic power density by taking the ratio of $F_d(\theta, \phi)$ with the flux density produced by an isotropic radiator at the same distance and fed with the same power P_t. Thus, the antenna gain is given by

$$G(\theta,\phi) = \frac{F_d(\theta,\phi)}{F_0} \tag{5.2}$$

The nature of the gain function $G(\theta,\phi)$ will depend on the type of antenna. Normally, antennas are used to concentrate the power flux in some specific direction and along this boresight direction the antenna gain will be a maximum. Typically, wire antennas of the Yagi type operating in the UHF bands yield gains of approximately 15 dBi, whereas large microwave earth-station

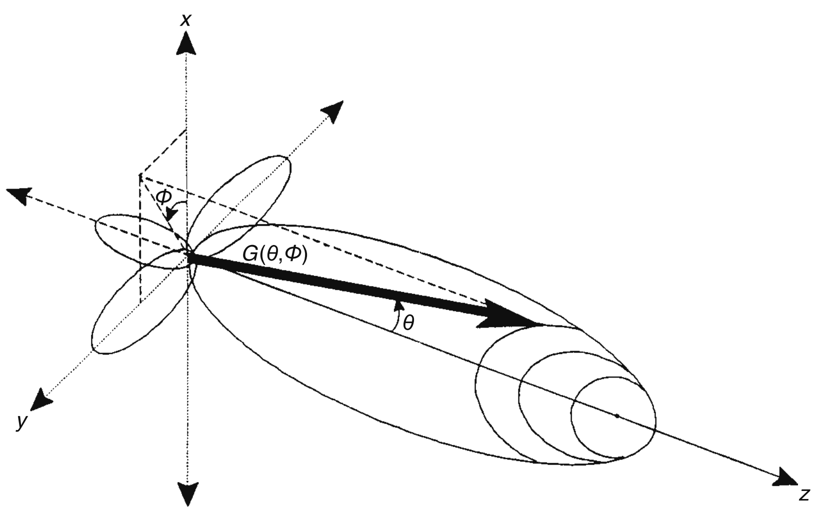


Figure 5.1 An antenna radiation pattern

antennas of about 3 m diameter can produce gains in the order of 50 dBi. Here the i denotes the gain with respect to that of an isotropic antenna.

5.2.2 Receiving antennas

The power density at a distance d in a given direction from a transmitting antenna is $F_d(\theta,\phi)$. A useful expression for this quantity can be found by combining Eqns 5.1 and 5.2, to yield:

$$F_d(\theta,\phi) = \frac{P_t G_t(\theta,\phi)}{4\pi d^2} \tag{5.3}$$

where G_t, is the gain of the transmitting antenna. A receiving antenna placed in this power flux will deliver a power P_r into a load connected to the antenna terminals or waveguide port. To facilitate the calculation of P_r we may consider the antenna to be presenting an effective collecting-aperture area A_e, to the incoming field. Hence, the received power becomes:

$$P_r = \frac{P_t G_t(\theta,\phi) A_e}{4\pi d^2} \tag{5.4}$$

The concept of an effective aperture area is easily interpreted for reflector antennas, where a paraboloid of diameter D presents a projected area $\pi D^2/4$ to a normally incident wave. Even if ohmic losses in the antenna are neglected, some of the incident field will be reflected and hence not all of the incident energy will be absorbed in the antenna load. If the paraboloid is rotated, the incident wave will no longer be normal to the aperture. Hence, additional incident energy will be reflected and thus the received power in the load will be further reduced in a manner consistent with the antenna pattern. These considerations illustrate that the effective aperture area of an antenna is not necessarily equal to the physical area $A = \pi D^2/4$. Moreover, the effective area will depend on the direction of the incident field. It is possible to relate the effective aperture to the antenna gain through the relation:

$$A_e(\theta,\phi) = G(\theta,\phi)\,\frac{\lambda^2}{4\pi} \tag{5.5}$$

The ratio between the effective aperture and the physical area, in the direction of maximum gain, is a measure of antenna efficiency η, so that $\eta = A_e/A$. Combining this definition with Eqn. 5.5 yields a useful expression for the maximum antenna gain, namely:

$$G = \left(\frac{\pi D}{\lambda}\right)^2 \eta \tag{5.6}$$

Although the above discussion has been concerned with reflector antennas, the concept of effective aperture area is also commonly applied to wire antennas.

5.2.3 Free-space transmission loss

Combining Eqn. 5.4 with Eqn. 5.5 permits the received power to be written in terms of the transmitting and receiving antenna gains G_t, and G_r, respectively. The resulting expression takes the form:

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi D}\right)^2 \tag{5.7}$$

To interpret this result, note that if both transmitting and receiving antennas possess a gain of unity, then in free space the ratio P_r/P_t depends on $(\lambda/4\pi D)^2$, which is simply the free-space loss.

5.2.4 Link power budget

Radio engineers usually express Eqn. 5.7 in decibels, so that we may write:

$$P_r(\text{dBw}) = P_t(\text{dBw}) + G_t(\text{dBi}) + G_r(\text{dBi}) - L_{bf}(\text{dB}) \tag{5.8}$$

where:

$$L_{bf} = 20\{\log_{10}(4\pi d) - \log_{10}(\lambda)\}$$

noting that $\lambda = c/f$ (c is the velocity of light in a vacuum = 3×10^8 ms^{-1}) and taking f to be in MHz and d in km, the following standard result is obtained:

$$L_{bf} = 32.44 + 20\log_{10} f + 20\log_{10} d \tag{5.9}$$

(Note that dBw is with respect to 1 W; dBi is the gain with respect to an isotropic antenna.)

Eqn. 5.8 can be extended to treat propagation between antennas which are not in free space by making allowance for additional sources of path loss. For example, propagation over a path with terrain features introduces loss in addition to that of free space. This additional diffraction loss, L_d, can easily be included into the link budget, so that Eqn. 5.8 becomes

$$P_r = P_t + G_t + G_r - L_{bf} - L_d \tag{5.10}$$

Clearly, this approach can be extended to include other loss mechanisms, such as antenna feeder losses. Eqn. 5.10 forms the basis of many radio-system-link budget calculations.

5.3 Basic electromagnetic theory

The above discussion has only been concerned with the use of conservation-of-energy arguments where the electromagnetic field has been treated as a scalar power flux. In reality this model is inadequate and better agreement with experimental observation is achieved by treating electromagnetic effects in terms of the interaction of charged particles with a vector field. For engineering applications it is often sufficient to describe these electromagnetic phenomena by means of a classical field theory, namely Maxwell's equations. These equations relate the electric field $\boldsymbol{E}$ (V m^{-1}) and the magnetic field $\boldsymbol{H}$ (A m^{-1}) to the current density $\boldsymbol{J}$ (A m^{-2}) and the charge density ρ (C m^{-3}).

They may be written in the form:

$$\nabla \times \boldsymbol{E} = -\mu \frac{\partial \boldsymbol{H}}{\partial t} \tag{5.11a}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \varepsilon \frac{\partial \boldsymbol{E}}{\partial t} \tag{5.11b}$$

$$\nabla.(\varepsilon \boldsymbol{E}) = \rho \tag{5.11c}$$

$$\nabla.(\mu \boldsymbol{H}) = 0 \tag{5.11d}$$

where ε and μ are the permittivity and permeability of the medium. They are often written as $\varepsilon = \varepsilon_0\ \varepsilon_r$, and $\mu = \mu_0\ \mu_r$, *where* $\mu_r = \varepsilon_r = 1$ in free space, and $\varepsilon_0 = 8.854 \times 10^{12}$ F m^{-1}, $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1}.

Significant simplification of Maxwell's equations can be achieved in free space where both $\boldsymbol{J}$ and ρ are zero. Taking the curl of Eqn. 5.11*a* and using Eqn. 5.11*b* to eliminate $\boldsymbol{H}$ gives:

$$\nabla \times \nabla \times \boldsymbol{E} = -\mu_0 \varepsilon_0 \frac{\partial^2 \boldsymbol{E}}{\partial t^2} \tag{5.12}$$

This may be further simplified by means of a standard vector identity to give

$$-\nabla^2 \boldsymbol{E} + \nabla(\nabla.\boldsymbol{E}) = -\mu_0 \varepsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} \tag{5.13}$$

Since in Eqn. 5.11*c* $\rho = 0$ and $\varepsilon = \varepsilon_0$ (constant), Eqn. 5.13 may be further reduced to:

$$\nabla^2 \boldsymbol{E} - \mu_0 \varepsilon_0 \frac{\partial^2 \boldsymbol{E}}{\partial t^2} = 0 \tag{5.14}$$

This is the well known free-space wave equation for $\boldsymbol{E}$.

5.3.1 Plane-wave solutions

In general, solutions to Eqn. 5.14 are difficult to obtain for many important engineering problems. However, a number of very useful properties of the electromagnetic field can be deduced by considering simple plane-wave solutions. The equation for a plane wave which is travelling in the positive $\boldsymbol{z}$ direction, with polarisation in the $\boldsymbol{x}$ direction, may be written as:

$$\boldsymbol{E} = E_0 \cos(\omega t - kz)\boldsymbol{x} \tag{5.15}$$

where E_0 is the amplitude of the wave, ω is the angular frequency $2\pi f$ and k the wave number $2\pi/\lambda$. Here, f (Hz) and λ (m) are the frequency and wavelength, respectively.

If we select a point along the wave of arbitrary constant phase, satisfying $\omega t - kz =$ constant, then the velocity c of this point on the wave is $dz/dt = \omega/k = \lambda f$. It is easy to check that Eqn. 5.15 is a solution of Eqn. 5.14 provided that:

$$k^2 = \omega^2 \mu_0 \varepsilon_0 \tag{5.16}$$

Therefore, the wave velocity is given by:

$$C = \frac{\omega}{k} = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$$

and it may be verified that $c = 3 \times 10^8$ m s^{-1} in free space.

To find the magnetic field corresponding to this plane-wave solution we substitute Eqn. 5.15 into Eqn. 5.11a to obtain:

$$-\mu_0 \frac{\partial \boldsymbol{H}}{\partial t} = \nabla \times \mathbf{E} = \frac{\partial}{\partial z} \{E_0 \cos(\omega t - kz)\} \hat{\boldsymbol{y}}$$

$$-\mu_0 \boldsymbol{H} = \hat{\boldsymbol{y}}.E_0 k \sin(\omega t - kz) dt$$

so that

$$\boldsymbol{H} = H_0(\omega t - kz)\hat{\boldsymbol{y}} \tag{5.17}$$

where

$$H_0 = \frac{E_0 k}{\omega \mu_0}$$

It can be seen that, for this plane-wave solution, the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields are orthogonal, as illustrated in Figure 5.2.

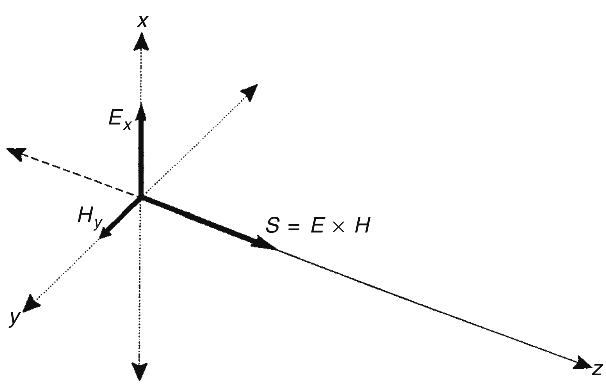


Figure 5.2 Relationship between the E and H fields

5.3.2 Wave impedance

It is evident that the ratio of the $\boldsymbol{E}$- and $\boldsymbol{H}$-field amplitudes in Eqn. 5.15 and Eqn. 5.17 is a constant, namely:

$$\frac{|\boldsymbol{E}|}{|\boldsymbol{H}|} = \frac{E_x}{H_y} = \frac{\omega\mu_0}{k} = \sqrt{\frac{\mu_0}{\varepsilon_0}} \tag{5.18}$$

This ratio is called the free-space wave impedance since it has units of ohms, and is usually denoted by Z_0, which for this free-space case has an approximate value of 377Ω. It is analogous to the ratio of voltage and current in an electric circuit.

5.3.3 Power flow and Poynting's theorem

In the same way that the product of voltage and current may be used to calculate power in an electric circuit, a vector product of $\boldsymbol{E}$ and $\boldsymbol{H}$ gives the power density in the electromagnetic field. For the plane-wave solution, the power flow along the Z direction is $(|\boldsymbol{E}|\,|\boldsymbol{H}|)/2$ W m^{-2}. Here the factor 1/2 in this expression has been introduced because E_0 and H_0 are peak values, whereas the power density is in root mean square units. Using Eqns. 5.15 and 5.17, this expression reduces to:

$$\frac{E_0^2}{2z_0} = \frac{Z_0}{2} H_0^2$$

In general the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields may not be orthogonal and in this case the above result is not valid. However, it is possible to show that the power flow in an electromagnetic field may be represented by the Poynting vector $\boldsymbol{S}$ [122], which points in the direction of power flow and has a magnitude equal to the power density, and is defined by:

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} \tag{5.19}$$

5.3.4 Exponential notation

Plane-wave solutions occur commonly in engineering problems. Indeed, it is possible to represent arbitrary field distributions by a sum of plane waves of varying amplitude, phase and direction. It is also possible to simplify the manipulation of plane-wave expressions by introducing exponential notation, namely:

$$\mathrm{Re}[\exp\{j(\omega t - kz)\}] = \cos(\omega t - kz) \tag{5.20}$$

Hence, Eqn. 5.15 is written as:

$$\boldsymbol{E} = E_0\,\mathrm{Re}\{\exp(j\omega t)\exp(-jkz)\}\hat{\boldsymbol{x}} \tag{5.21}$$

The $\exp(j\omega t)$ term is present because the field varies harmonically with time. Such steady-state solutions arise commonly in radio engineering problems. In Maxwell's equations the adoption of the harmonic time dependence of Eqn.

5.21 implies that differentiation with respect to time is replaced by $j\omega$, so that Eqns. 5.11*a–d* reduce to:

$$\nabla \times \boldsymbol{E} = -j\omega\mu\boldsymbol{H} \tag{5.22a}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + j\omega\varepsilon\boldsymbol{E} \tag{5.22b}$$

$$\nabla . \varepsilon\boldsymbol{E} = \rho \tag{5.22c}$$

$$\nabla . \mu\boldsymbol{H} = 0 \tag{5.22d}$$

The wave equation corresponding to Eqn. 5.14 follows by taking the curl of Eqn. 5.22*a* and using Eqn. 5.22*b* to eliminate $\boldsymbol{H}$, and is given by:

$$\nabla^2 \boldsymbol{E} + k^2 \boldsymbol{E} = 0 \tag{5.23}$$

It is easily shown that Eqn. 5.21 is a solution of Eqn. 5.23. Thus, using this shorthand notation, Eqn. 5.21 becomes:

$$\boldsymbol{E} = E_0 \exp(-jkz)\hat{\boldsymbol{x}} \tag{5.24}$$

When interpreting this equation, or any solution of Eqn. 5.23, the need to multiply by $\exp(-jkz)$ and take the real part should be borne in mind.

5.4 Radiation from current distributions

So far the plane-wave solutions considered have been obtained without any reference to the current distribution which excites them. This is manifest by the arbitrary nature of E_0 in Eqn. 5.15. In reality, E_0 would be determined by the source currents. More generally, if the field were represented by a superposition of plane waves, known as a plane-wave spectrum, the amplitude, phase and direction of each plane wave would be determined by the source-current distribution.

When calculating the field radiated by a current distribution, such as that supported on an antenna structure, it is often more convenient to work with a vector potential, namely $\boldsymbol{A}$. This potential may be introduced through the standard result $\nabla . \nabla \times \boldsymbol{A} = 0$, and it may be deduced from Eqn. 5.11*d* that:

$$\mu\boldsymbol{H} = \nabla \times \boldsymbol{A} \tag{5.25}$$

Substituting Eqn. 5.25 into Eqns 5.11*a–d* yields a wave equation for $\boldsymbol{A}$ of the form:

$$\nabla^2 \boldsymbol{A} + k^2 \boldsymbol{A} = -\mu_0\boldsymbol{J} \tag{5.26}$$

If the current distribution is bounded by a volume v_0, the field $\boldsymbol{A}$ in the free-space region outside the volume v_0 can be written explicitly in the form:

$$\boldsymbol{A(r)} = \frac{\mu}{4\pi}\int \boldsymbol{J}(\boldsymbol{r}_0)\frac{\exp(-jk|\boldsymbol{r}-\boldsymbol{r}_0|)}{|\boldsymbol{r}-\boldsymbol{r}_0|}dv_0 \tag{5.27}$$

Eqn. 5.27 is therefore a general solution to Eqn. 5.26 for any current distribution and, as such, is a very useful result. The integration in Eqn. 5.27 is taken over the source co-ordinates which are denoted by the subscript 0. The details of the derivation of Eqns. 5.26 and 5.27 are beyond the scope of this Chapter, but may be found in Reference [1]. The field quantities $\boldsymbol{E}$ and $\boldsymbol{H}$ can be derived once $\boldsymbol{A}$ is known, first by application of Eqn. 5.25 to find $\boldsymbol{H}$ and then, from Eqn. 5.22*b*, $\boldsymbol{E}$ is deduced, noting that in free space $\boldsymbol{J} = 0$.

5.4.1 Radiation from a short current element

Consider the short current element of length dl as shown in Figure 5.3, which is assumed to carry a constant current I. Since the element is z directed, we may write down the expression for $\boldsymbol{A}$ in the form:

$$A_z = \frac{\mu_0}{4\pi} \int_{-dl/2}^{+dl/2} I \frac{\exp(-jkr)}{r} dz, \qquad A_x = A_y = 0 \tag{5.28}$$

Eqn 5.28 follows directly from Eqn. 5.27 by using the approximation $dl \ll r$, which implies that the observation distance is much greater than the length of the current element. The integral in Eqn. 5.28 is easily evaluated to yield:

$$A_z = \frac{\mu_0}{4\pi} I.dl \frac{\exp(-jkr)}{r} \tag{5.29}$$

The fields of antennas are usually presented in spherical co-ordinates; hence we convert Eqn. 5.29 as follows:

$$\boldsymbol{A} = \frac{\mu_0 Idl}{4\pi} \frac{\exp(-jkr)}{r} (\hat{\boldsymbol{r}} \cos\theta - \hat{\theta} \sin\theta)$$

To find $\boldsymbol{H}$, Eqn. 5.25 is applied to give:

$$\boldsymbol{H} = \frac{1}{\mu_0} \nabla \times \boldsymbol{A} = \frac{Idl \sin\theta}{4\pi} \left(\frac{jk}{r} + \frac{1}{r^2} \right) \exp(-jkr)\hat{\phi} \tag{5.30}$$

Eqn. 5.30 shows that the magnetic field is arranged in loops around the z axis of the current element. Furthermore, the field appears to fall off with distance at two rates, as controlled by the terms $(1/r)$ and $(1/r^2)$. A more complex expression results when Eqn. 5.22*b* is used to calculate $\boldsymbol{E}$, namely:

$$\boldsymbol{E} = \frac{jz_0 Idl \cos\theta}{2\pi k_0} \left(\frac{jk}{r^2} + \frac{1}{r^3} \right) \exp(-jkr)\hat{\boldsymbol{r}} - \frac{jz_0 Idl \sin\theta}{4\pi k_0} \left(-\frac{k_0^2}{r} + \frac{jk}{r^2} + \frac{1}{r^3} \right) \exp(-jkr)\hat{\theta} \tag{5.31}$$

It is evident that the $\boldsymbol{E}$ field possesses both transverse and longitudinal components, together with a distance dependence which extends up to $(1/r^3)$ terms.

When r is large relative to the wavelength, the only important terms are those that vary as $(1/r)$. In this radiation, or far-field, zone the $\boldsymbol{E}$ and $\boldsymbol{H}$ fields reduce to:

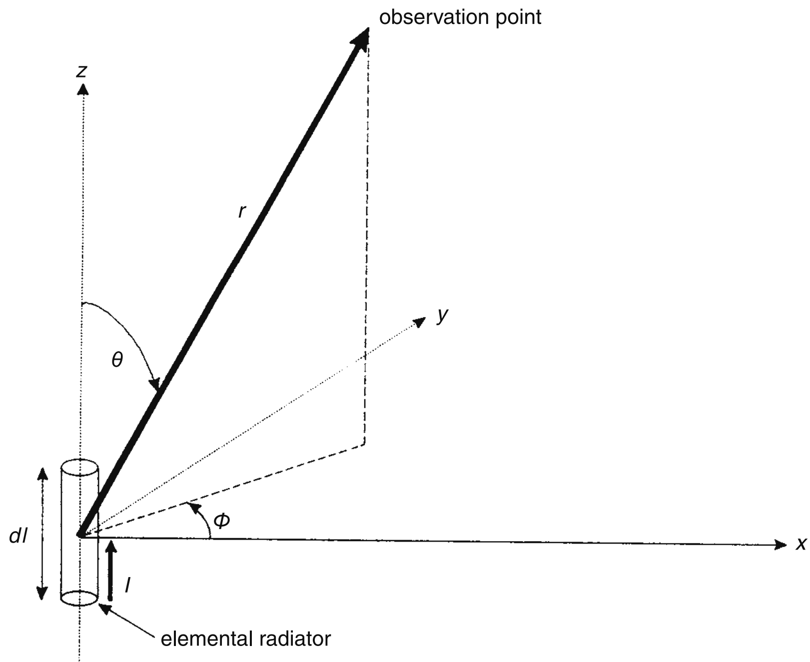


Figure 5.3 Radiation from a short current element

$$\boldsymbol{E} = jZ_0 Idlk_0 \sin\theta \frac{\exp(-jkr)}{4\pi r}\hat{\theta} \tag{5.32a}$$

$$\boldsymbol{H} = jIdlk_0 \sin\theta \frac{\exp(-jkr)}{4\pi r}\hat{\phi} \tag{5.32b}$$

When $r < \lambda$, the terms $(1/r^2)$ are important and cannot be neglected in the calculation. These near-field components are reactive and equate to the storage of magnetic- and electric-field energy. Hence the antenna may be visualised as being surrounded by reactive stored energy for $r < \lambda$, with radiation fields which become more dominant as r increases. As would be expected, the radiation field possesses a Poynting vector which points radially outward from the current element. This is evident if one calculates the radiated power $\boldsymbol{S}_{\mathrm{r}}$ (RMS) from Poynting's theorem applied to Eqn. 5.32, namely:

$$\boldsymbol{S} = \frac{\boldsymbol{E} \times \boldsymbol{H}}{2} = \frac{|\boldsymbol{E}|\,|\boldsymbol{H}|}{2}(\hat{\theta} \times \hat{\phi}) = \frac{|\boldsymbol{E}|\,|\boldsymbol{H}|}{2}\hat{\boldsymbol{r}} \tag{5.33}$$

$$\boldsymbol{S}_{\mathrm{r}} = \frac{I^2 dl^2 Z_0 k^2 \sin^2\theta}{32\pi^2 r^2}\hat{\boldsymbol{r}} \tag{5.34}$$

This result shows that the power falls off as an inverse-square law and flows radially away from the current element, as expected.

To calculate the total radiated power P_r it is necessary to integrate Eqn. 5.34 over a complete sphere enclosing the current element, which yields the surface integral:

$$P_r = \frac{I^2 dl^2 Z_0 k^2}{32\pi^2} \int_0^{2\pi} \int_0^{\pi} \frac{\sin^2\theta}{r^2} (r^2 \sin\theta) d\theta d\phi$$

Integration over ϕ is straightforward, but to facilitate the integration over θ the substitution $u = \cos\theta$ is made, to yield:

$$P_r = \frac{I^2 dl^2 Z_0 k^2}{16\pi} \int_{-1}^{+1} (1-u^2) du$$

which reduces to

$$P_r = \frac{I^2 dl^2 Z_0 k^2}{12\pi} \tag{5.35}$$

5.4.2 Radiation resistance

If a source is connected to the antenna terminals and supplies a given power P_R, we may regard the power as being supplied to an effective radiation resistance R_{rad}. Although this is not a physical resistance, it absorbs a power equal to that being radiated in to the far field. Since I is a peak value, equating the power supplied to R_{rad} to that which is radiated gives:

$$\frac{1}{2} I^2 R_{rad} = \frac{I^2 \, dl^2 Z_0 k^2}{12\pi}$$

Therefore

$$R_{rad} = \frac{Z_0 (k dl)^2}{6\pi} = 80\pi^2 \left(\frac{dl}{\lambda}\right)^2 \tag{5.36}$$

As an example, at a frequency of 1 MHz (λ = 300 m) and with an antenna of length dl = 1 m, R_{rad} = 0.0084 Ω. In practice it would be difficult to match an antenna with such a small radiation resistance to a source and achieve efficient power transfer. As a result, practical wire antennas are usually of the order of one wavelength in length.

5.4.3 The halfwave dipole

The starting point for calculating the fields radiated by the short current element was an assumed current distribution. Since the short current element is assumed to be small compared with the wavelength, taking the current to be uniform along the element is a good approximation. However, for the halfwave dipole a physically realisable distribution would take the form of a cosinusoidal distribution falling to zero at the wire ends.

Such a distribution can be justified as a good approximation theoretically and may be written as:

$$I = I_0 \cos\left(\frac{2\pi}{\lambda} Z\right) - \quad \frac{\lambda}{4} \leq Z \leq \frac{\lambda}{4} \tag{5.37}$$

Again, the antenna is assumed to be located along the z axis, as illustrated in Figure 5.3.

Although the algebra is more complex with the current distribution of Eqn. 5.37, a calculation analogous to that used for the short current element may be repeated to derive the halfwave dipole fields. For brevity, the results are merely stated here and in the far field are given by [75]:

$$\boldsymbol{E} = \frac{jI_0 Z_0}{2\pi r} \exp(-jkr) \frac{\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta} \hat{\theta} \tag{5.38a}$$

$$\boldsymbol{H} = \frac{jI_0}{2\pi r} \exp(-jkr) \frac{\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta} \hat{\phi} \tag{5.38b}$$

The radiation resistance of the halfwave dipole can be calculated by following the approach adopted for the short current element, and may be shown to be approximately 73 Ω.

5.5 Reference

1 KRAUS, and CARRER: ‘Electromagnetics’, International Students edition (McGraw Hill, New York, 1973, 2nd Edn.)

Chapter 6

Reflection and scattering from rough surfaces

David Bacon

6.1 Introduction

Radiowaves are reflected by the ground and from objects such as buildings. This can have various effects on radio systems. Reflections from buildings, etc., can permit a radio service to exist where the signal would otherwise be excessively attenuated by shadowing. Conversely, reflections can cause interference where shadowing alone would provide adequate attenuation of an unwanted signal. Reflections are also a major cause of multipath propagation. This can sometimes be exploited, particularly in a mobile radio system using code division multiple access (CDMA). For point-to-point links, on the other hand, ground reflections are generally viewed as an impairment, and every effort is made to minimise their effect.

6.2 Reflection from a plane surface

Surface reflection occurs at an abrupt change in the electrical properties of the medium through which electromagnetic radiation passes. The general situation for a plane smooth surface is shown in Figure 6.1.

The line A–A′ is the interface between two media of different electrical properties. Radio energy in medium 1 meets this interface at incidence angle θ_i. Due to the change in wave impedence at the interface, part of the incident energy is reflected, and by geometrical optics this will be at angle θ_r, equal to θ_i. The remainder will penetrate medium 2 at angle θ_p which in general will not equal θ_i. The change in direction of the penetrating ray is referred to as refraction, and the penetrating ray is often referred as the refracted ray. In radiowave propagation studies one of the two media is nearly always air.

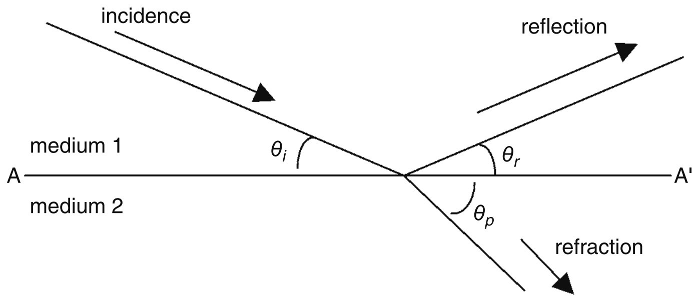


Figure 6.1 Reflection and refraction at a plane boundary between two media

6.2.1 The complex reflection coefficient

The reflection coefficient of a surface can be viewed as the transfer function of the reflection process. Since both amplitude and phase changes are possible at the point of reflection, the coefficient is in general complex. Thus if E_i and E_r are the complex values of the incident and reflected phasors immediately before and after the reflection point, we may write:

$$E_r = \rho\, E_i \tag{6.1}$$

where ρ = the complex reflection coefficient.

A reflection coefficient is a function of the electrical properties of the two media, the angle of incidence and of the frequency and polarisation of the incident signal.

6.2.2 Definition of reflection angles

The angles shown in Figure 6.1 are measured from the plane interface between the media. This is often used for ground reflection since it is more convenient to measure ray elevation angles relative to the horizontal, and such angles are often small. However, reflection angles are sometimes measured to the normal to the interface, which is also the convention used in optics. Thus in radiowave propagation calculations care should be taken to ensure that the correct definition of angles is being used. In most cases changing from one to the other only requires exchanging sines and cosines in the relevant equations.

6.2.3 Designation of polarisation

Reflection theory normally accounts for two orthogonal linear polarisations. For other polarisation angles, and for circular polarisation, the incident wave is resolved into these orthogonal linear components, and then recombined after reflection.

If Figure 6.1 is interpreted in terms of reflection from plane ground, the incident ray, if linearly polarised, will probably be either horizontal or vertical.

In the case of horizontal polarisation the electric vectors will be normal to the paper; in the case of vertical polarisation the electric vectors will be parallel to the paper.

However, the designations horizontal and vertical are less useful in the case of reflection from the vertical side of a building, and even less so for a sloping roof. Thus for the general case of reflection in a three-dimensional world the terms perpendicular and parallel are normally used. These refer to the relationship of the incident electric vectors with the plane containing the incident and reflected rays. Thus in the case of ground reflection, horizontal polarisation can also be described as perpendicular, and vertical as parallel.

Both terminologies are used in this book. When discussing ground reflection it is often more convenient to refer to horizontal and vertical polarisation, but perpendicular and parallel are more exact descriptions, and should be used where there can be any doubt as to the orientation of the reflecting surface.

It can be noted that in a vertically-polarised wave the electric vectors are only approximately vertical; they actually lean forwards or backwards according to the elevation angle of the ray.

6.3 Reflection by perfectly-conducting surfaces

A perfectly-conducting surface is assumed to reflect perfectly, that is, there is no penetration and all of the incident energy is reradiated in the reflected wave. This is based on the principle that the tangential electric vector at a perfectly conducting surface must be zero. This is an idealisation which can only be approximated in reality, but is nevertheless a useful notion.

Before describing perfect reflection for the two incident polarisations it will be useful to introduce the concept of reflection images.

6.3.1 Theory of images

An image is a hypothetical source of radiation which can be useful in practical calculations concerning reflections by plane surfaces.

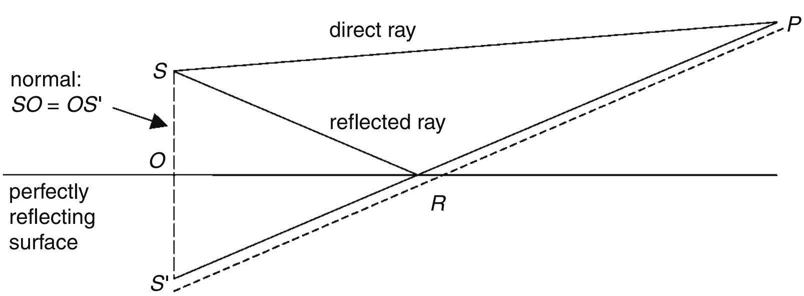


Figure 6.2 Reflection image, and hypothetical image ray

In Figure 6.2 the source S launches direct and reflected rays to a field point P. The image S' of the source lies on the normal to the surface through S, and is equidistant to S on the far side of the surface.

The basic principle of the image is that a hypothetical direct ray from S' to P, shown dashed in Figure 6.2, replaces the reflected ray and the surface. Note that the hypothetical ray $S'P$ has the same length as the total reflected path $SR + RP$.

This might appear to contravene conservation of energy, since the field point can now see two sources of radiation. However, the principle applies only to the half space on the S side of the reflecting surface.

6.3.2 Perfect reflection for perpendicular polarisation

Figure 6.3*a* illustrates reflection from a perfectly-conducting surface for perpendicular polarisation.

The symbol ⊕ at reflection point R represents horizontal polarisation with the electric field pointing into the plane of the paper at a particular instance. Since the tangential electric field is zero, the image ray must produce an equi-amplitude opposite-phase field at the same point. This condition is satisfied by an image source having the same amplitude but opposite phase to S. Note that this satisfies the condition for all positions of R, since the real and image incident path lengths, SR and $S'R$, respectively, are equal for all points on the surface.

6.3.3 Perfect reflection for parallel polarisation

Figure 6.3*b* illustrates the same principle applied to parallel polarisation. In this case only the components of the electric fields tangential to the reflecting surface need to be considered.

The real incident electric field E at reflection point R can be resolved into its components normal and tangential to the reflecting surface, E_n and E_t, respectively. The tangential component E_t will be cancelled if the image source is of equal amplitude and in phase with S. The image tangential component E'_t will

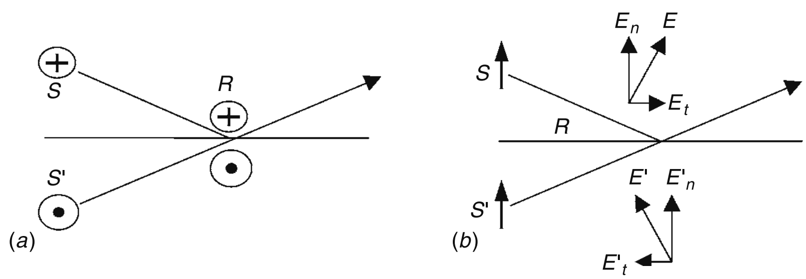


Figure 6.3 Perfect reflection
(a) Perpendicular polarisation
(b) Parallel polarisation

then be equal and opposite to E_t, and again this will be true for any point on the surface.

6.3.4 Discussion of perfect-reflection results

The above results suggest that for a perfect conductor the reflection coefficient will be $-1 + j.0$ for perpendicular polarisation and $1 + j.0$ for parallel polarisation. In fact, as shown in the following section of this Chapter, for small incidence angles (relative to the surface) reflection coefficients tend in practice to be close to -1 for both polarisations.

6.4 Reflection by finitely-conducting surfaces

The reflection coefficient at a general plane interface between two media can be derived from first principles as described below. Before doing so it will be convenient to define the relevant electrical properties of media involved in the reflection.

6.4.1 Electrical properties relevant to reflection

The following electrical properties must be known for both media in order to calculate the complex reflection coefficient at a plane boundary between them:

Permittivity:

$\varepsilon \;=\; \varepsilon_0 \varepsilon_r$

where:

ε_0 = permittivity of vacuum
= $8.854.10^{-12}$ F m^{-1}
ε_r = relative permittivity
= 1.0 for vacuum (and to a good approximation for air)

Permeability:

$\mu \;=\; \mu_0 \mu_r$

where:

μ_0 = permeability of vacuum
= $4\,\pi.10^{-7}$ H m^{-1}
μ_r = relative permeability
= 1.0 for vacuum (and $\approx$ 1.0 for nonferromagnetics)

Two important characteristics of radiowaves are derived from these quantities:

wave impedance = $\sqrt{(\mu/\varepsilon)}$
impedance of vacuum = $\sqrt{(\mu_0/\varepsilon_0)}$ = 376.7 Ω

wave velocity = $\sqrt{(1/\mu \,.\, \varepsilon)}$
velocity in vacuum = $\sqrt{(1/\mu_0 \,.\, \varepsilon_0)}$ = $2.998.10^8$ m s^{-1}

6.4.2 Snell's law for angle of refraction

Snell's law, which gives the relationship between the angles of incidence and refraction, is described at this point since it will be needed when deriving reflection coefficients. Figure 6.4 shows the geometry of Snell's law.

The change in angle on refraction is a function of the velocity of propagation in the two media. In Figure 6.4 the velocity is lower in medium 2. Although the wavefront in medium 1 travels d_1, the wavefront in medium 2 travels d_2. Thus:

$$\frac{\text{velocity}_1}{\text{velocity}_2} = \frac{d_1}{d_2} = \frac{\cos\theta_1}{\cos\theta_2} = \frac{\mu_2\varepsilon_2}{\mu_1\varepsilon_1} \tag{6.2}$$

If, as is normal, $\mu_1 = \mu_2$, then:

$$\frac{\cos\theta_1}{\cos\theta_2} = \sqrt{\frac{\varepsilon_{r2}}{\varepsilon_{r1}}} \tag{6.3}$$

Where medium 1 is a vacuum (or air), Snell's law is often quoted in terms of the refractive index of the other material. Refractive index n is given by:

$$n = c/v \tag{6.4}$$

where c and v are the velocities in vacuum and the material, respectively. Thus:

$$n = \sqrt{\varepsilon_r} \tag{6.5}$$

6.4.3 Continuity of tangential electric fields

Figure 6.5 shows a vanishingly-thin rectangular loop cutting the boundary between two media.

In Figure 6.5, the line integral of electric field $E = E_1 - E_2$ around the loop, by Maxwell's equations, is proportional to the rate of change of magnetic field H through it. But since the loop is vanishingly thin the area of the loop, and thus

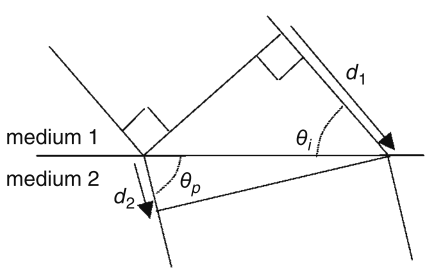


Figure 6.4 Geometry of Snell's law

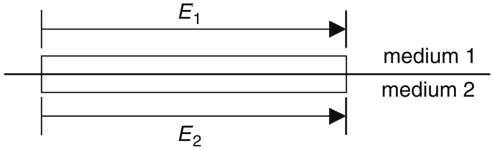


Figure 6.5 Continuity of tangential electric field at boundary

the surface integral of the field through it, is vanishingly small. Thus we can write:

$$\oint E \cdot dl = -\frac{\delta}{\delta t}\iint \mu \cdot H \cdot ds = 0 \tag{6.6}$$

It follows that $E_1 = E_2$. This result is true for all media. Note that if either medium is a perfect conductor then $E_1 = E_2 = 0$.

6.4.4 Continuity of tangential magnetic field

Referring to Figure 6.5 again with E_1 and E_2 replaced by H_1 and H_2, respectively, the line integral of magnetic field $H = H_1 - H_2$ around this loop, by Maxwell's equations, is proportional both to the rate of change of electric field and of actual current through it. For a vanishingly thin loop these will both also be vanishing, and thus:

$$\oint H \cdot dl = -\frac{\delta}{\delta t}\iint_s \varepsilon \cdot E \cdot ds + \iint_s J \cdot ds = 0 \tag{6.7}$$

where J = current density in A.m^{-2}.

6.4.5 Complex permittivity

For a dieletric, $J = 0$. In the case of a finitely-conducting medium and a harmonic signal, Maxwell's equation for curl.H can be written:

$$\begin{aligned} \text{curl}.H &= j\omega\varepsilon E + J \\ &= j\omega E + \sigma E \\ &= j\omega\,(\varepsilon + \sigma/j\omega)E \end{aligned} \tag{6.8}$$

Thus the permittivity of a finitely-conducting material is complex, and may be written:

$$\text{complex permittivity} = (\varepsilon - j\sigma/\omega) \tag{6.9}$$

where:

$\omega = 2\pi f$ where f = frequency in Hz
σ = conductivity in S m^{-1}
j = complex operator = $\sqrt{-1}$

6.4.6 General complex reflection coefficients

We can now derive general reflection coefficients. The derivation will be given for perpendicular polarisation, and an expression simply stated for parallel polarisation.

Figure 6.6 shows the incident, reflected and refracted rays, with the E phasors pointing out of the paper, and the H vectors parallel to the paper as required for the direction of power transmission S for each ray.

From the continuity of tangential electric fields:

$$E_i + E_r = E_p \tag{6.10}$$

From the continuity of tangential magnetic fields and noting that $\theta_i = \theta_r$:

$$(H_i - H_r)\sin\theta_i = H_p \sin\theta_p \tag{6.11}$$

which is equivalent to:

$$(E_i - E_r)\sin\theta_i / Z_1 = E_p \sin\theta_p / Z_2 \tag{6.12}$$

where Z_1 and Z_2 are the impedances of medium 1 and 2, respectively.

Assuming that $\mu_1 = \mu_2$:

$$(E_i - E_r)\sin\theta_\mathrm{i}\sqrt{\varepsilon_{r1}} = E_p \sin\theta_p \sqrt{\varepsilon_{r2}} \tag{6.13}$$

Combining Eqns. 6.11 and 6.13 to eliminate E_p:

$$\frac{E_r}{E_i} = \frac{\sin\theta_i\sqrt{\varepsilon_{r1}} + \sin\theta_i\sqrt{\varepsilon_{r2}}}{\sin\theta_i\sqrt{\varepsilon_{r1}} - \sin\theta_i\sqrt{\varepsilon_{r2}}} \tag{6.14}$$

It may be noted here that, if medium 1 is air or a vacuum, $\theta_p > \theta_i$ and thus E_r always has the opposite sign to E_i.

Using Snell's law (Eqn. 6.3) to write θ_p in terms of θ_i:

$$\frac{E_r}{E_i} = \frac{\varepsilon_{r1}\sin\theta_i - \sqrt{\varepsilon_{r2} - \varepsilon_{r1}\cos^2\theta}}{\varepsilon_{r1}\sin\theta_i + \sqrt{\varepsilon_{r2} - \varepsilon_{r1}\cos^2\theta}} \tag{6.15}$$

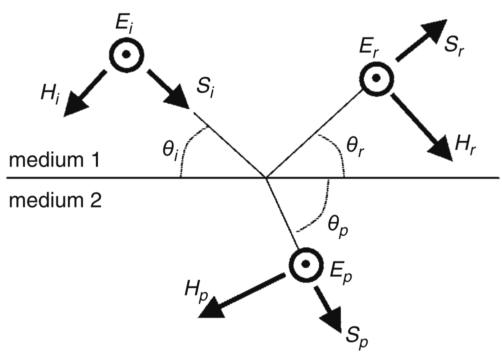


Figure 6.6 Vector relationships in reflection and refraction

If medium 1 is air or vacuum, for which $\varepsilon_{r1} = 1$, this can be simplified to:

$$\frac{E_r}{E_i} = \frac{\sin\theta_i - \sqrt{\varepsilon_{r2} - \cos^2\theta}}{\sin\theta_i + \sqrt{\varepsilon_{r2} - \cos^2\theta}} \tag{6.16}$$

If the reflecting surface has nonzero finite conductivity, as shown in 6.4.5, ε_{r2} can be replaced by the complex equivalent $(\varepsilon_2 - j.\sigma_2/\omega)/\varepsilon_0$, from which:

$$\frac{E_r}{E_i} = \frac{\sin\theta_i - \sqrt{(\varepsilon_{r2} - j\sigma_2/\omega\varepsilon_0) - \cos^2\theta_i}}{\sin\theta_i + \sqrt{(\varepsilon_{r2} - j\sigma_2/\omega\varepsilon_0) - \cos^2\theta_i}} \tag{6.17}$$

The complex term in Eqn. 6.17 is often written in the form $(\varepsilon_r - j\chi)$ where:

$$\chi = \frac{\sigma_2}{\omega\varepsilon_0} = \frac{\sigma_2}{2\pi f} \cdot \frac{1}{8.854{\cdot}10^{12}} = 18 \cdot 10^9\ \sigma_2/f \tag{6.18}$$

Normally medium 1 is air, and thus suffices are not needed for the electrical properties of medium 2. The complex reflection coefficient for perpendicular incidence on a surface is therefore given by:

$$\rho_{\text{perp}} = \frac{\sin\theta - \sqrt{(\varepsilon_r - j\chi) - \cos^2\theta}}{\sin\theta + \sqrt{(\varepsilon_r - j\chi) - \cos^2\theta}} \tag{6.19}$$

Reasoning similar to that in Eqns. 6.10 to 6.19 can be used to show that the complex reflection coefficient for parallel polarisation is given by:

$$\rho_{\text{para}} = \frac{(\varepsilon_r - j\chi)\sin\theta - \sqrt{(\varepsilon_r - j\chi) - \cos^2\theta}}{(\varepsilon_r - j\chi)\sin\theta + \sqrt{(\varepsilon_r - j\chi) - \cos^2\theta}} \tag{6.20}$$

where, in Eqns. 6.19 and 6.20:

θ = incidence angle measured between the ray and surface
ε_r = relative permittivity of the reflecting material
$\chi = 18{\cdot}10^9\sigma/f$
σ = the conductivity of the reflecting material in S m^{-1}
f = frequency in Hz

Figure 6.7 illustrates the general behaviour of these coefficients for electrical properties typical of good and poor ground conductivity at 1 MHz.

For perpendicular polarisation (horizontal in the case of ground reflection) the amplitude of the reflection coefficient falls from unity at grazing incidence to a lower value at normal incidence, depending on frequency and the electrical properties of the ground. The phase angle is always 180°, that is, a phase reversal on reflection.

For parallel polarisation (vertical in the case of ground reflection) the behaviour of the reflection coefficient is more complicated. At grazing and normal incidence the amplitude is the same as that for perpendicular. This is necessarily so at normal incidence since both polarisations represent the same geometry. Between the extremes of grazing and normal incidence, however,

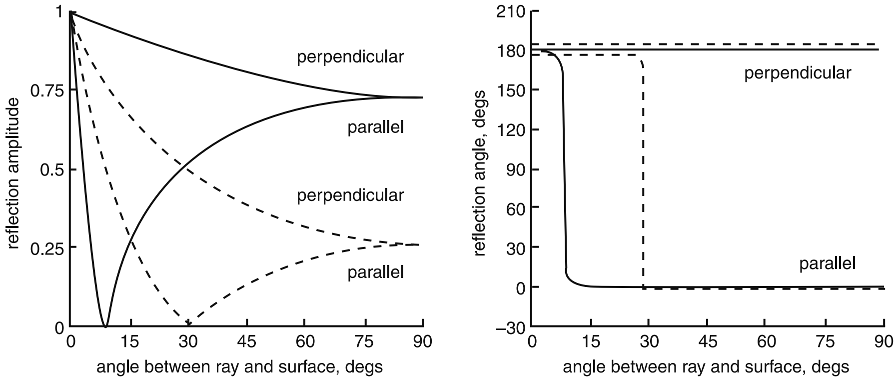


Figure 6.7 Reflection coefficient amplitude and phase at 450 MHz for: well conducting ground, $\sigma = 0.03$ S m^{-1}, $\varepsilon_r = 40$, solid curves; poorly-conducting ground, $\sigma = 0.0001$ S m^{-1}, $\varepsilon_r = 3$, dashed curves.

parallel polarisation falls to a reflection minimum at what is termed the pseudo-Brewster angle. For smaller incidence angles the phase change for parallel polarisation is 180°, as for perpendicular, but above the pseudo-Brewster angle it switches to zero and remains at this value up to normal incidence. The opposite phase angle between perpendicular and parallel polarisation at normal incidence arises from the convention used to define polarisation; the geometry of the rays and the surface are the same in both cases.

The pseudo-Brewster angle moves towards grazing incidence for better conductors, and particularly as the relative permittivity of the reflecting surface increases. For a perfect conductor the incidence angle becomes zero, and the switch in phase for parallel polarisation does not take place. Thus, although for a perfectly-conducting surface the complex reflection coefficients can be taken as $-1 + j.0$ for perpendicular polarisation and $1 + j.0$ for parallel, as noted in Section 6.3.4, in practice at grazing incidence the complex reflection coefficient can be approximated to $-1 + j.0$ in both cases. This is a common assumption in terrestrial propagation models, which often take no account of polarisation.

Reflection can have a marked effect on a circularly-polarised wave, since its two linear components can be viewed as having perpendicular and parallel polarisation. Thus above the pseudo-Brewster angle the handedness of the wave will be reversed (left-handed changed to right-handed, and *vice versa*), and in general a change from circular to elliptical will take place. At the pseudo-Brewster angle one linear component will be suppressed, and the reflected wave will thus be linearly polarised.

6.5 The plane-earth two-ray reflection model

For short radio paths it is sometimes sufficiently accurate to ignore earth curvature. For low antennas and uncluttered terrain it can also be assumed that ground reflection will be taking place at grazing incidence. In these circumstances the plane-earth two-ray model, which assumes a reflection coefficient of –1, can be a useful approximation.

6.5.1 Explicit calculation

Figure 6.8 shows the geometry, where s_1 and s_2 are the slope distances of the direct and reflected rays, given by:

$$s_1 = \sqrt{d^2 + (h_1 - h_2)^2} \tag{6.21a}$$

$$s_2 = \sqrt{d^2 + (h_1 + h_2)^2} \tag{6.21b}$$

where d is the horizontal distance and h_1 and h_2 are the heights of the antennas above ground, all in the same units.

The field strength resulting from the combination of the direct and reflected rays can be written in complex notation as:

$$e = A\left\{\frac{\exp(-jks_1)}{s_1} + \rho \cdot \frac{\exp(-jks_2)}{s_2}\right\} \tag{6.22}$$

where:

ρ = the complex reflection coefficient of the ground
A = is a normalising constant.

Figure 6.9 shows Eqn. 6.22 plotted in dB(μV m^{-1}) for 1 kW e.r.p. at 900 MHz over a range of (horizontal) distances for a flat path with both antennas 3 m above ground level, and with $\rho = -1$. For comparison the free-space field strength is shown dashed.

The variation of field strength due to the combined direct and reflected rays, as path length increases, shows the following features:

(a) Up to about 50 m the path length difference between the two rays is greater than a wavelength, but decreases as the path length increases. Thus the two

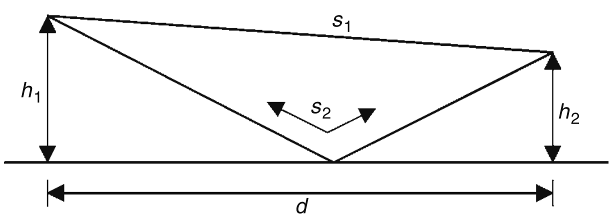


Figure 6.8 Geometry of plane-earth two-ray reflection

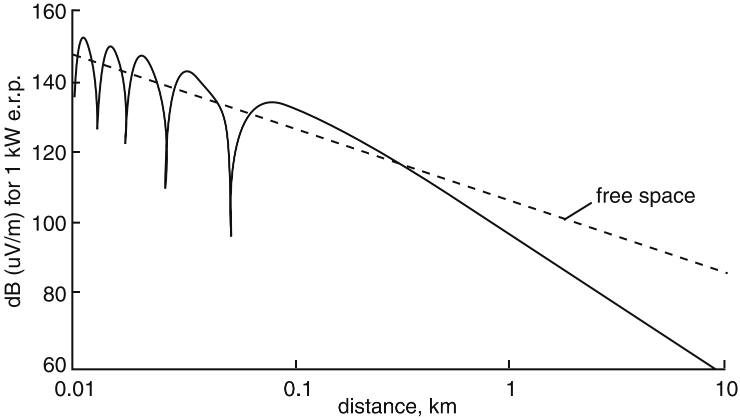


Figure 6.9 Plane-earth two-ray reflection for: 1 kW e.r.p. at 900 MHz, ground reflection coefficient = –1, both antennas 3 m above ground level

components of the signal pass periodically in and out of phase, causing alternate reinforcement and cancellation known as lobing. The peak reinforced signal levels are 6 dB above free space, and the signal is 3 dB or more above free space for 50 per cent of locations. This apparent doubling of power is due to signals received from the real transmitter and its image in the reflecting surface. As noted earlier, this occurs only in the half space on one side of the reflecting surface, and thus does not violate conservation of energy.

(b) At about 55 m the path length difference is one wavelength. Due to the phase reversal upon reflection this causes the last cancellation of the signal.
(c) At about 100 m the path length difference is half a wavelength. With the phase reversal of the reflected ray this causes the last reinforcement of the signal.
(d) At greater distances the path length difference is less than half a wavelength and progressively reducing towards zero, and thus the signal continuously approaches a further cancellation.

The situation described in (*d*) can be modelled by a simple expression, derived in Section 6.6.

6.6 Approximation to two-ray reflection

For $h_1 << d >> h_2$ it can be shown that the difference in path length between the reflected and direct rays can be approximated by $2h_1h_2/d$, for which the resulting phase difference $\Delta\phi$ is $\pi h_1 h_2 f / 75d$ radians, where the heights and d are in metres and f is in MHz. Figure 6.10 shows the phasor diagram for distances well beyond the final reinforcement where $\Delta\phi$ is small. Note that the directions of the arrows representing the direct and reflected phasors take into account phase reversal upon reflection.

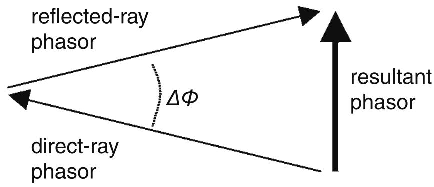


Figure 6.10 Phasor diagram for path length difference << wavelength

When $\Delta\phi$ is small, the amplitude of the resultant phasor is approximately equal to $\Delta\phi$ multiplied by the length of the direct-ray phasor. Thus the field strength relative to free-space E_{fs} as is given by:

$$E_{rf} = 20\log(\pi h_1 h_2 f / 75d) \quad \text{dB} \tag{6.23}$$

Evaluating the constant and converting to distance in km, E_{fs} can be written:

$$E_{rf} = -87.6 + 20\log(h_1 h_2 f) - 20\log(D) \quad \text{dB}(\mu\text{V m}^{-1}) \tag{6.24}$$

where D is the path length in km.

Since the free-space field strength in dB(μV m^{-1}) for 1 kW e.r.p. is given by:

$$E_{fs} = 106.9 - 20\log(D) \quad \text{dB}(\mu\text{V m}^{-1}) \tag{6.25}$$

the field strength in dB(μV m^{-1}) for 1 kW e.r.p. for the approximation to the two-ray model is:

$$E_2 = 19.3 + 20\log(h_1 h_2 f) - 40\log(D) \quad \text{dB}(\mu\text{V m}^{-1}) \tag{6.26}$$

Figure 6.11 shows the same information as Figure 6.10 with the addition of the approximate plane-earth two-ray reflection model as given by Eqn. (6.26).

The line-of-sight (LOS) model in ITU-R P.1411 Section 4.1 for propagation within a street canyon is based on the plane-earth two-ray model. It is important

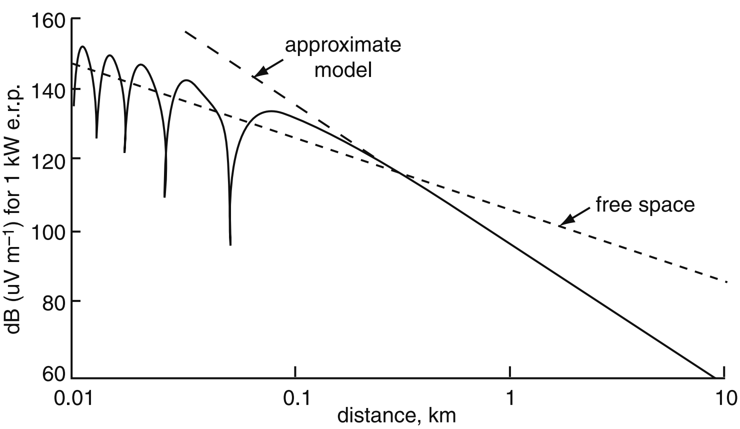


Figure 6.11 Plane earth two-ray reflection – comparison of exact and approximate models

to note the restricted range of validity for Eqn. 6.26. It is necessary to limit the prediction of field strength at short distances such that it does not greatly exceed free space. At long distances care should be taken to ensure that the assumption of a line-of-sight path with uncluttered ground reflection does not become unrealistic.

6.7 Reflection and scattering from rough surfaces

It has been assumed up to this point that the reflecting surface is flat and smooth. This gives rise to reflections which are termed specular, exactly analagous to optical reflections in a mirror.

In practice many reflecting surfaces along a radio path, such as the ground and the walls and roofs of buildings, will not be smooth. Depending on the degree of roughness in relation to the wavelength the effect will vary from specular reflection to diffuse scattering.

Figure 6.12 shows the geometry of the Rayleigh roughness criterion, which is widely used to assess whether reflection or scattering will take place.

In figure 6.12 the roughness has been idealised to rectangular steps of height Δh. For a plane wave incident at θ_i (measured to the surface) the phase difference $\Delta\phi$ between reflections from the upper and lower parts of the surface, due to the difference in path length AC minus AB, is given by:

$$\Delta\phi = 4\,\pi\,\Delta h \sin(\theta_i)/\lambda \quad \text{radians} \tag{6.27}$$

If $\Delta\phi$ is small, e.g. less than 0.3 radians, the surface will support specular reflection. If it is large, say greater than three radians, the surface is rough and will produce scattering. Real surfaces do not normally have such idealised irregularities, and the Rayleigh criterion is only an approximate guide.

Figure 6.13 illustrates totally diffuse scattering from a surface, in which the specular component is negligible. This is sometimes referred to as Lambertian, following an observation by Lambert that an optically illuminated surface has the same apparent brightness when viewed from any angle. The dependence on both angles, θ_i and θ_s, is necessary to maintain reciprocity, and is consistent with Lambert's observation. Thus in Figure 6.13, the energy incident on area δA at

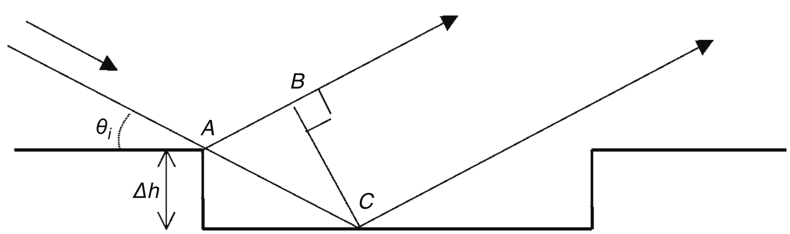


Figure 6.12 Geometry of the Rayleigh roughness criterion

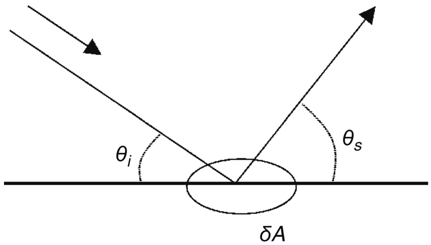


Figure 6.13 Geometry of Lambertian scattering

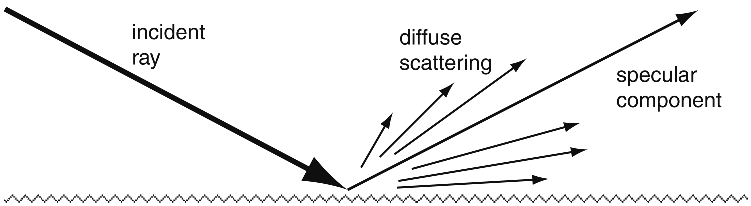


Figure 6.14 Combination of specular reflection and diffuse scattering

angle θ_i is proportional to $\sin(\theta_i)$ and the energy radiated from area δA at angle θ_s is proportional to $\sin(\theta_s)$.

For surfaces with intermediate roughness, the incident energy can be modelled as though divided into a specularly-reflected component and a diffusely-scattered component. The reduction in amplitude of the specularly-reflected field can be modelled by:

$$f_r = \exp(-\Delta\phi^2) \tag{6.28}$$

where $\Delta\phi$ is given by Eqn. 6.28 with Δh replaced by the standard deviation of surface roughness. It may be noted that the criteria of 0.3 and 3.0 radians quoted above for the Rayleigh roughness test will result in reductions of 0.8 and 78 dB in the specular component, according to this model.

Figure 6.14 illustrates a combination of specular and diffuse scattering in terms of power flux. As a simple approximate model, if W_i (W m^{-2}) is incident on the surface, $\rho f_r^2\ W_i$ (W m^{-2}) will be radiated as a coherent specular reflection, and the remaining power $\rho\ (1 - f_r^2)\ W_i$ will consist of incoherent radiation in other directions.

The directional distribution of the scattered power in Figure 6.14 will depend on the angle of incidence, and the nature and roughness of the surface.

Chapter 7

Clear-air characteristics of the troposphere

K.H. Craig

7.1 Introduction

This Chapter considers the effects of refractive index variations on the propagation of radiowaves in the troposphere, and in particular those mechanisms which lead to propagation beyond the normal line of sight. Clear air implies that the effects of condensed water (clouds, rain etc.) are ignored, although gaseous absorption is included. The influence of terrain diffraction is covered in Chapter 8, but terrain reflections are discussed here insofar as they contribute to the clear-air space wave. The frequencies of interest are above about 100 MHz; below this frequency refractive index variations are not strong enough to cause significant effects, and the ground wave and ionospheric mechanisms dominate at transhorizon ranges.

The emphasis is on the meteorological mechanisms that give rise to anomalous propagation, and the basic models that have been developed to predict the effects of refractive index variations on radiowave propagation. Statistical procedures for the prediction of radio link reliability are the subject of Chapter 20.

7.2 Causes and effects of refraction in the troposphere

7.2.1 Electromagnetic waves

Electromagnetic waves propagating in the troposphere are refracted and scattered by variations in the radio refractive index n. Recall that the electromagnetic field of a plane wave propagating in a medium of constant refractive index, n, has a space, $\boldsymbol{r}$, and time, t, variation given by:

$$E(\boldsymbol{r}, t) = E_0 \exp[i(n\boldsymbol{k}_0 \,.\, \boldsymbol{r} - \omega t)] \tag{7.1}$$

where $\omega = 2\pi \times$ frequency and $\boldsymbol{k}_0$ is a vector normal to the wavefront with a magnitude equal to the free space wavenumber ($= 2\pi$/wavelength).

In the troposphere, the refractive index is not constant. At microwave frequencies, however, it varies slowly on the scale of a wavelength. In this case it is still possible to write:

$$E(\boldsymbol{r},t) \approx E_0 \exp[i(n(\boldsymbol{r})\boldsymbol{k}_0 \cdot \boldsymbol{r} - \omega t)] \tag{7.2}$$

although the magnitude of $E(\boldsymbol{r},t)$ will in general vary with position. The value and variations of $n(\boldsymbol{r})$ are fundamental to understanding the way in which electromagnetic waves propagate through the troposphere. For example, Snell's law of refraction and the Fresnel coefficients for reflection and transmission at an interface follow from Eqn. 7.2 by applying appropriate boundary conditions across a boundary separating media of different refractive indices. In this Chapter we are principally interested in refractive effects. We first consider the determination of n in the troposphere.

7.2.2 Radio refractive index

The radio refractive index of the troposphere is due to the molecular constituents of the air, principally nitrogen, oxygen, carbon dioxide and water vapour. The value of n deviates from unity because of the polarisability of these molecules due to the incident electromagnetic field, and quantum mechanical molecular resonances. The latter effect is limited to narrow frequency bands (for example, around 22 GHz and 60 GHz). We first discuss the former effect which is independent of frequency at the frequencies of interest (up to millimetre waves).

The deviation of n from unity is very small in absolute terms, a typical value being 1.0003 at the earth's surface. Because of the closeness of n to unity, it is usual to work with the refractivity, N, defined by:

$$N = (n - 1) \times 10^6 \tag{7.3}$$

N is dimensionless, but for convenience is measured in N units. N depends on the pressure P (mbar), the absolute temperature T (K) and the partial pressure of water vapour e (mbar):

$$N = 77.6\,\frac{P}{T} + 3.73 \times 10^5\,\frac{e}{T^2} \tag{7.4}$$

This is derived from the Debye formula [1] for the polarisability of polar (i.e. with a strong permanent electric dipole moment) and nonpolar molecules. The first (dry) term is due principally to the nonpolar nitrogen and oxygen molecules, and the second (wet) term is from the polar water vapour molecules. The constants are empirically determined, based on experimental measurements [2]. An excellent discussion of Eqn. 7.4 is given in Reference [3].

The variation of P, T and e can be considered at various scales:

(a) on the largest (global) scale the troposphere is stratified in horizontal layers due to the effect of gravity;
(b) on the medium scale (100 m–100 km) the ground and meteorology (local or mesoscale) can produce spatial and temporal variations;
(c) on the small scale (<100 m) turbulent mixing causes scattering and scintillation.

The macroscopic, large-scale, structure of the troposphere varies much more rapidly vertically than horizontally. Figure 7.1 shows contours of potential refractivity (i.e. refractivity reduced to a standard pressure level) derived from aircraft measurements made over the English Channel. Bearing in mind the greatly exaggerated vertical scale, the variations are about two orders of magnitude greater in the vertical direction than in the horizontal. (Actually the inhomogeneities in Figure 7.1 are more severe than in many locations because of the influence of the coastal zones at each side of the Channel.) An assumption of horizontal stratification of the troposphere on this scale is therefore justified. In practice, the same stratification may persist over a horizontal region tens or hundreds of kilometres in extent. The inhomogeneities can have significant effects on radiowave propagation, however, as will be illustrated later.

In an atmosphere at rest with no heat sources, the pressure can be shown to decrease exponentially with height, dropping to a fraction $1/e$ of its value at the surface at a height of approximately 8 km (the scale height). In unsaturated air, the temperature falls linearly with height at about 1 °C per 100 m (the dry

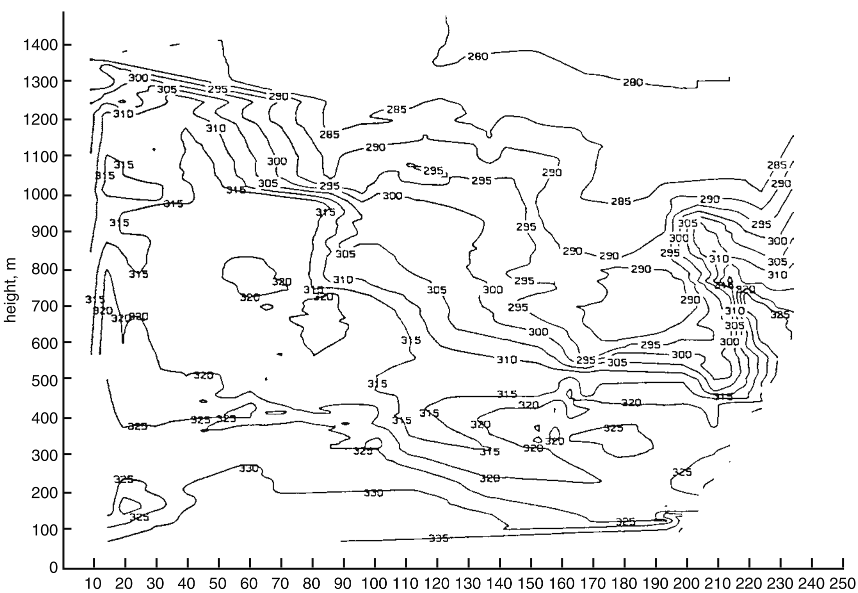


Figure 7.1 Contours of potential refractivity measured over the English Channel

adiabatic lapse rate). The behaviour of water vapour pressure is more complicated. Ignoring condensation, it would fall exponentially at the same rate as the pressure. However, air at a given temperature can hold only a limited amount of water vapour; the limit occurs at the saturated water vapour pressure, e_s, which varies from 43 mbar at 30 °C to 6 mbar at 0 °C. Above this limit, water condenses to form water droplets (clouds). Since the saturation vapour pressure decreases as temperature decreases, and temperature decreases with height, condensation will occur above a certain height, reducing the water vapour content of the air. Thus the water vapour pressure decreases more rapidly with height than pressure and for practical purposes is negligible above 2 or 3 km. Above the condensation level, the temperature follows the saturated adiabatic lapse rate which is less than the dry adiabatic lapse rate because of the latent heat released by the condensation process.

For reference, note that several quantities other than water vapour pressure are used to characterise the water vapour content of the air: examples are relative humidity, (= e/e_s, expressed as a percentage), water vapour density (= 216.7 e/T, g m^{-3}) and humidity mixing ratio (= 622 e/P, g kg^{-1}); it may also be specified in terms of the dew point or wet bulb temperatures.

The net effect of the variations in P, T and e is that N decreases with height. On average N decreases exponentially in the troposphere:

$$N = N_s \exp(-z/H) \tag{7.5}$$

where N_s is the surface value of refractivity, z is the height above the surface and H is the scale height. Average mid-latitude values are $N_s = 315$ and $H = 7.35$ km [4]. Maps showing the geographical and seasonal variation of N_0 (the value of N_s at sea level) are given in References [3] and [4].

7.2.3 Effect of the refractive index on radiowaves

If the refractive index were constant, radiowaves would propagate in straight lines. For present purposes an adequate physical picture is to consider the radiowaves as propagating out from a transmitter along ray paths. Initially, we assume stratification of the atmosphere and ignore variations in the horizontal direction. Changes in refractive index with height then determine the bending of the ray paths in a vertical plane. The amount of bending between level 1 and level 2 is determined by Snell's Law (i.e. $n_1 \sin\theta_1 = n_2 \sin\theta_2$, where $n_{1,2}$ and $\theta_{1,2}$ are, respectively, the refractive indices and the angles that the rays make to the vertical in levels 1 and 2). Since n decreases with height, rays are bent downwards toward the earth. An immediate consequence is that the radio horizon lies further away than the visible horizon (Figure 7.2).

For a radio path extending through the atmosphere, this refractive bending causes the elevation angle of a ray at the ground to be greater than if the atmosphere were not present. Figure 7.3 shows the computed relationship between the elevation angle correction and the true elevation angle for a slant path through the atmosphere in a tropical and a polar climate [5]. The curve for

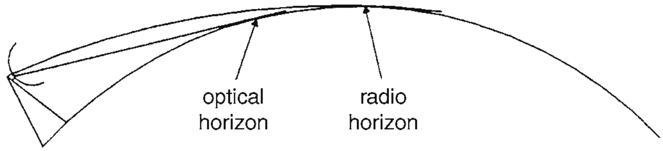


Figure 7.2 Extension of radio horizon due to tropospheric bending (greatly exaggerated scale)

the UK lies close to the mean of those shown. These elevation angle offsets can be important, for example, in estimating target heights from radar returns.

For heights much less than the scale height, the exponential in Eqn. 7.5 can be approximated by the first term in its expansion, giving a linear decrease of refractivity with height at a rate of about 40 N units per kilometre at mid latitudes. This approximation is excellent for terrestrial paths but is inadequate for airborne radar calculations and earth–satellite paths at low elevation angles.

It can be shown that the radius of curvature, C, of a ray is very well approximated by:

$$C = -\frac{dn}{dz} \tag{7.6}$$

at low elevation angles. The curvature of the earth is $1/a$ where a is the earth's radius (6378 km). Thus the curvature of the ray relative to the curvature of the Earth is $(-\,dn/dz - 1/a)$. Since we are often mainly interested in this relative

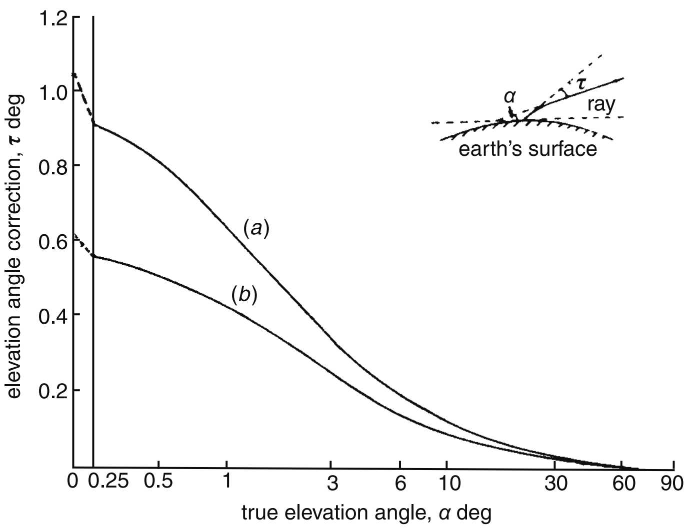


Figure 7.3 Error in elevation angle due to tropospheric refraction
(a) tropical maritime air (July)
(b) polar continental air (April)

curvature, it is useful to introduce the concept of an effective earth radius $a_e = ka$, where k is known as the k factor. Then we have:

$$-\frac{dn}{dz}-\frac{1}{a}=-\frac{dn_{eff}}{dz}-\frac{1}{a_e} \tag{7.7}$$

where n_{eff} is the effective refractive index associated with the effective earth radius a_e. We have already seen that $C = -dn/dz = 40 \times 10^{-6}$ km^{-1} in the average mid-latitude atmosphere, while $1/a = 157 \times 10^{-6}$ km^{-1}. (Note that the curvature of the earth substantially exceeds the downward curvature of the ray.) Straight line ($dn_{eff}/dz = 0$) ray propagation relative to the effective earth radius can then be arranged by setting $1/a_e = 117 \times 10^{-6}$ km^{-1}, corresponding to $k = 4/3$. This is the origin of the well known 4/3 earth's radius construction so useful in engineering calculations: a ray propagating in a straight line over terrain based on a 4/3 effective earth radius is equivalent to a ray propagating in an atmosphere with the average lapse rate of 40 N km^{-1} over the actual terrain. Figure 7.4 shows different k factor representations of the same picture. Rays emanate from a transmitter on the left at a height of 25 m into a standard atmosphere (–40 N km^{-1}): note that the rays curve downwards slightly for $k = 1$ (when $dn_{eff}/dz = -40 \times 10^{-6}$ km^{-1}) and are straight for $k = 4/3$ (when $dn_{eff}/dz = 0$). For terrestrial radio links it is a simple matter to check for terrain clearance or obstruction by joining potential transmitter and receiver positions by straight lines on 4/3 earth-radius graph paper.

There is a third viewpoint, useful in ducting studies: replace the earth with a flat earth ($k = \infty$) and modify the curvature of the ray so that the relative curvature between ray and earth is preserved. In this case n_{eff} is known as the modified refractive index, m, and the refractivity N is replaced by the modified refractivity M:

$$M = N + 10^6 \times z/a = N + 157z \tag{7.8}$$

where the height z is given in kilometres, i.e.

$$\frac{\partial M}{\partial z}=\frac{\partial N}{\partial z}+157 \tag{7.9}$$

Note that rays curve upwards relative to a flat earth (Figure 7.4).

Although N decreases by about 40 N km^{-1} (M increases by about 117 N km^{-1}) in average conditions at mid latitudes in the lower troposphere, significant deviations from the average do occur. Figure 7.5 shows the distribution of mean refractivity gradient in the UK from the surface up to a range of heights [5]. If the lapse rate of N is less than 40 N km^{-1}, the downward curvature of radio rays will decrease, shortening the radio horizon and reducing the clearance above terrain on terrestrial paths; this is known as subrefraction. On the other hand, if the lapse rate of N exceeds 40 N km^{-1}, the ray curvature will increase, extending the radio horizon and increasing path clearance; this is known as super-refraction (see Figure 7.6).

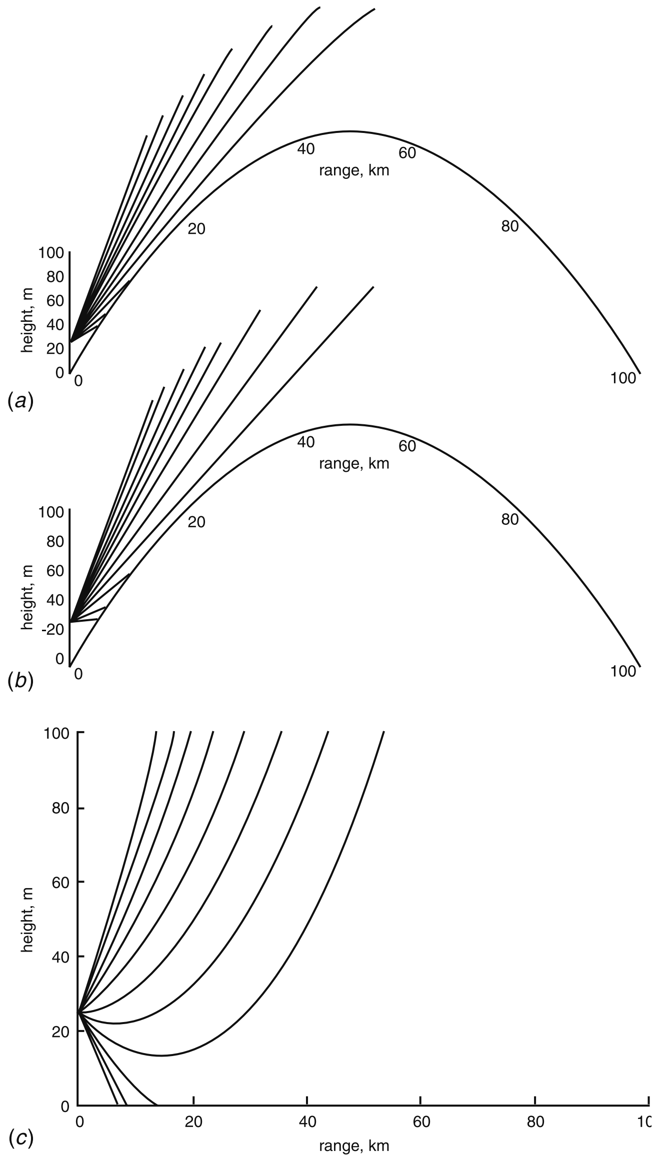


Figure 7.4 k-factor representations of rays propagating in a standard atmosphere:
(a) $k = 1$
(b) $k = 4/3$
(c) $k = \infty$ (flat earth)

When the lapse rate of N exceeds 157 N km^{-1}, i.e. $\partial N/\partial z < -157$, or equivalently, $\partial M/\partial z < 0$, then the rays are bent towards the earth more rapidly than the earth's curvature. This is known as ducting and can cause rays to propagate to extremely long ranges beyond the normal horizon. The usual classification of

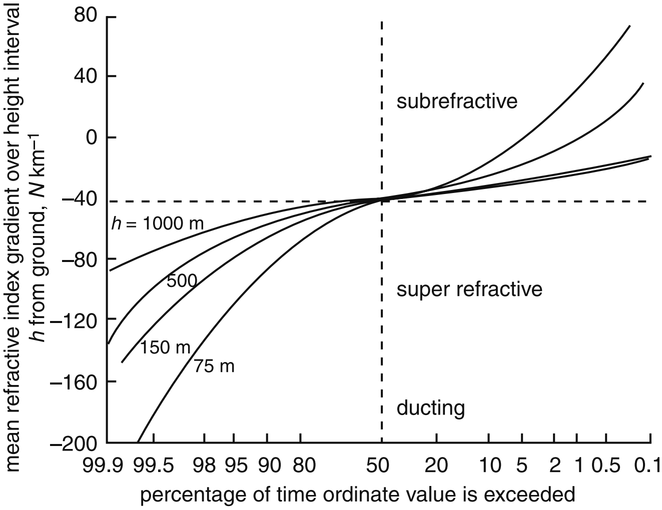


Figure 7.5 Distribution of mean refractive index gradient in the UK with height interval above ground level [5]

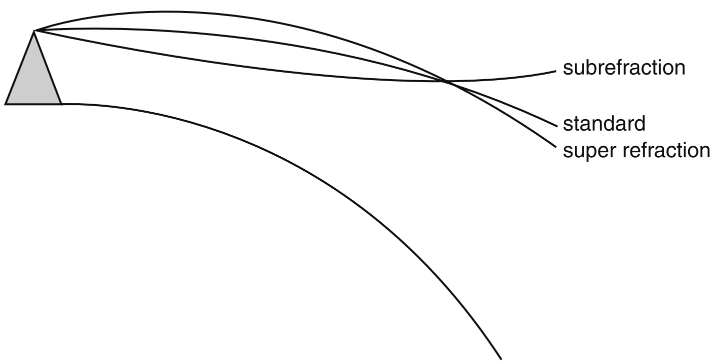


Figure 7.6 Classification of refractive conditions

propagation conditions in terms of refractivity gradients is given in Figure 7.7. The simple criterion for ducting in terms of modified refractivity gradients (the existence of a negative slope on the $M - z$ graph, irrespective of the scales of the axes) is one reason why M is the most useful quantity for ducting studies. Ducting is discussed later.

7.2.4 Gaseous absorption and complex refractive index

In Section 7.2.2 it was pointed out that molecular resonances make a significant contribution to the radio refractive index in certain frequency bands. Only oxygen and water vapour are relevant at frequencies below 350 GHz. These

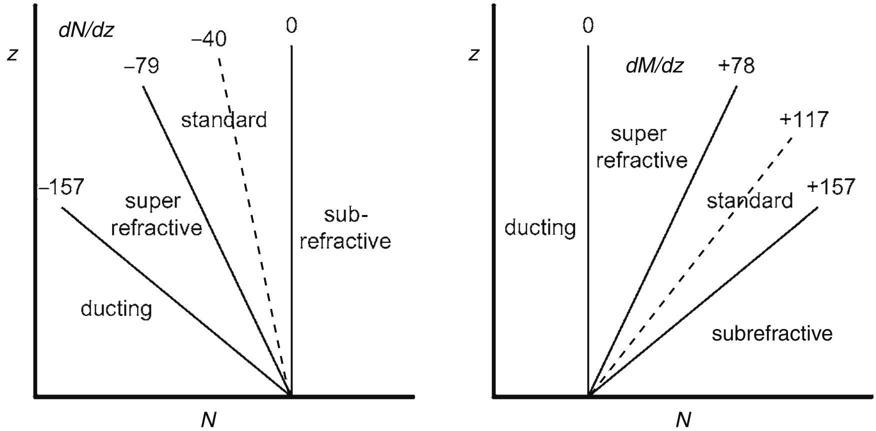


Figure 7.7 N–z and M–z plots of refractive condition classes (gradients are in N km^{-1} or M km^{-1})

resonance lines can cause significant absorption of radiowaves at frequencies near the lines.

The oxygen molecule has a permanent magnetic dipole moment due to paired electron spins. Changes in orientation of the combined electron spin relative to the orientation of the rotational angular momentum give rise to a closely spaced group of spin-flip or hyperfine transitions near 60 GHz, and a single line at 119 GHz. The water molecule has a permanent electric dipole moment, and rotations of the molecule with quantised angular momentum give rise to spectral lines at 22, 183 and 325 GHz.

Figure 7.8 shows the attenuation rate per kilometre at ground level caused by oxygen and water vapour, calculated using the methods given in Reference [6]. At low altitudes the lines are greatly widened by pressure (collision) broadening, and the complex of separate lines at 60 GHz cannot be resolved individually. At lower pressures near the top of the troposphere, the separate lines can be resolved. The absorption spectrum of water vapour has very intense lines in the far infra-red region; the low frequency tails of these lines are seen as the sloping baseline of the water vapour spectrum shown in Figure 7.8. Very significant absorption can occur, notably around the oxygen line complex at 60 GHz. This limits path lengths to a few kilometres at these frequencies. Unlike oxygen, the amount of water vapour in the atmosphere is variable even at the ground, and consequently the attenuation near the water vapour lines can vary significantly from place to place and day to day.

From the point of view of modelling, it is worth noting that gaseous absorption and refraction can be dealt with in a unified way. If the refractive index is regarded as a complex number ($n = \Re\,[n] + i\Im[n]$), Eqn. 7.2 can be written:

$$E(\boldsymbol{r},t) \approx E_0 \exp[i(\Re[n(\boldsymbol{r})]\boldsymbol{k}_0 \,.\, \boldsymbol{r} - \omega t]\exp[-\Im[n(\boldsymbol{r})]\boldsymbol{k}_0 \,.\, \boldsymbol{r}] \qquad (7.10)$$

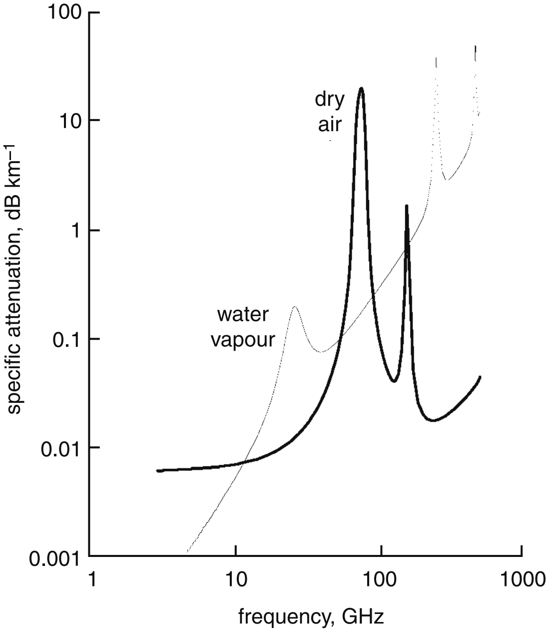


Figure 7.8 Specific attenuation at ground level due to gaseous absorption by oxygen and water vapour (water vapour density = 7.5g m^{-3}, temperature = 15 °C)

The imaginary part of n causes the field to decay exponentially with range, characteristic of absorption. In fact the specific attenuation is just $0.182 \times 10^6 f\Im[n]$ (dB km^{-1}) where f is the frequency (GHz).

7.2.5 *Refractive index measurements*

The most widely available source of refractive index data is the radiosonde. A radiosonde consists of an instrument package carried aloft by a freely ascending, gas-filled balloon. The instruments measure (directly or indirectly) pressure, temperature and humidity, from which height and refractive index along the ascent can be derived. The data is sampled and transmitted to a receiver on the ground. Radiosondes are launched twice daily at a large number of meteorological stations around the world for the purpose of weather forecasting. The vertical resolution of traditional radiosondes is poor in the first kilometre: thin ducting layers tend to be missed, or at best smoothed out, so underestimating the effect of ducting layers. The newer minisonde systems give better vertical resolution, are portable and can be launched by one person.

An alternative *in situ* measurement system is the refractometer. This instrument is basically an open-ended resonant microwave cavity, whose resonant frequency is determined by the refractive index of the air within the cavity [7]. Advantages of refractometers over radiosondes include rapid response time (and hence high resolution) and direct measurement of refractivity. However, the bulk and cost of these systems rule them out as throw-away sensors for balloon ascents. They have principally been used as high precision instruments mounted on aircraft or helicopters for detailed case studies.

Remote sensing of the troposphere is an active area of study. Techniques being investigated for refractivity sensing include Doppler radars, lidars (laser sounders) [8], sodars (acoustic sounders) and satellite-borne instruments. Although useful qualitative results have been achieved (particularly on the spatial and temporal distribution of ducting layers), none of these methods has yet been developed into a practical, quantitative tool.

Numerical weather models hold out promise for the widescale forecasting of refractivity. Recent mesoscale models covering limited regions have good vertical resolution. Although not specifically designed for refractivity purposes, all the necessary parameters exist in the models. Much testing and refinement is required before these can be used for radiowave prediction purposes.

7.3 Anomalous propagation: multipath and ducting

7.3.1 Types of duct

We have seen that if the vertical lapse rate of refractivity exceeds 157 $N\,\text{km}^{-1}$, then the waves bend downwards with a curvature greater than that of the earth, and ducting can occur. Radio energy can become trapped between a boundary or layer in the troposphere and the surface of the earth or sea (surface duct) or between two boundaries in the troposphere (elevated duct). In this waveguide-like propagation, very high signal strengths can be obtained at very long range (far beyond line of sight). Indeed, the signal strength may well exceed its free-space value. Eqn. 7.4 shows that two processes can cause the formation of high lapse rates: a rapid decrease in water vapour pressure with height, and an increase in temperature with height (a temperature inversion); these mechanisms often occur together. The vertical pressure gradient never deviates much from its standard value – large scale air movements (winds) rapidly restore pressure equilibrium.

The sensitivity of N to variations in the meteorological parameters can be found by differentiating Eqn. 7.4. Assuming typical atmospheric conditions ($P = 1000$ mbar, $T = 293$ K, $e = 15$ mbar), the variation in N is given by:

$$\delta N = 0.26\delta P + 4.3\delta e - 1.4\delta T \tag{7.11}$$

Differences of a few degrees in T and a few millibars in e can occur between adjacent air masses in certain meteorological conditions. This can lead to changes of several tens of N units over a height interval of tens of metres, and the formation of a ducting layer.

There are three important types of layer/duct which are illustrated in Figure 7.9 using actual radiosonde data. As discussed in Section 7.2.3, a ducting layer is

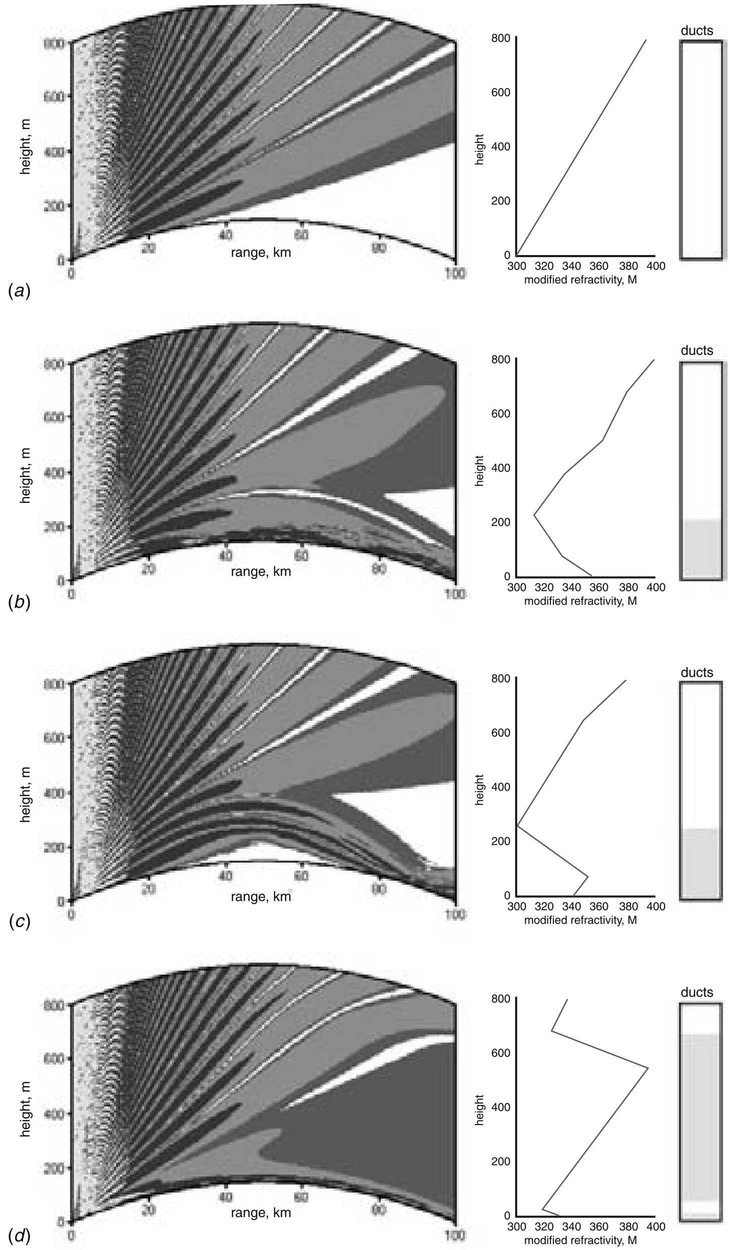


Figure 7.9 Definition of duct types, and their effect on a 3 GHz transmitter at 20 m height; the ducts are indicated by the vertical bars.
(a) standard atmosphere
(b) surface layer, surface duct
(c) elevated layer, surface duct
(d) elevated layer, elevated duct

immediately identifiable by a negative slope in the modified refractivity–height curve; the duct (marked by a vertical bar in Figure 7.9) extends from the top of the ducting layer down to the ground (in the case of a surface duct) or down to the height at which the modified refractivity returns to the same value as at the top of the layer (in the case of an elevated duct).

Figure 7.9*a* shows the standard atmosphere for reference. Figure 7.9*b* shows a surface duct caused by a surface layer. Figures 7.9*c* and *d* show elevated layers: Figure 7.9*c* forms a surface duct, while Figure 7.9*d* forms an elevated duct. It should be borne in mind that the refractivity profiles can be complicated and multiple layers do occur (such as the low surface duct in Figure 7.9*d* accompanying the elevated duct); nevertheless the concept of surface and elevated ducts is useful. In northern Europe, surface ducts occur for about 5 per cent, and elevated ducts for 5–10 per cent, of the time, averaged over the year. The percentage of the time that a radio link is affected by ducting will in general be less than these figures, because of the importance of the path geometry relative to the ducting layers.

The contour plots show the effect of the associated layers on a 3 GHz transmitter located on the left at a height of 20 m over the sea; the contours are given in terms of path loss. The common feature in all the plots is the lobing caused by interference between energy arriving at a point *via* the direct and ground reflected paths. The term multipath is often used to describe the situation where radiowaves propagate from transmitter to receiver by more than one path, especially on a line-of-sight link. The secondary paths may occur by ground reflection (as here) or by refraction. The distinction between ducting and multipath caused by interference between two or more refractive paths through the atmosphere is more a descriptive convenience than a fundamental difference in mechanism. At longer (transhorizon) ranges, the number of multiple paths becomes very large, and it is more appropriate to use models based on ducting theory. The effects of multipath on terrestrial line-of-sight paths are discussed in Chapter 20.

The distance to the normal radio horizon, d (km), is given in terms of the transmitter height, h (m) by:

$$d = 4.12\sqrt{h} \tag{7.12}$$

Both types of surface duct are seen to propagate energy well beyond the normal radio horizon (at 18.4 km range for a 20 m high transmitter). However, note the skip zone that occurs in the elevated layer, surface duct example.

A ducting layer will only trap radiowaves if certain geometrical constraints apply. In particular, the angle of incidence of electromagnetic energy at the layer must be very small. A simple rule of thumb, derived from the total-internal-reflection condition of geometrical optics, is that the maximum angle of incidence θ_{max} (degrees) is related to the change in refractivity ΔN (N units) across the layer by:

$$\theta_{max} = 0.081\sqrt{|\Delta N|} \tag{7.13}$$

As ΔN rarely exceeds 50 N units, θ_{max} will be limited to 0.5–1°. Simple geometrical considerations show that even energy launched horizontally will intercept an elevated layer at a nonzero angle due to the earth's curvature. It follows that ducting layers higher than about 1 km will not significantly affect terrestrial radio links. For example, although the elevated duct of Figure 7.9*d* distorts the interference lobes significantly, it is not strong enough to trap the energy completely.

The dependence of ducted field strength on frequency and path geometry is complicated. At long ranges from the transmitter, a simple single mode propagation model of a surface duct (see Section 7.4.3) predicts that the field strength should decrease exponentially with distance (i.e. linear in decibels) because of mode attenuation. This is indeed reflected in the ITU-R prediction method for the prediction of interference levels caused by ducting and layer reflection (see Chapter 20, Figure 20.17, where γ_d represents the specific attenuation of the duct). However, the attenuation rate should theoretically become smaller at higher frequencies for a uniform duct; this is opposite to what is observed (see Section 20.9.5.2), indicating that horizontal inhomogeneities and small scale scatter begin to be important at microwave frequencies. The strongest coupling of energy into or out of a duct occurs when the transmitter or receiver itself lies within the duct: the ITU-R prediction method incorporates an antenna height-dependent factor to account for coupling into coastal advection ducts.

If the change in refractivity between two air masses is very abrupt, it is more appropriate to consider the radiowaves to be reflected by the layer, rather than refracted by it. A layer of thickness t can be considered to be a discontinuity if:

$$t < 14\,\lambda/\theta \tag{7.14}$$

where λ is the wavelength (in the same units as t) and θ is the angle of incidence (degrees). In this case the strength of the field reflected and transmitted by the layer can be calculated by means of Fresnel's formulae for reflection and transmission coefficients. From Eqn. 7.14 the distinction between refraction and reflection is frequency dependent, and is really a choice in the way the mechanism is modelled, rather than a difference in the mechanism itself. The layer reflection model is most useful at lower (VHF) frequencies, while refraction/ducting is a better model at UHF and above.

The meteorological mechanisms causing duct formation are now discussed. More details can be found in References [3] and [9].

7.3.2 Evaporation

A shallow surface-based duct, the evaporation duct, exists for most of the time over the sea (and other large bodies of water), caused by the very rapid decrease of water vapour pressure with height in the lowest few metres above the sea surface. The mean duct thickness varies with geographical location, ranging from 5–6 m in the North Sea through 13–14 m in the Mediterranean to over 20 m in the Gulf area. The evaporation duct is the dominant propagation mechanism for ship-based radar and communication systems. It can also have an important modifying effect at coastal sites in the presence of other surface ducts, such as the advection duct described below.

Because of its influence on naval sensor performance, much effort has gone into understanding the evaporation duct. There are well established boundary layer models [10–13] based on the similarity theory of turbulence; these enable the evaporation duct height to be estimated from a bulk measurement of sea temperature, and the air temperature, water vapour pressure and wind velocity a few metres above the surface. In open ocean conditions where the sea temperature and meteorology vary relatively slowly with range, the evaporation duct can extend for hundreds of kilometres with almost constant duct height. In coastal regions, and in enclosed areas such as the Gulf area, significant variations in duct height can occur.

Figure 7.10 shows the effect of a 15 m evaporation duct on the coverage of a radar at a height of 25 m above the sea at 3 GHz (S band) and 18 GHz (Ku band). The contours here are given in terms of propagation factor (field strength relative to free space); this is a better quantity than path loss for comparing systems operating at different frequencies. Note that the effects are stronger at the higher frequency, and that the interference lobes are trapped in the duct even though the antenna lies above the duct.

7.3.3 Nocturnal radiation

Radiative heat loss from the ground during clear, still nights produces a temperature inversion. This can result in the formation of a duct over inland or coastal regions, depending on the humidity profile. If there is sufficient water vapour present, condensation can occur, forming a radiative fog; the temperature inversion will cause an increase in water vapour pressure with height, leading to subrefraction. On the other hand, if the air is dry, the temperature inversion causes super refraction and ducting. In dry, hot climates (such as North Africa or the Middle East), radiation inversions cause severe problems to broadcast services [14].

The inversion layer weakens after sunrise, and is finally destroyed by solar heating. Hence periods of anomalous propagation associated with nocturnal radiation ducting are fairly short (one or two hours). In northern Europe this type of ducting tends to be localised, since it depends on the nature of the ground cover and of local topographical features. It is therefore less likely than other forms of ducting to cause interference over long distances. However, it can cause a severe multipath problem for low lying terrestrial links, particularly when the ducting layers are in the process of forming and breaking up.

7.3.4 Subsidence inversion

Elevated ducts can be formed by large scale atmospheric subsidence occurring during anticyclonic conditions. Large cooler air masses associated with high

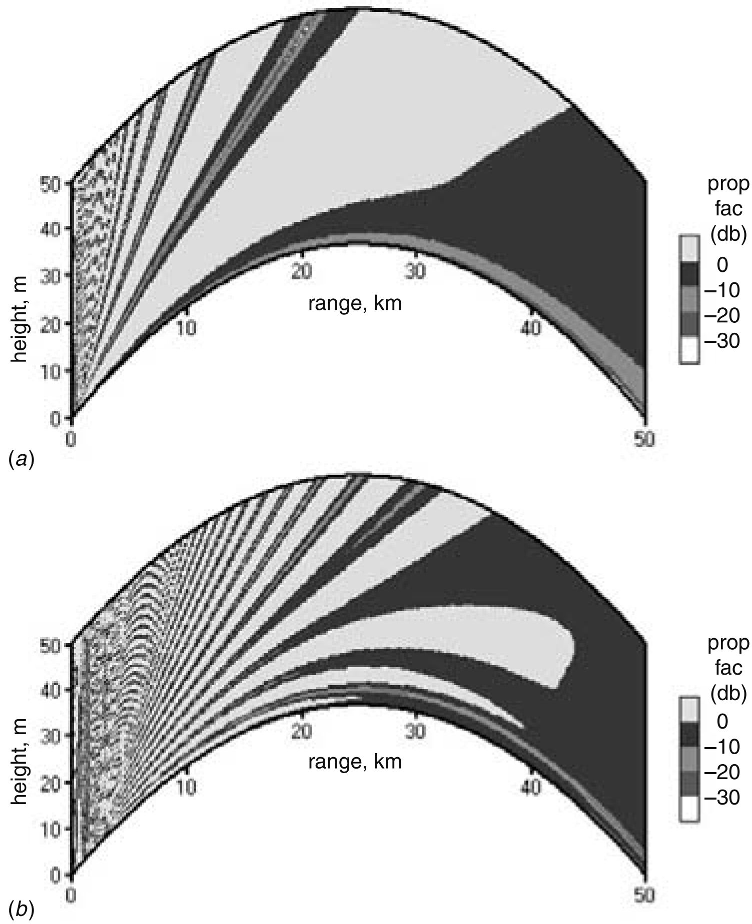


Figure 7.10 Effect of a 15 m evaporation duct on a radar at a height of 25 m over the sea
(a) 3 GHz
(b) 18 GHz

pressure systems are heated by adiabatic compression as they descend to lower levels. A strong temperature inversion forms between the descending air and the well mixed air near the surface. (The inversion layer is often visible at the ground due to the trapping of atmospheric pollutants, or the formation of a layer of cirrus cloud at the inversion.) The inversion may be accompanied by a sharp decrease in humidity. Both mechanisms tend to cause the formation of an elevated duct.

In the early stages of anticyclonic subsidence, the height of the inversion is generally too great (1–2 km or higher) to cause significant ducting. As the anticyclone evolves, the edges of the subsidence may descend close to the ground, causing ducting. There is also evidence of strong diurnal variation in the inversion height, with a lower inversion height (less than 500 m) and anomalous propagation occurring during the night. Anticyclonic subsidence can be very widespread (500–1000 km) although the part of the anticyclone contributing to ducting is unlikely to exceed a few hundred kilometres in extent. The inversion layer can be relatively homogeneous, although it tends to slope down from the centre to the edge of the subsidence dome. Figure 7.11 shows the synoptic chart and radiosonde profiles for a period of anticyclonic subsidence over northern Europe; note the sloping elevated duct.

There are other causes of subsidence that can generally be neglected because of their limited geographical occurrence or their short duration. For example, sporadic areas of subsidence at levels lower than normal may occur when a weather front interacts with an anticyclone. This process can yield thin inversion

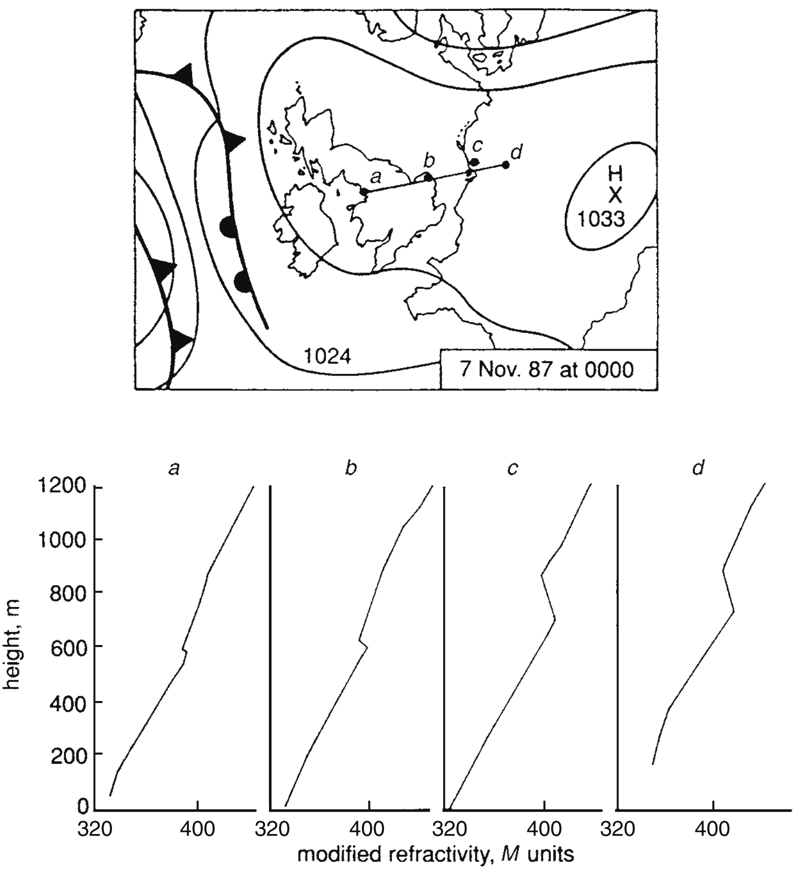


Figure 7.11 *Synoptic chart and radiosonde-derived modified refractivity profiles at the indicated locations showing a sloping elevated duct during anticyclonic subsidence*

layers with very strong gradients of humidity and temperature, and hence very severe periods of anomalous propagation. This type of subsidence is short lived and localised, and appears more like a perturbation of the general anticyclonic pattern.

7.3.5 Advection

Advection ducts are of considerable importance in coastal regions such as the North Sea or enclosed seas with adjoining hot, dry land areas such as the Mediterranean and the Gulf area. Advection is the large scale motion of air masses. A duct can be formed when a warm, well mixed air mass flows off a land surface over a cooler sea. In Northern Europe, advection is generally associated with anticyclonic subsidence: when a summer anticyclone is positioned over continental Europe, warm dry air is carried by advection from the continent out over the North Sea. The interaction of the continental air with the underlying cooler, moister air just above the sea surface results in the formation of high humidity lapse rates and temperature inversions, producing marked refractivity gradients. Advection ducts are usually surface based, typically less than 200 m thick. (Figure 7.9*b* shows an example.)

Whereas subsidence ducts are characterised by an almost trilinear shape, the typical advection duct profile is smoother. As a result, these ducts are weaker and leakier than subsidence ducts. Nevertheless they constitute a major cause of anomalous propagation in coastal areas as the surface duct couples strongly to terrestrial links and advection ducts tend to persist for several days at a time. Like subsidence, advection shows some diurnal variation, with a peak of activity expected in the early evening. Advection ducts weaken away from the coast, but can still be significant several hundred kilometres from it.

7.4 Propagation models

7.4.1 Statistical and deterministic models

Statistical models are often used for planning purposes to estimate the reliability of a system, or the level of interference to be expected to a service. They can be based on the underlying physics of the problem, but because of a shortage or absence of real-time input meteorological and other environmental data, and the need for wide applicability and short computation times, these models tend to be semi-empirical, the model parameters being deduced from experimental data.

A different requirement arises when the operating environment of a system is well characterised, the expense of obtaining real-time meteorological measurements is justified and there is a requirement to predict the performance of a system in near real-time; this may include the need for a propagation forecast a day or two ahead. The performance of military surveillance radars and communication systems also comes into this category. Here a more complete model based on physical principles is required.

We now describe the most important deterministic methods that have been used to model radiowave propagation in clear-air conditions. Statistical models of clear-air propagation are described in Chapters 19 and 20.

7.4.2 Geometrical optics

Probably the simplest conceptual model for radiowave propagation is geometrical optics. When the medium changes slowly on the scale of a wavelength, the electromagnetic field is a plane wave locally, and the energy propagates along rays which are trajectories orthogonal to the wavefronts. The rays are traced outwards from the transmitter, with the radius of curvature of the ray paths depending on the local refractive index gradients. Differential equations describing the ray paths can be derived from repeated application of Snell's law. These can be integrated numerically in general, or in the case of one-dimensional vertical refractive index profiles, can be integrated semi-analytically, yielding a computationally efficient algorithm.

Figure 7.12 shows an example of ray tracing for a transmitter at 25 m height in the presence of a simple elevated ducting layer. Ground reflected rays have been included. Note that the lower turning points of the ducted rays occur above the ground, typical of an elevated duct. There are several limitations of this naïve ray trace:

(a) when an individual ray encounters the layer it either penetrates it (such as the ground reflected rays) or is turned round by it; there is no concept of a ray splitting into two components (leakage out of the duct, partial reflection);
(b) the ray trace is frequency independent;
(c) a single ray by itself carries no amplitude information.

The field strength can be derived from geometrical optics either by explicitly calculating the cross section of a pencil of rays, or by integrating a further set of differential equations – the transport equations. To calculate the field strength at a point, it is necessary to carefully categorise the rays into separate families (each family containing rays that have followed similar trajectories), and to add coherently the contributions from all ray families passing through the point. This coherent addition requires the phase of the field, obtained from the optical path length of the ray. Two-dimensional field plots based on these methods have been generated [15].

Apart from the difficulty of classifying rays into families automatically, there is the problem that geometrical optics fails at ray caustics. A caustic is a locus of points defining the boundary of a ray family. On one side of the caustic, two rays pass through each point, while on the other side there are no rays of that family. A caustic occurs between 60 and 80 km at a height of 20 m in Figure 7.12, and there is a second caustic above this one at the boundary of the ray envelopes; the

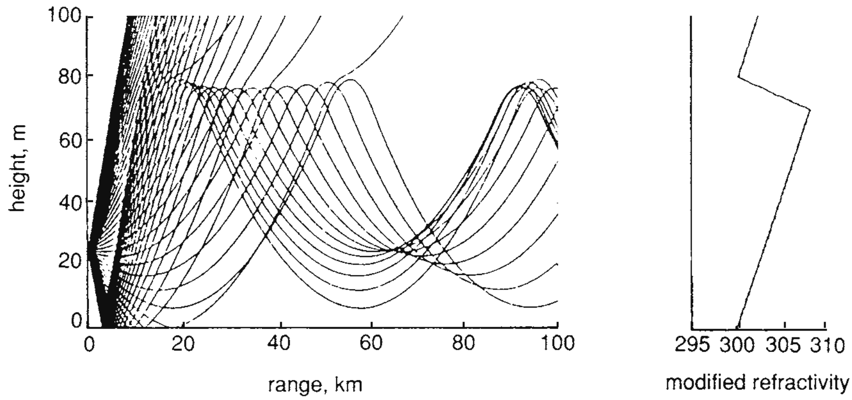


Figure 7.12 Ray trace for a transmitter at 25 m height in the elevated duct shown

two caustics meet at a point called a cusp. The ray equations predict infinite field strength at caustics and cusps, with no energy reaching the region beyond a caustic from rays in the family giving rise to it. The problem is that geometrical optics ignores the effects of diffraction. Unfortunately, these situations are likely to arise in association with the anomalous refractive index structures that give rise to ducting. At long ranges in the presence of ducts many rays may link transmitter and receiver and the number of caustics will build up with distance from the transmitter. In addition, the numerical accuracy required to calculate the optical path lengths of these long rays is difficult to achieve, even if it is assumed that the refractive index along the path is sufficiently well characterised for stable results, which it rarely is in practice. If surface reflected rays are present, as in the case of a surface duct, the problem is compounded by the difficulties of modelling the reflection, particularly if the surface is rough.

Although several sophisticated ray tracing programs are available which give good qualitative results, and quantitative results in the shorter range, line-of-sight region, geometrical optics has severe difficulties for quantitative calculations in the transhorizon region.

7.4.3 Mode theory

At ranges beyond a transmitter's radio horizon, or in regions where the assumptions of ray theory break down, field strength calculations can be made by full wave methods. These attempt to find exact solutions (in principle) to Maxwell's equations for the given refractive index profile and transmitter–receiver geometry. For real situations, of course, such a solution is computationally intractable, and approximating assumptions are inevitable.

Historically, a common approach was to impose strong constraints on the representation of the measured refractive index profile. For example, if the refractive index is assumed to vary only in the vertical direction, the wave equation can be separated and the imposition of matching conditions yields a one-dimensional equation, the mode equation. Even this one-dimensional equation is difficult to solve in general; to simplify the problem, the profile was often approximated by simple shapes (such as linear segments) to which there are standard solutions. Mode theory was first applied to tropospheric radio-wave propagation in Reference [16] and is fully described in Reference [17]; the application to elevated ducts was first given in Reference [18].

Mode theory in the troposphere is similar to mode propagation in a parallel plate waveguide. A duct of a given width will support a certain number of propagating (trapped) modes, the number being larger at higher frequencies. Higher-order modes are cut off and are evanescent. Each mode has an eigenvalue and eigenfunction: the eigenvalue defines the phase velocity and attenuation rate of the mode; the eigenfunction is a height gain function that defines the vertical distribution of energy in the mode. However, there is an important difference between tropospheric and waveguide modes: in the troposphere, at least one of the duct boundaries is soft, as it is defined by a change in the refractive index gradient, rather than by a perfectly conducting plate. The consequence is that all tropospheric modes are leaky to some extent (that is, they are attenuated as they propagate) although the attenuation rate for the lowest-order modes may be very small. A physical picture of tropospheric modes is given in Reference [19]; in particular Reference [19], Eqn. 9 shows that the total-internal-reflection criterion of geometrical optics given in Eqn. 7.13 is equivalent to the condition for a mode to be well trapped. The field at a receiver is obtained by summing the field contributions from all significant modes. Each contribution is a product of three terms: the height gain function evaluated at the transmitter height, expressing the strength of coupling of the transmitter to the mode; the height gain function evaluated at the receiver height; a term expressing the attenuation of the mode with range.

Mode theory has been successfully applied to long range propagation in the evaporation duct where there are a few, well trapped, modes and the transmitter and receiver are both strongly coupled into the duct. Mode theory becomes intractable at short range near the line of sight, and at high frequencies (or for deep ducts) where the number of trapped modes becomes very large.

Although a one-dimensional approximation may be adequate for evaporation ducts, it is inadequate for representing the meteorological structures which lead to elevated ducts. In general, horizontally homogeneous mode theory models overestimate the path loss for terminals that are located outside a duct. Inhomogeneities can give rise to higher field strengths due to mode mixing between the strongly coupled, but leaky modes and the low loss, well trapped modes. Extensions of the theory to account for mode conversion in a stepped layer model have been made, but these are complex and there seems little future for this approach in practical prediction models. Another disadvantage of mode theory is that there is no simple way of including the effects of terrain diffraction at the earth's surface.

7.4.4 Parabolic equation

A more recent alternative to mode theory is the parabolic equation (PE) model [20–24]. Parabolic equation methods are described in some detail in Chapter 10, where they are applied to terrain diffraction. However, radiowave PE methods were first developed for predicting the effects of atmospheric ducting over the sea. For this application, a two-dimensional split-step algorithm proved very effective and computationally efficient for calculating radar coverage diagrams/field strength contours on a personal computer. The method works at all ranges, thus avoiding the need for different models for the short and long distance regimes. It also works equally well with real or complex refractive index, so the effects of gaseous absorption (which can be variable from point to point) are very easy to include within the PE framework. PE and hybrid ray PE models are now incorporated in operational radar performance tools.

All the field contour plots in this Chapter were generated using the PE. Figure 7.13 shows the PE results at three frequencies (3 GHz, 9.5 GHz and 18 GHz) for the same refractive index profile as the ray trace of Figure 7.12, together with the standard atmosphere picture at 9.5 GHz for reference. In the radar's line-of-sight region the main feature is the interference lobing; the lobes are seen to be distorted significantly by the presence of the layer. (Note that the lobes become more numerous at higher frequencies; in fact the number of lobes in a sector is proportional to h/λ where h is the antenna height and λ is the wavelength.) As the frequency increases, the pictures become more optical and begin to resemble the ray trace as one would expect (compare Figures 7.13*d* and 7.12); note, however, that the full-wave PE calculation contains much information missing from the ray trace: (*a*) the PE results are quantitative, and are frequency dependent; (*b*) leakage through the top of the layer is visible at all frequencies; at 3 GHz the field is much more diffuse than the ray trace would suggest; (*c*) the PE solution is valid at caustics and cusps, giving finite results. One of the principle advantages of the PE method is its ability to handle two-dimensional refractive index data. Figure 7.14 shows the picture for the same case as Figure 7.13*c*, except that the ducting layer has been assumed to weaken with range, returning to a standard atmosphere at 100 km (such as could occur in coastal advection). The second layer bounce at 90 km no longer occurs.

7.5 Turbulent scatter

In the troposphere small-scale irregularities in the propagation medium are caused by small deviations of the temperature or humidity from the background value. The process of creating these irregularities begins with the shear forces between two moving air masses; this force creates large turbulent eddies at the outer scale of turbulence (typically 100 m). These large eddies spawn smaller and smaller eddies, transferring energy from the input range to smaller scale structures in the transformation or inertial range. In fully developed turbulence,

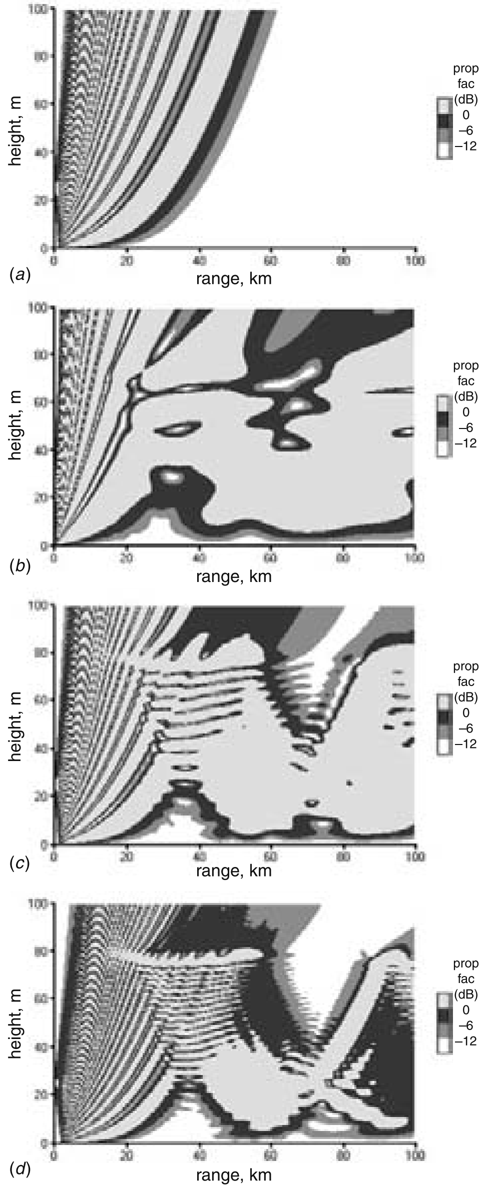


Figure 7.13 Parabolic equation derived contours for a radar at 25 m height over the sea
(a) 9.5 GHz radar in a standard atmosphere. b–d are for the same elevated duct as in Figure 7.12
(b) 3 GHz
(c) 9.5 GHz
(d) 18 GHz

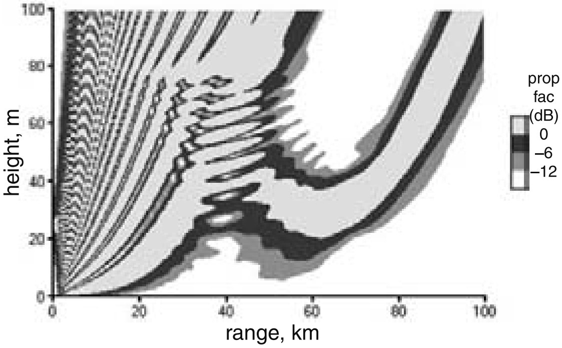


Figure 7.14 The same situation as Figure 7.13c except that the ducting layer weakens with range from the radar

this mixing continues right down to the inner scale of turbulence (about 1 mm) where the energy is dissipated; in this dissipation range, the eddies are dominated by the influence of viscosity and diffusion and cannot sustain turbulent activity. The theory of propagation in a turbulent medium is given in Reference [25]. It can be shown that in the inertial range of a turbulent atmosphere, the refractive index fluctuations have an energy proportional to the eddy size to the power of 11/6 (the Kolmogorov spectrum, Figure 7.15).

There are two ways in which turbulence affects radiowave propagation. First, within the radio beam, rapid amplitude and phase fluctuations occur, causing scintillation and wavefront distortion. These arise because the irregularities within the beam focus and defocus energy (see Figure 7.16). Fading occurs with a characteristic frequency spectrum (related to the Kolmogorov spectrum) due to the movement of these irregularities. The largest amplitude fluctuations observed in the aperture plane of an antenna are produced by eddies with scale sizes of the order of the Fresnel zone size. For antenna apertures larger than the Fresnel zone, the fluctuations are spatially averaged by the antenna: this antenna

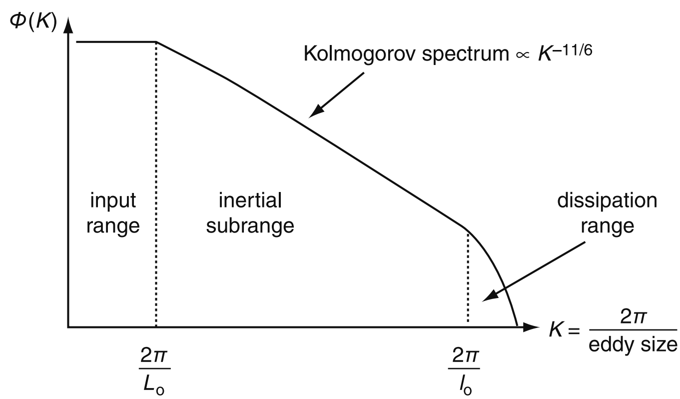


Figure 7.15 Power spectrum of refractive index fluctuations in a turbulent atmosphere

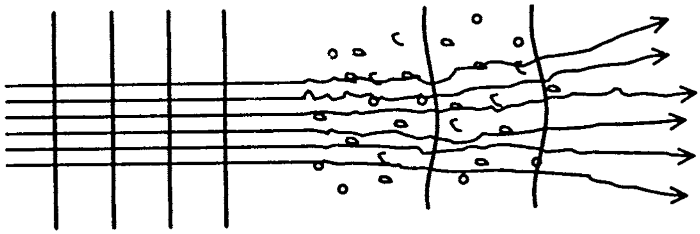


Figure 7.16 Scintillation, defocusing and wavefront distortion caused by turbulent scatter

aperture averaging results in an apparent loss of antenna gain, reducing the level of scintillation. Angle of arrival variations also occur, causing increased fading on terrestrial paths when narrow beam antennas are used. The impact of clear-air effects on earth–space propagation is discussed in Chapter 21.

The second effect of turbulence is scatter: a small portion of the incident energy is redirected and provides a weak signal outside the incident beam. Troposcatter is always present to some extent and can provide communications on beyond-the-horizon paths (Recommendation ITU-R P.617-1 [26]). Troposcatter can also cause interference to communication systems operating well beyond the horizon (see Chapter 20).

7.6 References

1 DEBYE, P.: 'Polar molecules', (Dover Publ., 1957, pp. 89–90)

2 SMITH, E.K., and WEINTRAUB, S.: 'The constants in the equation for atmospheric refractive index at radio frequencies', *Proc. IRE*, 1953, **41**, pp. 1035–1037

3 BEAN, B.R. and DUTTON, E.J.: 'Radio meteorology' (National Bureau of Standards Monograph 92, ch. 1)

4 Recommendation ITU-R P.453-8: 'The radio refractive index: its formula and refractivity data', 2001

5 HALL, M.P.M.: 'Effects of the troposphere on radio communications' (Peter Peregrinus Ltd, 1979)

6 LIEBE, H.J.: 'An updated model for millimeter wave propagation in moist air', *Radio Sci.*, 1985, **20**, pp. 1069–1089

7 YILMAZ, U.M., KENNEDY, G.R., and HALL, M.P.M.: 'A GaAs FET microwave refractometer for tropospheric studies'. AGARD CP-346, 1983, 9.1-5

8 MELFI, S.H., and WHITEMAN, D.: 'Observation of lower-atmospheric moisture structure and its evolution using a raman lidar', *Bull. Amer. Met. Soc.*, 1985, **66**, (10), pp. 1288–1292

9 COST 210 Final Report: 'Influence of the atmosphere on interference between radio communications systems at frequencies above 1GHz'. *EUR 13407*, Commission of the European Communities, 1991, catalogue no. CD-NA-13407-EM-C

10 JESKE, H.: 'Die Ausbreitung elecktromagnetischer Wellen im cm- bis

m-band über dem Meer unter besonderer Berücksichtigung der meteorologischen Bedingungen in der maritimen Grenzschicht', *Hamburger Geophysikalische Einzelschriften*, De Gruyter and Co, Hamburg, 1965
11 ROTHERHAM, S.: 'Radiowave propagation in the evaporation duct', *Marconi Review* 1974, **37**, pp. 18–40
12 PATTERSON, W.L.: 'Comparison of evaporation duct and path loss models', *Radio Science*, 1995, **20**, pp. 1061–1068
13 BATTAGLIA, M.R.: 'Modelling the radar evaporative duct'. *RANRL Technical Note* 3/85, Department of Defence, DSTO, RAN Research Laboratory, Australia, 1985.
14 BASHIR, S.O.: 'Three years statistics of refractive index gradient and effective Earth radius factor for the State of Bahrain'. *IEE Conf. Pub. 301*, (ICAP 89), 1989, Part 2, pp. 220–223
15 HERRING, R.N.: 'General raytracing techniques for predicting fields in an arbitrarily varying troposphere'. *IEE Conf. Pub. 248*, (ICAP 85), 1985, pp. 279–282
16 BOOKER, H.G., and WALKINSHAW, W.: 'The mode theory of tropospheric refraction and its relation to wave-guides and diffraction', in 'Meteorological factors in radio wave propagation'. Physical Society Report, London, 1946, 80–127
17 BUDDEN, K.G.: 'The waveguide mode theory of wave propagation' (Logos Press, 1961)
18 WAIT, J.R., and Spies, K.P.: 'Internal guiding of microwave by an elevated tropspheric layer', *Radio Sci.*, 1969, **4**, pp. 319–326
19 CRAIG, K.H.: 'Roots of the mode equation for propagation in an elevated duct'. *IEE Conf. Pub. 248*, (ICAP 85), 1085, pp. 274–278
20 KO, H.W., SARI, J.W., THOMAS, M.E., HERCHENROEDER, P.J. and MARTONE, P.J.: 'Anomalous propagation and radar coverage through inhomogeneous atmospheres'. AGARD CP-346, 1984, 25.1-14
21 CRAIG, K.H.: 'Propagation modelling in the troposphere: parabolic equation method', *Electron. Lett.*, 1988, **24**, pp. 1136–1139
22 DOCKERY, G.D.: 'Modeling electromagnetic wave propagation in the troposphere using the parabolic equation', *IEEE Trans.*, 1988, **AP-36**, pp. 1464–1470
23 CRAIG, K.H. and LEVY, M.F.: 'Parabolic equation modelling of the effects of multipath and ducting on radar systems', *IEE Proc. F*, 1991, **138**, (2), pp. 153–162
24 LEVY, M.: 'Parabolic equation methods for electromagnetic wave propagation'. IEE Electromagnetic Waves Series 45, 2000
25 TATARSKI, V.I.: 'Wave propagation in a turbulent medium' (McGraw-Hill, New York, 1961)
26 Recommendation ITU-R P.617-1: 'Propagation prediction techniques and data required for the design of trans-horizon radio-relay systems', 1992

Chapter 8

Introduction to diffraction

David Bacon

8.1 Underlying principles

Diffraction can be treated as a highly theoretical subject, although the underlying process can be visualised quite readily. It is usually interpreted in terms of the wave nature of electromagnetic radiation, although there is an argument that quantum uncertainty provides a better physical description. Whatever the case, this Chapter will use the concept of radiowaves to describe the cause and effects of diffraction.

8.1.1 Signal variations at the edge of a radio shadow

Radio energy does not simply travel in straight lines. As an example, Figure 8.1 shows the variation of illumination at the edge of a radio shadow. On the left a plane radio wavefront travelling to the right is obstructed up to a certain height by a thin totally absorbing obstruction. The graph on the right shows the variation of signal level in dB (horizontal scale) against height (vertical scale) at a point beyond the obstruction.

In Figure 8.1 the dB scale is relative to the signal level that would exist if the obstruction were absent; it is clearly not the sharp-edged shadow we are used to with light. Just above the obstruction the signal level oscillates; at the edge of the physical obstruction the signal is at –6 dB, not the half-power value of –3 dB as might be assumed, and the signal level does not immediately fall to zero behind the obstruction.

However, exactly the same pattern does in fact exist with light, it is just too small to see under normal circumstances. Due to the much longer wavelengths, the patterns which exist at radio frequencies can be large and highly significant.

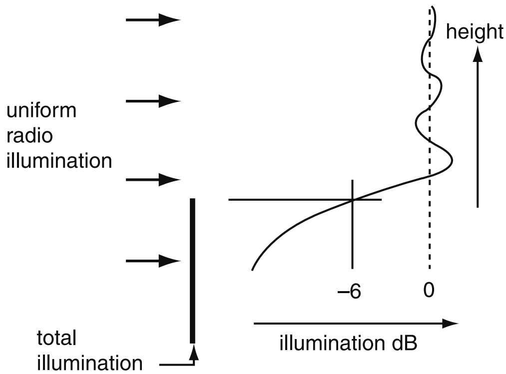


Figure 8.1 Variation of signal strength at the edge of a radio shadow

8.1.2 Huygens' construction

An early insight into the underlying mechanism comes from Huygens' construction, which was devised in the 17th century to predict the successive positions of an advancing wavefront. This is illustrated in Figure 8.2 where the wavefront is advancing from the left, with each point on the advancing wavefront radiating a spherical wave. The envelope of these waves, sometimes referred to as Huygens' wavelets, forms the new wavefront.

Huygens' wavelets suggest that there is an apparent sideways redistribution of energy as a wavefront advances. This leads to a useful description of diffraction known as the Cornu spiral.

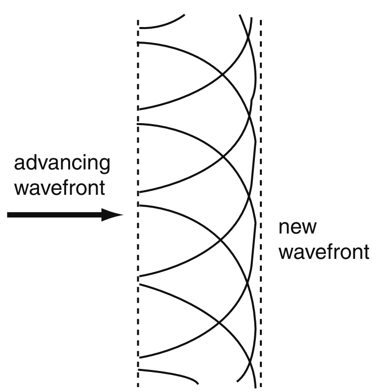


Figure 8.2 Huygens' construction

8.1.3 Fresnel knife-edge diffraction and the Cornu spiral

The diffraction caused by a thin screen obstructing part of an advancing wavefront is referred to as knife-edge diffraction. One of the simplest diffraction analysis methods uses Fresnel diffraction, which is based on a set of assumptions. Some of these concern geometry as illustrated in Figure 8.3. This shows a thin knife-edge obstacle between transmitter T and receiver R which has its edge at height h above the line of sight from T to R. Note that h can be negative. Fresnel diffraction theory applies when $h \ll d_1$, $h \ll d_2$, $\lambda \ll d_1$ and $\lambda \ll d_2$.

The Fresnel method integrates contributions from an advancing wavefront taking into account only the phase differences arising from the differences in path length, for example, the differences between the slope and direct distances s and d, respectively, as shown in Figure 8.3. When this is done for a uniformly illuminated plane wavefront the resulting phasor diagram can be represented by a double spiral known as the Cornu spiral. This is illustrated in Figure 8.4.

On the left of Figure 8.4 five rays from wavefront AA′ converge at point B. These originate from equally spaced points, and thus they represent equal magnitudes. On the right the resulting phasors are summed graphically by placing them in tandem. The phasor from the central point 0 is arbitrarily oriented parallel to the real axis. The phasors from the other points will be at phase angles determined by the additional path lengths compared to the central point. Since $h \ll d$ for each point, the additional path length can be approximated by $h^2/2d$, the angle of each phasor will be proportional to the square of h, and the two arms of the phasor summation for the upper and lower parts of the wavefront will thus curve into spirals. This leads to the shape shown in Figure 8.5.

A small point is noted here which sometimes causes confusion. Phasors are considered to rotate anticlockwise, and the delayed contributions in Figure 8.4 are oriented progressively in this direction on the grounds that they have further

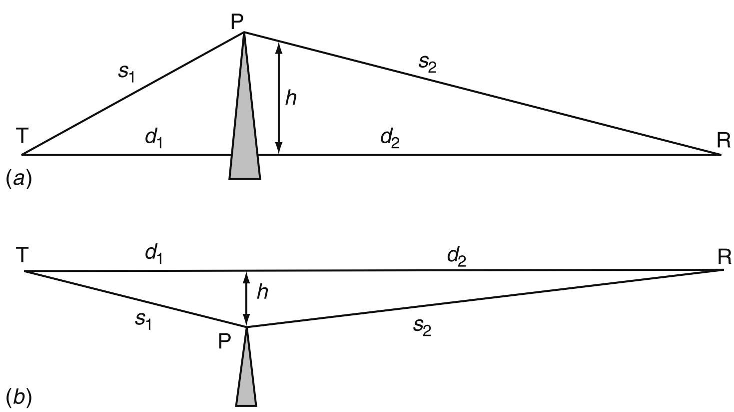


Figure 8.3 Geometry of knife-edge Fresnel diffraction
(a) h positive
(b) h negative

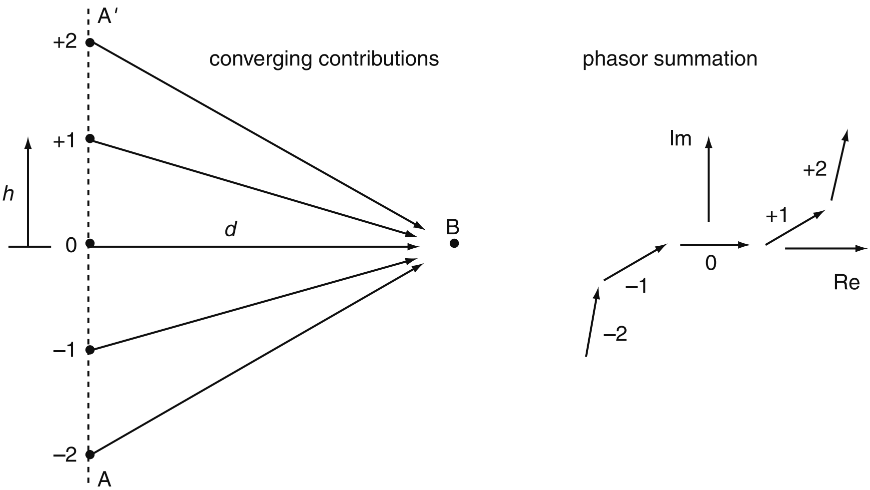


Figure 8.4 Phasor summation of contributions from a wavefront

to travel and thus arrive late. However, phasors can be validly combined only if they coincide in space and time. The contributions arriving from longer distances must in fact be those which started earlier, and should thus be rotated clockwise.

However, the Cornu spiral is nearly always shown as in Figure 8.5, and the associated Fresnel integral is shown converging towards $0.5 + j0.5$, as in Figure 8.7. Providing that either form is used systematically, the same results are obtained for knife-edge diffraction.

If the summation is taken far enough, each end of the spiral will close in on the crosses. Thus the resulting signal at point B in Figure 8.4 is the phasor extending between these points.

If a thin obstruction progressively removes contributions from one end of the spiral, it can be seen that the net signal amplitude will initially oscillate, and then

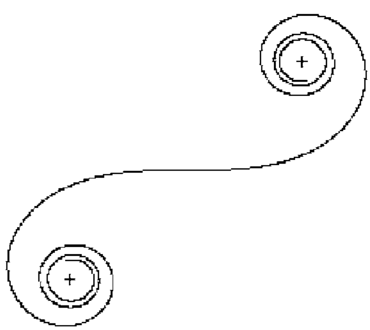


Figure 8.5 The Cornu spiral

continuously decrease in amplitude as the mid point is passed and the second half of the spiral is removed.

8.2 Mathematical formulation for knife-edge diffraction

An evaluation of the Cornu spiral can be based on the integration process illustrated in Figure 8.6.

The left side of Figure 8.6 shows the geometry for one contribution from the wavefront at height h above the direct line, to be integrated at distance d along the direct line. Noting that $h << d$, the additional path length Δd is given by the parabolic approximation as:

$$\Delta d = \frac{h^2}{2 \cdot d} \tag{8.1}$$

The right side of Figure 8.6 shows the element of length δs along the Cornu spiral with components $\delta s \cdot \cos(\phi)$ and $\delta s \cdot \sin(\phi)$ parallel with the real and imaginary axes, respectively, where the phase angle ϕ is given by:

$$\phi = \frac{2\pi \Delta d}{\lambda} = \frac{\pi h^2}{\lambda d} = \frac{\pi v^2}{2} \tag{8.2}$$

where

$$v = 2\sqrt{\frac{\Delta d}{\lambda}} \tag{8.2a}$$

This form of the integral results in the dimensionless parameter v, which is purely a function of the geometry, being the curved length along the spiral from its centre. The corresponding complex spiral itself is given by the integral:

$$F(v) = \int_0^v \exp(j \cdot \frac{\pi s^2}{2}) \cdot ds \quad \text{or} \quad F(v) = C(v) + j \cdot S(v) \tag{8.3}$$

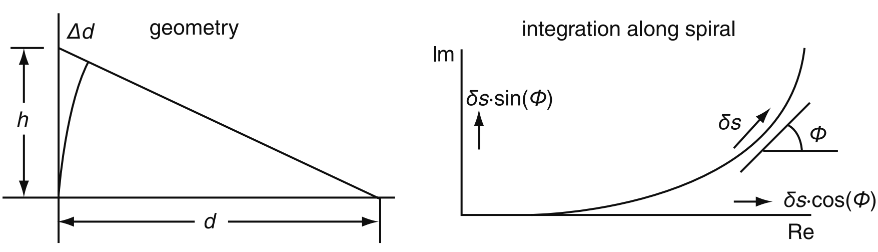


Figure 8.6 Integration along the Cornu spiral

where

$$C(v) = \int_0^v \cos(\frac{\pi s^2}{2}) \cdot ds \quad \text{and} \quad S(v) = \int_0^v \sin(\frac{\pi s^2}{2}) \cdot ds \qquad (8.3a)$$

and where it is useful to note that $C(v)$ and $S(v)$ converge upon 0.5 as v approaches infinity.

Figure 8.7 shows the functions $C(v)$ and $S(v)$ plotted on the real and imaginary axes to give a quantitative view of the positive half of the Cornu spiral for v up to 80, with the lengths $v = 1$, 2 and 3 marked on the spiral, and with the spiral converging upon $0.5 + j \cdot 0.5$.

The above formulation of the Cornu spiral can be used to calculate the normalised complex field in front of a plane wavefront partially obstructed by a knife edge. The geometry is illustrated in Figure 8.8, where E_0 is the phasor representing the unobstructed field, or free-space propagation. Since the ends of the spiral are at:

$$-0.5 - j.0.5 \text{ (lower end)}$$
$$+0.5 + j.0.5 \text{ (upper end)}$$

the complex value of the unobstructed phasor is given by:

$$E_0 = 1 + j$$

If the wavefront is obstructed from the lower end up to v, E represents the obstructed field. E can be expressed immediately as an infinite integral, but since

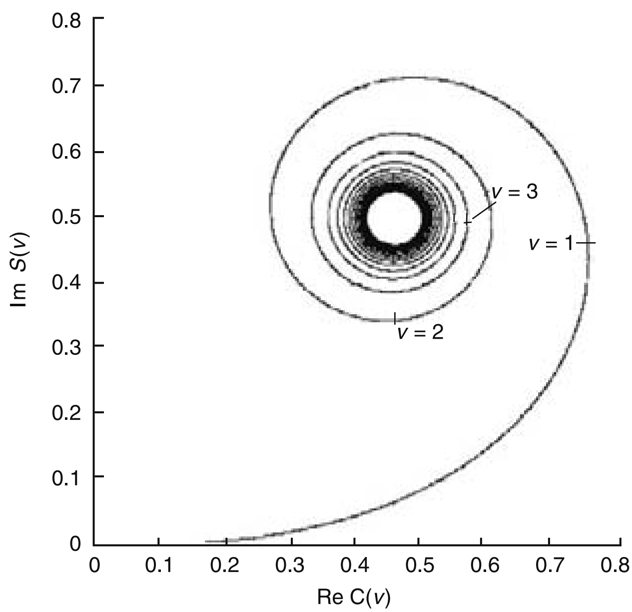


Figure 8.7 Plot of C(v) vs S(v) forming the Cornu spiral

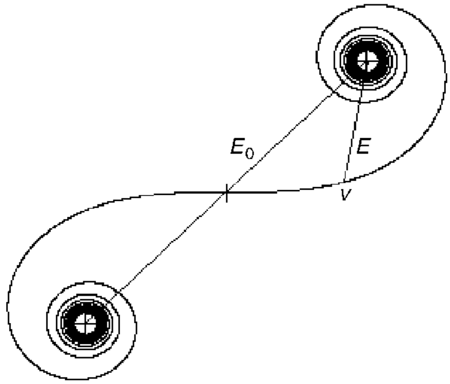


Figure 8.8 Normalised complex field from the Cornu spiral

its far end is known to be at $0.5 + j0.5$, it can be transformed into an expression in terms of the finite integrals $C(v)$ and $S(v)$:

$$E = \int_v^\infty \exp\left(\frac{j\pi s^2}{2}\right)ds = (0.5 + j0.5) - \int_0^v \exp\left(\frac{j\pi s^2}{2}\right)ds \tag{8.4}$$

$$= [0.5 - C(v)] + j\,[0.5 - S(v)]$$

Thus the obstructed field relative to free-space E_f is given by:

$$E_f = \frac{E}{E_0} = \frac{[1 - C(v) - S(v)] + j[C(v) - S(v)]}{2} \tag{8.5}$$

Figure 8.9 shows the amplitude of E_f in dB plotted against v.

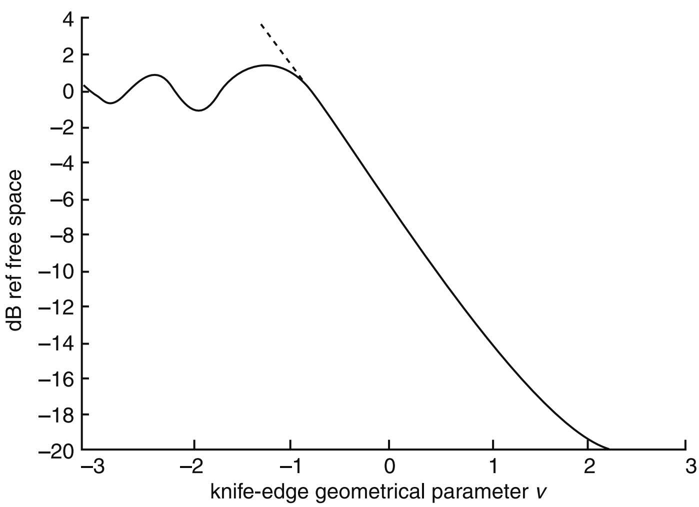


Figure 8.9 Field strength in dB relative to free space as function of v (v is defined in Eqn. 8.2a)

The dashed curve in Figure 8.9 is the approximation to knife-edge diffraction loss given in Recommendation ITU-R P.526 for $v > -0.7$ as:

$$J(v) = 6.9 + 20\log[\sqrt{(v - 0.1)^2 + 1} + v - 0.1] \tag{8.6}$$

As can be seen, $J(v)$ fits the explicit expression well for $v > -0.7$ but gives rapidly increasing errors below this value. A convenient way to use this approximation is to implement $J(v)$ for any value of v, and to limit the result such that it does not exceed zero. This gives a complete function for knife-edge diffraction loss which does not implement the ringing for line-of-sight situations, but is otherwise a good approximation.

8.3 Application of knife-edge diffraction

8.3.1 Diffraction loss

Despite being an approximation, knife-edge diffraction using the methods described above is widely used in practical situations. It can provide useful accuracy for diffraction due to terrain obstruction, despite the substantial differences between a theoretical knife edge and actual hilltops.

For instance, if a radio path is obstructed by a well defined hill the geometry of Figure 8.3 can be used where the point P represents the top of the hill, and the height h is measured from the straight line drawn through the transmitting and receiving antennas T and R, respectively, taking earth curvature into account. The value of v is given by:

$$v = h\sqrt{\frac{2}{\lambda}\left(\frac{1}{d_1} + \frac{1}{d_2}\right)} \tag{8.7}$$

Eqn. 8.6 can then be used to obtain a good approximation to the diffraction loss. Further details of this type of approach to diffraction are given by the ITU in Recommendation ITU-R P.526.

A particular limitation concerns diffraction over successive obstructions. Fresnel diffraction starts with a uniformly illuminated wavefront. It is obvious from the amplitude and phase characteristics of Eqn. 8.5 that the wavefront in the shadow of the first obstruction will be nonuniform. Thus in general Fresnel knife-edge diffraction cannot be simply cascaded.

A number of authors have developed geometrical methods for modelling cascaded terrain obstructions as knife edges. Possibly the best known of this class of propagation model is Deygout's construction [1] illustrated in Figure 8.10.

Deygout's method is based upon finding the point on the profile which, treated as a single knife-edge obstruction for the whole path (ignoring all other points) gives the highest value of v. This is the principal point, and in Figure 8.10 is point B. The corresponding diffraction loss is calculated for T–B–R. The path is then divided into two parts, one on each side of the principal point, and the process is repeated. Assuming that the secondary principal points are at A

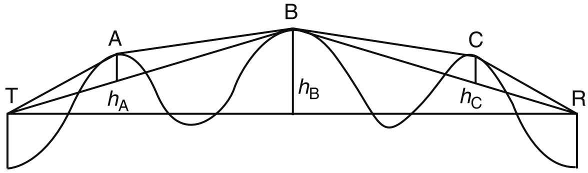


Figure 8.10 Deygout's knife-edge method for multiple obstructions

and C, diffraction losses are calculated for T–A–B and B–C–R and added to the total. This process is recursive, and can be continued until there are no further significant points. In practice it is normal to limit the process using a suitable criterion.

A number of other methods based on a geometrical construction for multiple knife-edge modelling of terrain have been published (e.g., Epstein & Peterson [2] and Giovanelli [3]). They are all subject to the problem that irregular terrain is not always suitable to be modelled as a series of knife edges. The Epstein–Peterson method has been extended by Sharples & Mehler [4] to model hill tops as horizontal cylinders.

A further difficulty is that a hill top may have closely-spaced multiple peaks; it can make a large difference to the result if such a hill is treated as one or several knife edges or cylinders. Several authors have published corrections for closely-spaced hill tops, but the issue remains something of a problem. Moreover, in computer implementations it is normal to use a terrain profile consisting of height samples at discrete points along the radio path. The results of geometrical construction methods can be sensitive to profile point spacing.

Despite all of these difficulties, edge-based methods are computationally simple, can give adequate results under the right circumstances and are widely used.

8.3.2 Fresnel clearance

Point-to-point links are designed to operate over line-of-sight paths. From the foregoing it is clear that optical line of sight alone is insufficient clearance to avoid diffraction losses.

The clearance along a radio path may be tested by constructing a Fresnel zone, which is the locus of all points having the same additional path length compared to the straight line between the antennas. By convention the *N*th Fresnel zone has $N \cdot \lambda/2$ additional path length.

Strictly, a Fresnel zone is an ellipsoid of revolution with the two antennas at the foci. Using the normal approximation for small path differences the radius of the *N*th zone at distance d_1 from the transmitter and d_2 from the receiver is given by:

$$F_N = \sqrt{\frac{N\lambda d_1 d_2}{d_1 + d_2}} \tag{8.8}$$

as illustrated in Figure 8.11.

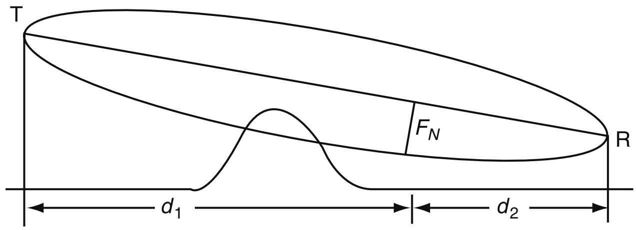


Figure 8.11 Fresnel clearance around the direct ray

A common criterion for point-to-point links is that 0.6 of the first Fresnel zone should be unobstructed. This implies a maximum value of v of about –0.85. Where it is necessary to take atmospheric refraction into account, a conservative approach is to use the earth-curvature factor k exceeded for about 99 per cent of time, which is typically also about 0.6.

8.4 Ray-based diffraction methods

Ray-based diffraction methods include the geometrical theory of diffraction (GTD) and the uniform theory of diffraction (UTD). These are based on mathematical formulations relating to rays, and can be applied in a much wider range of situations than knife-edge Fresnel diffraction. Unlike Fresnel knife-edge diffraction, ray-based methods take account of polarisation.

The principles of ray-based methods are illustrated in Figure 8.12 for a radiowave radiating from source S passing over a wedge-shaped obstacle.

For a given wedge and position of the source S, it is possible to define two boundaries:

(i) shadow boundary B_s below which a direct ray is not available;
(ii) reflection boundary B_r below which a reflected ray is not available.

The field at field point P is obtained by phasor addition of the contributions from the appropriate rays. The direct and reflected rays are included if they exist in the optical sense, even though neither will have full Fresnel clearance at the radio frequency when P is close to a boundary. A complex diffraction coefficient

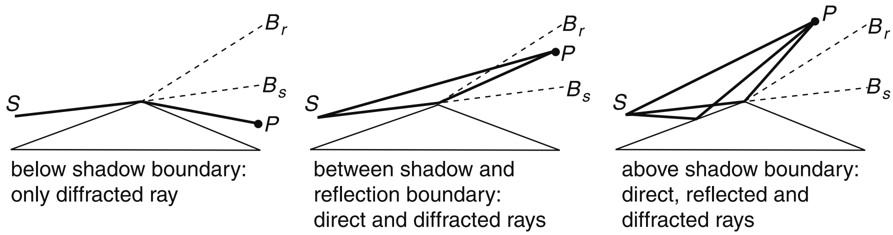


Figure 8.12 Ray sets available for positions of field point P

applied to the diffracted ray accounts for all diffraction effects.

Thus the direct and reflected rays, when they exist, can be treated simply. The essence of ray-based methods lies in the diffracted ray and the diffraction coefficient. Since the overall result must be continuous, the diffraction coefficient must be discontinuous at each boundary to compensate for the addition or removal of the direct or reflected ray.

8.4.1 GTD/UTD in two dimensions

Figure 8.13 shows the geometry for a GTD or UTD calculation for propagation over a general wedge-shaped obstacle in the plane normal to the diffracting edge E. The diffracted ray travels distance s_1 from the source S to E, and then a further distance s_2 to the field point P. The wedge faces are usually referred to as the 0 and n faces, and the external angle of the wedge ϕ_n is given by n, where $\phi_n = n.\pi$. The ray angles ϕ_1 and ϕ_2 are also measured externally from the 0 face.

The basic GTD/UTD formulation for the complex electric field phasor at P is given by:

$$e = e_0 \frac{\exp(-jks_1)}{s_1} \begin{bmatrix} D_{per} \\ D_{par} \end{bmatrix} \frac{\exp(-jks_2)}{\sqrt{(s_1+s_2)s_2/s_1}} \tag{8.9}$$

where:

e_0 = normalising amplitude
k = wavenumber = $2\pi/\lambda$
D_{per} = Diffraction coefficient for perpendicular polarisation
D_{par} = Diffraction coefficient for parallel polarisation

The second and fourth (last) terms in Eqn. 8.9 represent free-space propagation of a spherical wave from S to E, and a cylindrical wave from E to P, respectively.

The GTD form of the diffraction coefficient [5] approximates the required integral, and produces errors when the field point is close to a boundary.

8.4.2 A specific UTD formulation

The UTD form [6, 7], includes Fresnel integrals in the diffraction coefficient and gives accurate results arbitrarily close to the boundaries. Luebbers' diffraction coefficient is given by:

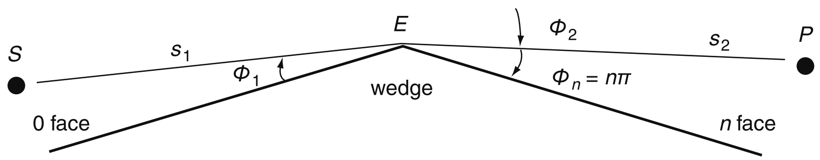


Figure 8.13 GTD geometry for wedge diffraction

$$D = \frac{-\exp(-j\pi/4)}{2n\sqrt{2\pi k}} \begin{bmatrix} \cot\left(\dfrac{\pi + (\phi_2 - \phi_1)}{2n}\right) F(kLa^{+}(\phi_2 - \phi_1)) \\ + \cot\left(\dfrac{\pi - (\phi_2 - \phi_1)}{2n}\right) F(kLa^{-}(\phi_2 - \phi_1)) \\ + R_0 \cot\left(\dfrac{\pi - (\phi_2 + \phi_1)}{2n}\right) F(kLa^{+}(\phi_2 + \phi_1)) \\ + R_n \cot\left(\dfrac{\pi + (\phi_2 + \phi_1)}{2n}\right) F(kLa^{-}(\phi_2 + \phi_1)) \end{bmatrix} \tag{8.10}$$

where:

$$L = \frac{s_1 s_2}{s_1 + s_2} \tag{8.10a}$$

$$a \pm (\beta) = 2\cos^2\left(\frac{2n\pi N^{\pm} - \beta}{2}\right) \tag{8.10b}$$

$$\beta = \phi_1 \pm \phi_2 \tag{8.10c}$$

$N^{\pm}$ $= \text{round}\left(\dfrac{\beta \pm \pi}{2\text{–}n\pi}\right)$, where round selects the nearest integer to the argument (8.10*d*)

R_0, R_n = reflection coefficients of the 0 and n faces, respectively, taking polarisation into account

$F(x)$ = a Fresnel integral in the form:

$$F(x) = 2j\sqrt{x}\,\exp(jx) \int_{\sqrt{x}}^{\infty} \exp(-jt^2)\,.\,dt \tag{8.10e}$$

In this form of the Fresnel integral the ends of the associated Cornu spiral are at $\pm\sqrt{\frac{\pi}{8}}(1 - j)$. Thus Eqn. (8.10*e*) can be rewritten to avoid integration to infinity as:

$$F(x) = 2j\sqrt{x}\,\exp(jx)\left\{\sqrt{\frac{\pi}{8}}(1 - j) - \int_{0}^{\sqrt{x}} \exp(-jt^2)\,.\,dt\right\} \tag{8.10f}$$

Alternatively a useful approximation is given by Boersma [8, 9]:

$$\int_{\sqrt{x}}^{\infty} \exp(-jt^2)\,.\,dt = \sqrt{\frac{\pi}{2}}A(x) \tag{8.11}$$

where:

$$A(x) = \begin{cases} \dfrac{1-j}{2} - \exp(-jx)\sqrt{\dfrac{x}{4}}\displaystyle\sum_{n=0}^{11}\left[(a_n + jb_n)\left(\frac{x}{4}\right)^n\right] & \text{if } x > 4 \\ -\exp(-jx)\sqrt{\dfrac{4}{x}}\displaystyle\sum_{n=0}^{11}\left[(c_n + jd_n)\left(\frac{4}{x}\right)^n\right] & \text{otherwise} \end{cases} \quad (8.11a)$$

and the coefficients a, b, c, d have the values given in Table 8.1.

Table 8.1 Coefficients for Boersma's approximation

	Coefficients			
Subscript	*a*	*b*	*c*	*d*
0	+1.595769140	−0.000000033	+0.000000000	+0.199471140
1	−0.000001702	+4.255387524	−0.024933975	+0.000000023
2	−6.808568854	−0.000092810	+0.000003936	−0.009351341
3	−0.000576361	−7.780020400	+0.005770956	+0.000023006
4	+6.920691902	−0.009520895	+0.000689892	+0.004851466
5	−0.016898657	+5.075161298	−0.009497136	+0.001903218
6	−3.050485660	−0.138341947	+0.011948809	−0.017122914
7	−0.075752419	−1.363729124	−0.006748873	+0.029064067
8	+0.850663781	−0.403349276	+0.000246420	−0.027928955
9	−0.025639041	+0.702222016	+0.002102967	+0.016497308
10	−0.150230960	−0.216195929	−0.001217930	−0.005598515
11	+0.034404779	+0.019547031	+0.000233939	+0.000838386

It should be noted that the cotangent terms in Eqn. 8.10 result in mathematical singularities exactly at each boundary, and that these points are not computable. The advance which UTD makes over GTD is that the diffraction coefficient is accurate arbitrarily close to these points. Just how close this can be depends upon the numerical precision used in calculation; in practical terms this need not be a limitation.

8.4.3 Sample UTD results

Figure 8.14 shows results obtained using the above method for a 90° perfectly-reflecting wedge where the source is kept stationary at an incidence angle of 45° and the field point is rotated from 90° to 240° from the incidence edge. This geometry permits the field point to pass through both the reflection and shadow boundaries, B_r and B_s, respectively.

The results show complicated lobing at angles less than the reflection boundary due to the combination of direct, reflected and diffracted rays. As the field point passes through the reflection boundary the lobing reduces to a

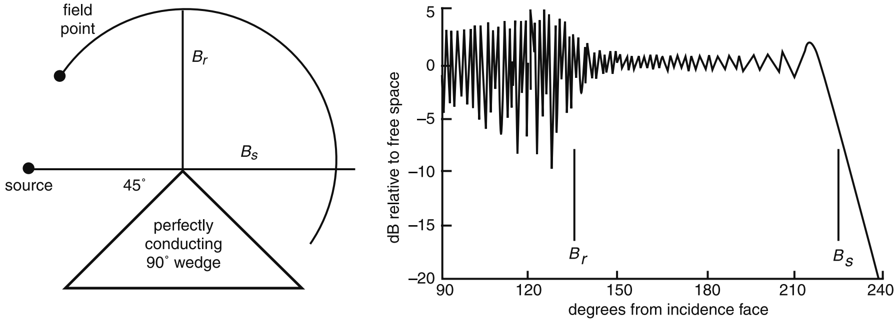


Figure 8.14 UTD results for a 90° perfectly-reflecting wedge

lower-amplitude oscillation due to the interaction between the direct and diffracted rays. As the field point passes through the shadow boundary the results are similar to those for knife-edge diffraction.

In general, UTD results are polarisation dependent. In Figure 8.14, however, it has been assumed that the wedge is perfectly conducting, and the same results will be obtained for any polarisation.

For a 0° perfectly-absorbing wedge the above UTD formulation gives the same results as Fresnel knife-edge diffraction.

8.4.4 Diffraction in three dimensions

The description of ray-based methods given above applies to the special case where the incident rays impinge normally on the diffracting edge. In general, the geometry will be as shown in Figure 8.15 where the incident rays meet the edge at angle θ. The possible diffracted rays leave the edge in a cone of internal half-angle θ lying on the far side of the plane normal to the edge. Sections 8.4.1 and 8.4.2 are specialised for $\theta = 90°$.

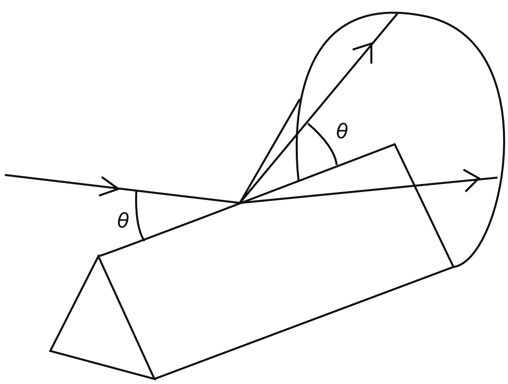


Figure 8.15 Edge diffraction in three dimensions

8.4.5 Ray-tracing methods

The UTD method is particularly useful for urban and indoor propagation models which must take account of reflection and diffraction through large angles in a three-dimensional environment. A great advantage of this approach is that combinations of reflection and diffraction in the same ray can be accommodated. However, although the UTD and reflection calculations are relatively simple, identifying the significant rays can be computationally intensive.

A particular point to note concerns the method used to combine different rays. It is tempting to think that phasor addition, i.e., taking account of amplitude and phase, will reproduce the fine structure of the multipath field. In practice this is only realistic if the topographic data (buildings shape and size) and the electrical properties of the diffracting and reflecting surfaces are known very accurately. Under normal conditions, where data on the local environment tends to be something of a simplification or approximation, it is normal to perform the final combination of rays by power summation. This approximates well to a prediction of the field strength with multipath fading averaged.

8.4.6 Further reading

GTD and UTD methods are not limited to the wedge, although this is particularly useful for radiowave propagation. Other shapes have received attention, including cylinders [3]. The UTD approach to diffraction is applicable to a wide range of situations, and has received considerable attention [6, 10–12].

8.5 References

1 DEYGOUT, J.: 'Multiple diffraction of microwaves', *IEEE Trans*; 1966, **AP 14**, pp. 480
2 EPSTEIN, J., and PETERSON, D.W.: 'An experimental study of wave propagation at 850 MC', *Proc. Inst. Radio Eng.*, 1053, **41**(5), p.595
3 GIOVANELLI, C.L.: 'An analysis of simplified solutions for multiple-knife-edge diffraction', *IEEE Trans*, 1984, **AP-32**, p.297
4 SHARPLES, P.A. and MEHLER, M.J.: 'Cascaded cylinder model for predicting terrain diffraction loss at microwave frequencies', *IEE Proc.*, 1989, **136**, (4)
5 KELLER, J.B.: 'Geometrical theory of diffraction', *J. Op. Soc. Am.*, 1962, **32**, (2)
6 KOUYOUMJIAN, R.G., and PATHAK, P.H.: 'A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface', *Proc.IEEE*, 1974, **62**, p.1448
7 LUEBBERS, R.J.: 'Finite conductivity uniform GTD versus knife edge diffraction in prediction of propagation path loss', *IEEE Trans.*, 1984, **AP-32**, (1)
8 BOERSMA, J.: 'On a numerical method for the computation of Fresnel integrals'. Report TW2, Math.Inst., Univ. of Groningen, 1960

9 BOERSMA, J.: 'Computation of Fresnel integrals', *J. Math. Comput.*, 1960, **14**, p.380
10 VOGLER, L.E.: 'An attenuation factor for multiple knife edge diffraction', *Radio Sci*, 1982, **17**, (6)
11 DEMETRESCU, C.: 'A general solution of the second-order difference equation occurring in diffraction theory', *Radio Sci.*, 1996, **31**, (3)
12 DEMETRESCU, C., *et al.*: 'TM electromagnetic scattering by a transparent wedge with resistive faces' *IEEE Trans.* 1999, **AP-47**, (1)

Chapter 9

Short-range propagation

Les Barclay

9.1 Introduction

Short-range radio systems are used for many purposes – for telemetry, remote control and games, as well as for communications. Communications uses include cordless telephony, radio local area networks (RLANs), radio fixed access and microcellular systems.

For some of these applications very low powers are used with poor and poorly located antennas where the user only expects the range to be of the order of a few metres. For example, for remote car door locks, the user will expect to point the transmitter towards the car and probably has some understanding of the need to provide a near line-of-sight path. For such applications there seems to be very little requirement to attempt to provide good propagation models.

For RLANs and other indoor applications there will be a need for some kind of generic modelling of the effects of the room size and shape, obstructions in the room, the construction materials and the penetration through walls and floors. For high-speed data transmission it may also be necessary to model the multipath time spreads.

But for outdoor microcellular systems and similar applications it will be necessary to model propagation at distances ranging out to a kilometre or so, where the longer-range area coverage prediction methods may take over. 1 km is the dividing distance used by the ITU-R in its Recommendations between short-range and longer-range prediction methods.

However, the above considerations mainly apply to VHF and UHF systems, where although the distances are short the propagation is in the far field regime and the usual propagation techniques are applicable. Another interpretation of short range would be to consider paths within the near field regime around the transmitting antenna. This is mainly of significance at lower frequencies.

It is also useful to include a discussion of propagation in tunnels and other restricted areas.

9.2 Outdoor propagation

9.2.1 Outdoor path categories

This Section presents the methods outlined in Recommendation ITU-R P.1411, and is restricted to a discussion of methods applicable to ranges up to 1 km. The methods may extend to somewhat longer ranges, see Chapter 11.

Account should be taken of the details of the environment and of the service application. Tables 9.1 to 9.3 summarise the factors to be considered.

Various propagation environments can be envisaged for short range outdoor propagation in urban areas.

Table 9.1 Physical operating environments

Environment	*Description and propagation impairments of concern*
Urban high-rise	• urban canyon, characterised by streets lined with tall buildings of several floors each • building height makes significant contributions from propagation over rooftops unlikely • rows of tall buildings provide the possibility of long path delays • large numbers of moving vehicles in the area act as reflectors adding Doppler shift to the reflected waves
Urban/suburban low-rise	• typified by wide streets • building heights are generally less than three stories making diffraction over rooftop likely • reflections and shadowing from moving vehicles can occur • primary effects are long delays and small Doppler shifts
Residential	• single and double story dwellings • roads are generally two lanes wide with cars parked along sides • heavy to light foliage possible • motor traffic usually light
Rural	• small houses surrounded by large gardens • influence of terrain height (topography) • heavy to light foliage possible • motor traffic sometimes high

Table 9.2 Definition of cell types for mobile systems

Cell type	*Cell radius*	*Typical position of base station antenna*
Small macrocell	0.5 km to 3 km	outdoor; mounted above average rooftop level, heights of some surrounding buildings may be above base station antenna height
Microcell	100 m to 500 m	outdoor; mounted below average rooftop level
Picocell	up to 100 m	indoor or outdoor (mounted below rooftop level)

Table 9.3 Typical velocities for mobile users

Environment	*Velocity for pedestrian users*	*Velocity for vehicular users*
Urban high-rise	1.5 m s^{-1}	typical city-centre speeds around 50 km h^{-1} (14 m s^{-1})
Urban/suburban low-rise	1.5 m s^{-1}	around 50 km h^{-1} (14 m s^{-1}) motorways up to 130 km h^{-1} (36 m s^{-1})
Residential	1.5 m s^{-1}	around 50 km h^{-1} (14 m s^{-1})
Rural	1.5 m s^{-1}	80–100 km h^{-1} (22–28 m s^{-1})

- base station mounted above rooftop level serving a small macrocell; in this case propagation from the base station is mainly over the roof tops
- base station mounted below rooftop level serving micro- or picocells; in these cases, propagation is mainly within street canyons
- mobile-to-mobile links, where both ends of the link can be assumed to be below rooftop level

9.2.2 Path loss models

9.2.2.1 Line-of-sight within a street canyon

UHF

In Chapter 6, propagation for two elevated antennas above a smooth plane earth was discussed, where the resultant signal is determined from the vector addition of the direct and a ground reflected ray. At short ranges the field strength has a series of minima corresponding to phase opposition of the two components, with intermediate maxima where the field has a 20 log$d$ variation. At longer distances, beyond the final null, the distance term becomes 40 log$d$.

ITU-R Recommendation P.1411 indicates that a relationship of this kind also applies approximately for some applications operating in obstructed environments. For operation over a largely line-of-sight path within a street canyon (i.e.

when the antennas are below roof height), reflections and scatter in the cluttered environment will largely fill in the minima due to wave interference. The field strength at short ranges, with some spatial averaging, will vary roughly as 20 log$d$, as if the path were in free space, although with an equivalent radiated power dependent on the environment around the transmitting antenna.

As an approximation it may be taken that the breakpoint, where 20 log$d$ changes to 40 log$d$, will occur typically at a distance given by $8\pi h_t h_r/\lambda$

ITU-R Recommendation P.1411 gives approximate lower and upper bounds for the transmission loss variation within a street canyon, as follows:

lower bound

$$L_{LOS,l} = L_{bp} + \begin{cases} 20\log_{10}\left(\dfrac{d}{R_{bp}}\right) & \text{for } d \le R_{bp} \\ 40\log_{10}\left(\dfrac{d}{R_{bp}}\right) & \text{for } d > R_{bp} \end{cases} \tag{9.1}$$

upper bound

$$L_{LOS,u} = L_{bp} + 20 + \begin{cases} 25\log_{10}\left(\dfrac{d}{R_{bp}}\right) & \text{for } d \le R_{bp} \\ 40\log_{10}\left(\dfrac{d}{R_{bp}}\right) & \text{for } d > R_{bp} \end{cases} \tag{9.2}$$

where R_{bp} is a breakpoint distance given by:

$$R_{bp} \approx \frac{4h_b h_m}{\lambda} \tag{9.3}$$

where λ is the wavelength (m), and L_{bp} is a value for the basic transmission loss at the break point, defined as:

$$L_{bp} = \left| 20\log_{10}\left(\frac{\lambda^2}{8\pi h_b h_m}\right)\right| \tag{9.4}$$

At very short distances it may also be important to take account of the vertical radiation pattern of the transmitting antenna. Where the antenna is elevated, a close receiver will be in the lower vertical sidelobe of the radiation pattern so that the gain reduces as the distance shortens. Thus for cellular mobile and for broadcasting there may be a reduction in field strength with decreasing distance very close below the antenna.

SHF

At SHF, up to, say, 15 GHz, the method given in the preceding section may be used, but a correction should be made to allow for an increase in the effective height of the road surface (i.e. for a reduction in the effective antenna heights) due to the presence of traffic, see Tables 9.4 and 9.5.

Table 9.4 The effective height of the road h_s (heavy traffic)

Frequency (GHz)	h_b (m)	Effective height of the road h_s (m) h_m 2.7	h_m 1.6
3.35	4	1.3	**
	8	1.6	**
8.45	4	1.6	**
	8	1.6	**
15.75	4	1.4	**
	8	*	**

* breakpoint beyond 1 km
** no breakpoint exists

Table 9.5 The effective height of the road h_s (light traffic)

Frequency (GHz)	h_b (m)	Effective height of the road h_s (m) h_m 2.7	h_m 1.6
3.35	4	0.59	0.23
	8	**	**
8.45	4	*	0.43
	8	*	**
15.75	4	*	0.74
	8	*	**

* breakpoint beyond 1 km
** no measurements taken

Millimetre wavelengths
At frequencies above, say, 10 GHz, the breakpoint distance R_{bp} is far beyond the expected maximum cell radius (500 m). This means that no fourth-power law is expected in this frequency band. Hence, the power distance decay rate will nearly follow the free-space law, except that the attenuation by rain (see Chapter 12) and oxygen and water vapour (see Chapter 7) has to be added.

9.2.2.2 Models for nonline-of-sight situations

For propagation over rooftops, the multiscreen diffraction model given below is valid if the rooftops are all about the same height. Assuming the rooftop heights differ only by an amount less than the first Fresnel-zone radius, the rooftop height to use in the model is the average rooftop height. If the rooftop heights vary by much more than the first Fresnel-zone radius, a preferred method is to use the highest buildings along the path in a knife-edge diffraction calculation, as described in ITU-R Recommendation P.526, instead of the multiscreen model.

For rooftops of similar height, the loss between isotropic antennas is expressed as the sum of free space loss, L_{bf}, the diffraction loss from rooftop to street L_{rts} and the reduction due to multiple screen diffraction past rows of buildings, L_{msd}.

In this model L_{bf} and L_{rts} are independent of the base station antenna height, and L_{msd} is dependent on whether the base station antenna is at, below or above building heights:

$$L_{NLOS1} = \begin{cases} L_{bf} + L_{rts} + L_{msd} & \text{for} \quad L_{rts} + L_{msd} > 0 \\ L_{bf} & \text{for} \quad L_{rts} + L_{msd} \leq 0 \end{cases} \tag{9.5}$$

The term L_{ori} describes the coupling of the wave propagating along the multiple-screen path into the street where the mobile station is located. It takes into account the width of the street and its orientation:

$$L_{rts} = -16.9 - 10 \log_{10}(w) + 10 \log_{10}(f) + 20 \log_{10}(\Delta h_m) + L_{ori} \tag{9.6}$$

$$L_{ori} = \begin{cases} -10 + 0.345\varphi & \text{for} \quad 0° \leq \varphi < 35° \\ 2.5 + 0.075(\varphi - 35) & \text{for} \quad 35° \leq \varphi < 55° \\ 4.0 - 0.114(\varphi - 55) & \text{for} \quad 55° \leq \varphi \leq 90° \end{cases} \tag{9.7}$$

where

$$\Delta h_m = h_r - h_m \tag{9.8}$$

L_{ori} is the street orientation correction factor, which takes into account the effect of rooftop-to-street diffraction into streets that are not perpendicular to the direction of propagation.

The multiple-screen diffraction loss from the base station due to propagation past rows of buildings depends on the base station antenna height relative to the building heights and on the incidence angle. A criterion for grazing incidence is the settled field distance d_s:

$$d_s = \frac{\lambda d^2}{\Delta h_b^{\,2}} \tag{9.9}$$

where

$$\Delta h_b = h_b - h_r \tag{9.10}$$

For the calculation of L_{msd}, d_s is compared with the distance l over which the buildings extend.

Calculation of L_{msd} for $l >> d_s$

$$L_{msd} = L_{bsh} + k_a + k_d \log_{10}(d/1000) + k_f \log_{10}(f) - 9 \log_{10}(b) \tag{9.11}$$

where

$$L_{bsh} = \begin{cases} -18 \log_{10}(1 + \Delta h_b) & \text{for} \quad h_b > h_r \\ 0 & \text{for} \quad h_b \leq h_r \end{cases} \tag{9.12}$$

is a loss term that depends on the base station height,

$$k_a = \begin{cases} 54 & \text{for} \quad h_b \leq h_r \\ 54 - 0.8\Delta h_b & \text{for} \quad h_b \leq h_r \text{ and } d \geq 500 \text{ m} \\ 54 - 1.6\Delta h_b \, d/1000 & \text{for} \quad h_b \leq h_r \text{ and } d < 500 \text{ m} \end{cases} \tag{9.13}$$

$$k_d = \begin{cases} 18 & \text{for} \quad h_b > h_r \\ 18 - 15\dfrac{\Delta h_b}{h_r} & \text{for} \quad h_b \leq h_r \end{cases} \tag{9.14}$$

$$k_f = \begin{cases} 0.7\,(f/925 - 1) & \text{for medium-sized city and suburban centres with medium tree density} \\ 1.5\,(f/925 - 1) & \text{for metropolitan centres} \end{cases} \tag{9.15}$$

Calculation of L_{msd} for $l < d_s$

In this case a further distinction has to be made according to the relative heights of the base station and the rooftops.

$$L_{msd} = -10 \log_{10}(Q_M^2) \tag{9.16}$$

where:

$$Q_M = \begin{cases} 2.35\left(\dfrac{\Delta h_b}{d}\sqrt{\dfrac{b}{\lambda}}\right)^{0.9} & \text{for} \quad h_b > h_r \\ \dfrac{b}{d} & \text{for} \quad h_b \approx h_r \\ \dfrac{b}{2\pi d}\sqrt{\dfrac{\lambda}{\rho}}\left(\dfrac{1}{\theta} - \dfrac{1}{2\pi + \theta}\right) & \text{for} \quad h_b < h_r \end{cases} \tag{9.17}$$

and

$$\theta = \arctan\left(\frac{\Delta h_b}{b}\right) \tag{9.18}$$

$$\rho = \sqrt{\Delta h_b^2 + b^2} \tag{9.19}$$

b is the average building separation.

For NLOS propagation within street canyons where both antennas are below rooftop level, diffracted and reflected waves at the corners of the street crossings have to be considered. In this case:

$$L_{NLOS2} = -10 \log_{10}(10^{L_r/10} + 10^{L_d/10}) \quad \text{dB} \tag{9.20}$$

where L_r is the reflection path loss defined by:

$$L_r = -20 \log_{10}(x_1 + x_2) + x_2 x_1 \frac{f(\alpha)}{w_1 w_1} - 20 \log_{10}\left(\frac{4\pi}{\lambda}\right) \quad \text{dB} \tag{9.21}$$

where:

$$f(\alpha) = \begin{cases} -41 + 110\alpha & \text{for} \quad \alpha \le 0.33 \\ -13.94 + 28\alpha & \text{for} \quad 0.33 < \alpha \le 0.42 \\ -5.33 + 7.51\alpha & \text{for} \quad 0.42 < \alpha \le 0.71 \\ 0 & \text{for} \quad \alpha > 0.71 \end{cases} \tag{9.22}$$

α is the angle of the street corner in radians, w_1 and w_2 are the widths of the streets meeting at the corner, and L_d is the diffraction path loss defined by:

$$L_d = -10 \log_{10}[x_2 x_1 (x_1 + x_2)] + 2D_a + 0.1\left(90 - \alpha\frac{180}{\pi}\right) - 20 \log_{10}\left(\frac{4\pi}{\lambda}\right) \text{dB} \tag{9.23}$$

$$D_a \approx -\left(\frac{40}{2\pi}\right)\left[\arctan\left(\frac{x_2}{w_2}\right) + \arctan\left(\frac{x_1}{w_1}\right) - \frac{\pi}{2}\right] \tag{9.24}$$

9.2.3 Influence of vegetation

The effects of propagation through vegetation (primarily trees) are important even for outdoor short path predictions. Two major propagation mechanisms can be identified:

(i) propagation through (not around or over) trees
(ii) propagation over trees

The first mechanism predominates for geometries in which both antennas are below the tree tops and the distance through the trees is small, and the latter predominates for geometries in which one antenna is elevated above the tree tops. The attenuation is strongly affected by multipath scattering initiated by diffraction of the signal energy both over and through the tree structures.

9.2.4 Default parameters for site-general calculations

If the data on the structure of buildings and roads is unknown (site-general situations) the following default values are recommended:

h_r = 3 × {number of floors} + roof height (m)
roof height = 3 m for pitched roofs, 0 m for flat roofs,
w (street width) = $b/2$
b (building separation) = 20 m to 50 m

9.3 Indoor propagation

Much of the study of propagation in indoor environments, which range from domestic rooms and offices, to open-plan office and industrial areas, corridors and stairwells and to exhibition halls and railway stations, has been limited to the modelling of specific buildings with difficulties in extrapolation to other buildings. The topic is covered in ITU-R Recommendation P.1238.

Propagation prediction for indoor radio systems differs in some respects from that for outdoor systems. The ultimate purposes, as in outdoor systems, are to ensure efficient coverage of the required area (or to ensure a reliable path, in the case of point-to-point systems), and to avoid interference, both within the system and to other systems. However, in the indoor case, the extent of coverage is well defined by the geometry of the building, and the limits of the building itself will affect the propagation. In addition to frequency reuse on the same floor of a building, there is often a desire for frequency reuse between floors of the same building, which adds a third dimension to the interference issues. Finally, the very short range, particularly where millimetre wave frequencies are used, means that small changes in the immediate environment of the radio path may have substantial effects on the propagation characteristics.

If the specific planning of an indoor radio system were to be undertaken, detailed knowledge of the particular site would be required, e.g. geometry, materials, furniture, expected usage patterns etc. However, for initial system planning, it is necessary to estimate the number of base stations to provide coverage to distributed mobile stations within the area and to estimate potential interference to other services or between systems. For these system planning cases, models that generally represent the propagation characteristics in the environment are needed.

Table 9.6 Applications and propagation impairments

Application	Propagation impairments of concern
Voice	path loss – temporal and spatial distribution
Data (low speed)	path loss – temporal and spatial distribution multipath delay
Data (high speed)	path loss – temporal and spatial distribution multipath delay ratio of desired-to-undesired mode strengths
Paging	path loss – temporal and spatial distribution
Fax	path loss – temporal distribution
Video	path loss – temporal and spatial distribution multipath delay

9.3.1 Propagation impairments and measures of quality in indoor radio systems

Propagation impairments in an indoor radio channel are caused mainly by:

- reflection from, and diffraction around, objects (including walls and floors) within the rooms
- transmission loss through walls, floors and other obstacles
- channelling of energy, especially in corridors at high frequencies
- motion of persons and objects in the room, including possibly one or both ends of the radio link

and give rise to impairments such as:

- path loss – not only the free-space loss but additional loss due to obstacles and transmission through building materials, and possible reduction of path loss by channelling
- temporal and spatial variation of path loss
- multipath effects from reflected and diffracted components of the wave
- polarisation mismatch due to random alignment of mobile terminal.

Table 9.6 lists the most significant characteristics for typical services.

9.3.2 Indoor path loss models

The indoor transmission loss model given below assumes that the base station and portable unit are located inside the same building. The indoor radio path loss is characterised by both an average path loss and its associated shadow fading statistics. The power loss coefficients include an implicit allowance for transmission through walls and over and through obstacles, and for other loss mechanisms likely to be encountered within a single floor of a building.

The basic model has the following form:

$$L_{total} = 20 \log_{10} f + N \log_{10} d + L_f(n) - 28 \quad \text{dB} \tag{9.25}$$

where

N = distance power loss coefficient
f = frequency (MHz)
d = separation distance (m) between the base station and portable unit
L_f = floor penetration loss factor (dB)
n = number of floors between base and portable

Typical parameters, based on various measurement results, are given in Tables 9.7 and 9.8. Additional general guidelines are given at the end of the section.

It should be noted that there may be a limit on the isolation expected through multiple floors or walls. The signal may find other external paths with less total loss.

The indoor shadow fading variation has a log-normal distribution with the standard deviation values (dB) as given in Table 9.9.

Table 9.7 Floor penetration loss factors, L_f (dB)

Frequency	*Residential*	*Office*	*Commercial*
900 MHz	—	9 (1 floor) 19 (2 floors) 24 (3 floors)	—
1.8–2.0 GHz	$4n$	$15 + 4 (n - 1)$	$6 + 3 (n - 1)$
52 GHz	—	16 (1 floor)	—

n is the number of floors penetrated, ($n \geq 1$)

Table 9.8 Power loss coefficients, N, for indoor transmission loss calculation

Frequency	*Residential*	*Office*	*Commercial*
900 MHz	—	33	20
1.2–1.3 GHz	—	32	22
1.8–2.0 GHz	28	30	22
4 GHz	—	28	22
5.2 GHz	—	31	—
60 GHz[1]	—	22	17

[1] 60 GHz values assume propagation within a single room or space, and do not include any allowance for transmission through walls. Gaseous absorption around 60 GHz is also significant for distances greater than about 100 m which may influence frequency reuse distances (see Recommendation ITU-R P.676)

Table 9.9 Shadow fading statistics, standard deviation (dB)

Frequency	*Residential*	*Office*	*Commercial*
1.8–2.0 GHz	8	10	10
5.2 GHz	—	12	—

A few general conclusions can be drawn, especially for the 900–2000 MHz band:

- paths with a line-of-sight component are dominated by free-space loss with a distance variation of about 20 log$d$
- large open rooms also have a similar distance variation; this may be due to a strong line-of-sight component to most areas of the room; examples include rooms located in large retail stores, sports arenas, open-plan factories, and open-plan offices
- corridors exhibit path loss less than that of free space, with a typical distance variation of about 18 log$d$; supermarkets with long, linear aisles exhibit similar characteristics
- propagation around obstacles and through walls adds considerably to the

loss, which can increase the distance variation up to about 40 log*d* for a typical environment; examples include paths between rooms in closed-plan office buildings

- for long unobstructed paths, the first Fresnel zone breakpoint may occur, as discussed for outdoor propagation.

9.3.3 Delay spread models

9.3.3.1 Multipath

The mobile/portable radio propagation channel varies in time, frequency and with spatial displacement. Even in the static case, where the transmitter and receiver are fixed, the channel can be dynamic, since scatterers and reflectors are likely to be in motion. The time delay associated with each multipath mode is proportional to path length. (A very rough estimate of the maximum delay time to be expected in a given environment may be obtained simply from the dimensions of the room and from the fact that the time in nanoseconds for a radio pulse to travel distance d (m) is approximately $3.3d$.)

9.3.3.2 RMS delay spread

If an exponentially decaying profile can be assumed for the impulse response, its form is given by:

$$h(t) = \begin{cases} e^{-t/S} & \text{for } 0 \le t \le t_{max} \\ 0 \ldots & \end{cases} \tag{9.26}$$

where S is the RMS delay spread, t_{max} is the maximum delay and $t_{\max} >> S$.

The advantage in using the RMS delay spread as the model output parameter is that the model can be expressed simply in the form of a table. Typical delay spread parameters, estimated from averaged delay profiles, for three indoor environments are given in Table 9.10. These values are based on measurements at 1.9 and 5.2 GHz using omnidirectional antennas. (There is little evidence of a strong frequency dependence in these parameters when omnidirectional antennas are used.) In Table 9.10, the lower values are not extreme and may occur rather frequently, but the highest values will occur only rarely. The values given in the table represent the largest room sizes likely to be encountered in each environment.

Within a given building, the delay spread tends to increase as the distance between antennas increases, and hence to increase as path loss increases. With greater distances between antennas, it is more likely that the path will be obstructed, and that the received signal will consist entirely of scattered paths.

9.3.3.3 Effect of polarisation and antenna radiation pattern

In an indoor environment, there is not only a direct path but also reflected and

Table 9.10 RMS delay spread parameters

Environment	*Frequency (GHz)*	*A lower values (ns)*	*B median (ns)*	*C highest values (ns)*
Indoor residential	1.9	20	70	150
Indoor office	1.9	35	100	460
	5.2	45	75	150
Indoor commercial	1.9	55	150	500

diffracted paths between the transmitter and receiver. The reflection characteristics of a building material depend on polarisation, incidence angle and the material's complex permittivity. The angles-of-arrival of multipath components are distributed, depending on the antenna beamwidths, building structures and siting of transmitter and receiver. Therefore, polarisation and the effective antenna radiation pattern can significantly affect indoor propagation characteristics.

It is widely accepted that, in line-of-sight (LOS) channels, directional antennas reduce RMS delay spread as compared to omnidirectional antennas, and that circular polarisation (CP) reduces it compared to linear polarisation (LP). Therefore, in this case a directional CP antenna offers an effective means of reducing the delay spread.

Since multipath propagation components have an angle-of-arrival distribution, those components outside the antenna beamwidth are spatially filtered out by the use of directional antennas, so that delay spread can be reduced. Indoor propagation measurement and ray-tracing simulations performed at 60 GHz, with an omnidirectional transmitting antenna and four different types of receiving antenna (omnidirectional, wide-beam, standard horn and narrowbeam antennas) directed towards the transmitting antenna, show that the suppression of delayed components is more effective with narrower beamwidths. Table 9.11 shows examples of the antenna directivity dependence of static RMS delay spread not exceeded at the 90th percentile obtained from a ray-tracing simulations at 60 GHz for an empty office. It may be noted that some CDMA schemes utilise the multipath components to give path or frequency diversity so that a reduction in RMS delay spread may not necessarily always be desirable.

When the direct path is obstructed, the polarisation and antenna directivity dependence of delay spread may be more complicated than those in the line-of-sight path.

9.3.3.4 Effect of building materials, furnishings and furniture

Indoor propagation characteristics are affected by reflection from and transmission through the building materials. The reflection and transmission characteristics of those materials depend on the complex permittivity of the

Table 9.11 Example of antenna directivity dependence of static RMS delay spread

Frequency (GHz)	TX antenna	RX antenna beamwidth (degrees)	Static RMS delay spread (90 percentile) (ns)	Room size (empty (m)
60	omni	omni	17	13.5~7.8
		60	16	
		10	5	
		5	1	

materials. Thus, site-specific propagation prediction models need the complex permittivity of building materials as well as the building structure data as basic input data.

The complex permittivities of typical building materials, obtained experimentally, are tabulated in Table 9.12. These permittivities indicate significant difference from one material to another, although showing little frequency dependence in the frequency range 60–100 GHz, except for floorboard which varied by ten per cent. At millimetre wave bands, surface finishes such as paint must be considered as one of the dielectric layers

From the complex permittivity, η, the reflection coefficient is given by:

E-vector normal to the reflection plane

$$R_N = \frac{\sin\theta - \sqrt{\eta - \cos^2\theta}}{\sin\theta + \sqrt{\eta - \cos^2\theta}} \tag{9.27}$$

E-vector parallel to the reflection plane

$$R_P = \frac{\sin\theta - \sqrt{(\eta - \cos^2\theta)/\eta^2}}{\sin\theta + \sqrt{(\eta - \cos^2\theta/\eta^2}} \tag{9.28}$$

circular polarisation

$$R_C = \frac{R_N + R_P}{2} \tag{9.29}$$

where the reflection plane is the plane in which both the incident and reflected rays lie, and θ is the angle between the incident ray and the plane of the reflecting surface.

Specular reflections from floor materials are significantly reduced at millimetre wavelengths when materials are covered by carpets with rough surfaces. Similar reductions may occur with window coverings such as draperies. Therefore, it is expected that the particular effects of materials will be more important as frequency increases.

Table 9.12 Complex permittivity of interior construction materials

	1 Ghz	*57.5 GHz*	*78.5 GHz*	*95.9 GHz*
Concrete	7.0 – *j*0.85	6.50 – *j*0.43	—	6.20 – *j*0.34
Lightweight concrete	2.0 – *j*0.50	—	—	—
Floorboard (synthetic resin)	—	3.91 – *j*0.33	3.64 – *j*0.37	3.16 – *j*0.39
Plasterboard	—	2.25 – *j*0.03	2.37 – *j*0.10	2.25 – *j*0.06
Ceiling board (rock wool)	1.2 – *j*0.01	1.59 – *j*0.01	1.56 – *j*0.02	1.56 – *j*0.04
Glass	7.0 – *j*0.10	6.81 – *j*0.17	—	—
Fibreglass	1.2 – *j*0.10	—	—	—

In addition to the fundamental building structures, furniture and other fixtures also significantly affect indoor propagation characteristics. These may be treated as obstructions and are covered in the path loss model above.

9.3.3.5 Effect of movement of objects in the room

The movement of persons and objects within the room causes temporal variations of the indoor propagation characteristics. These variations, however, are very slow compared with the data rate likely to be used, and can therefore be treated as virtually a time-invariant random variable. Apart from people in the vicinity of the antennas or in the direct path, the movement of persons in offices and other locations in and around the building has a negligible effect on the propagation characteristics.

Measurements performed when both of the link terminals are fixed indicate that fading is bursty (statistics are very nonstationary), and is caused either by the perturbation of multipath signals in areas surrounding a given link, or by shadowing due to people passing through the link.

Measurements at 1.7 GHz indicate that a person moving into the path of a line-of-sight signal causes a 6 to 8 dB drop in received power level, and the *K*-value of the Nakagami–Rice distribution is considerably reduced. In the case of nonline-of-sight conditions, people moving near the antennas did not have any significant effects on the channel.

In the case of a handheld terminal, the proximity of the user's head and body affect the received signal level. Measurements with dipole antennas at 900 MHz showed that the received signal strength decreased by 4 to 7 dB when the terminal was held at the waist, and by 1 to 2 dB when the terminal was held against the head of the user, as compared with the received signal strength when the antenna was several wavelengths away from the body.

When the antenna height is lower than about 1 m, for example in the case of a typical desktop or laptop computer application, the line-of-sight path may be shadowed by people moving in the vicinity of the user terminal. For such data applications, both the depth and the duration of fades are of interest.

Measurements at 37 GHz in an indoor office lobby environment have shown that fades of 10 to 15 dB were observed often. The duration of these fades due to body shadowing, with people moving continuously in a random manner through the line-of-sight, follows a log-normal distribution, with the mean and standard deviation dependent on fade depth. For these measurements, at a fade depth of 10 dB, the mean duration was 0.11 s and the standard deviation was 0.47 s. At a fade depth of 15 dB, the mean duration was 0.05 s and the standard deviation was 0.15 s.

9.4 Propagation in tunnels

A tunnel may be considered as a waveguide. Dependent on the propagation mode, the critical wavelength for the guide will be between 0.82 and 1.7 times the tunnel diameter, and the maximum wavelength for propagation will be about two-thirds of this. Thus VHF will not propagate in small tunnels. Of course tunnels do not necessarily have smooth, good conducting walls, the tunnels will dip and turn and will be partially filled from time to time by vehicles or trains. Thus losses may be higher than waveguide theory would predict. In addition the coupling of the tunnel entrance to the outside environment is not likely to be optimum and it may be difficult to launch signals into the tunnel.

Some success has been obtained with very large railway tunnels by arranging UHF antennas to beam signals directly into the tunnel entrance, but more generally the provision of radio communication in tunnels will be provided by leaky feeder systems.

For emergency communications where leaky feeders are not available, communication may be achieved by making use of conductors in the tunnel and by inductive coupling. For example, the FIGARO system used frequencies of about 3 MHz with loops coupling to railway lines and other metallic conductors. This technique required skill and experience in setting up and operating but could offer the only way of getting communication in extreme situations.

9.5 Leaky feeder systems

Leaky feeder systems may be of value not only in tunnels but also in lift shafts and other similar situations. In the past they have been used to provide walkie-talkie communication in offices, around an exchange floor for example, but this kind of application is likely to diminish with the widespread use of more flexible and versatile cellular and personal communication services.

Leaky feeders are coaxial lines with controlled leakage provided by slots in the outer conductor. In the past they have also been implemented with cables with a loosely woven braid etc. Cables sold for the purpose will have a manufacturer's specification for the coupling loss. This is sometimes specified as the loss when measured with a $\lambda/2$ dipole at 6 m from the feeder.

Systems may be implemented as shown in Figure 9.1. With two frequency duplex, the first system is shown as providing a transceiver system for signalling and control purposes and a repeater system for use between portables of the railway police. The second system shows a daisy-chained installation that might be arranged around the walls of a building. Since this kind of system would be vulnerable to a single break in the feeder, a second daisy-chain may be installed alongside, running in the opposite direction.

The design equation for these systems is of the form:

$$T_x - A_l - L_c - L_a - PS_l - COMB_l - SU = R_p \tag{9.30}$$

where

T_x is the walkie-talkie output power (dBW)
A_l is the portable antenna loss
L_c is the specified cable coupling loss
L_a is the leaky cable attenuation for the greatest distance before amplification
PS_l is the power splitter loss, if appropriate
$COMB_l$ is the combiner loss
$COAX_l$ is the loss in the main feed to the base station receiver
SU is a system use margin, say 20 dB
R_p is the received power at the base station, which should be compared with the receiver sensitivity

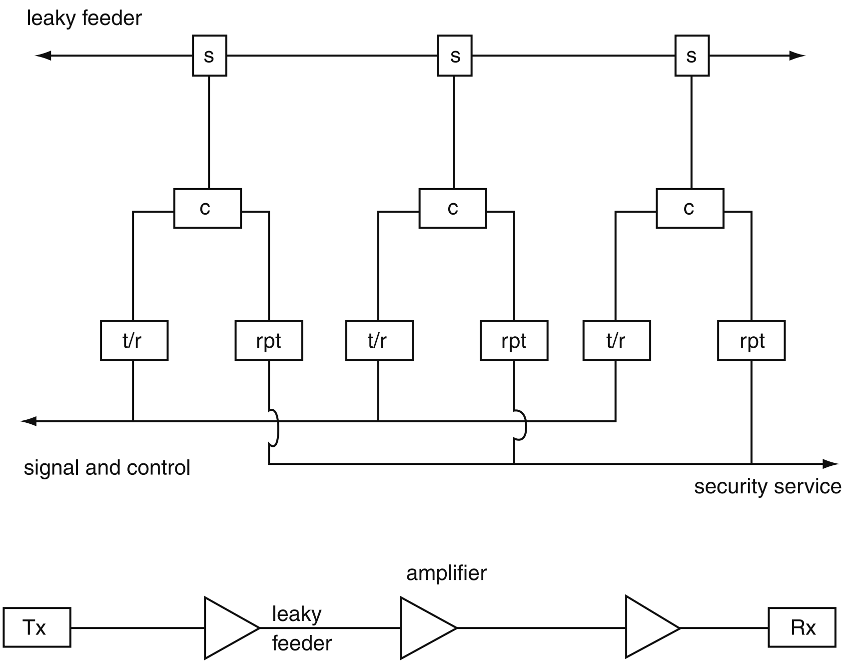


Figure 9.1 Leaky feeder systems

9.6 The near field

In earlier chapters it was indicated that, sufficiently far from the source, ray paths may be considered as parallel, and the wave front as plane; the impedance of free space is given by $\eta_0\sqrt{(\mu_0/\varepsilon_0)}$, i.e. 120 $\pi\Omega$. Near the transmitting antenna these relationships no longer apply.

The magnetic field intensity from a Hertzian dipole, carrying the current *I*, is given by:

$$H_\phi = \frac{Id\ell}{4\pi} . \beta^2 \sin\theta \left(j\frac{1}{\beta r} + \frac{1}{\beta^2 r^2} \right) e^{-j\beta r} \tag{9.31}$$

where θ is the angle from the direction of dipole, and $\beta = 2\pi/\lambda$.

The corresponding electric field intensity is:

$$E_\theta = \frac{Id\ell}{4\pi} . \eta\beta^2 \sin\theta \left(j\frac{1}{\beta r} + \frac{1}{\beta^2 r^2} - \frac{1}{\beta^3 r^3} \right) e^{-j\beta r} \tag{9.32}$$

$$E_{ri} = \frac{Id\ell}{4\pi} . \eta\beta^2 \sin\theta \left(\frac{1}{\beta^2 r^2} - j\frac{1}{\beta^3 r^3} \right) e^{-j\beta r} \tag{9.33}$$

Thus, although the far field varies as $1/r$, near to the antenna the terms in $1/r^2$ and $1/r^3$ dominate. The point where these higher order terms become insignificant is referred to as the boundary of the near and far zones. For a Hertzian dipole it is usually considered to occur where $r = \lambda/2\pi$

Where the source is a magnetic loop, there is an equivalent set of equations in which the electric field has an inverse square term and the magnetic field has both inverse square and inverse cube terms.

The wave impedance, given by the ratio H_θ/H_ϕ, is η in the far field, but in the near field it has a high impedance when the radiator is a dipole, and a low impedance with a magnetic loop

For practical purposes the boundary between the near and far field regions may be taken as the greater of 3λ and $2D^2/\lambda$ where *D* is the largest antenna dimension.

Chapter 10

Numerically intensive propagation prediction methods

C.C. Constantinou

10.1 Introduction

This Chapter aims to introduce a number of deterministic radiowave propagation prediction methods and discuss briefly the computational issues arising in their practical implementation.

Ideally, we wish to solve Maxwell's equations exactly by specifying a boundary value problem to a sufficient degree of accuracy, subject to some initial conditions. The boundary value problem in this case is the geometrical and electrical description of the radio environment down to subwavelength accuracy, whereas the initial conditions are the current distribution on the transmitting antenna.

10.1.1 Intractability of exact solutions

Let us consider the numerical complexity in attempting to solve the reduced vector wave equation, which can be derived from Maxwell's equations. Take a typical GSM900 macrocell of a 10 km radius. A stable numerical solution requires that we discretise the domain (macrocell) into a cubic grid whose size is at most $\lambda/2$ long. Taking the cell height to be 100 m so as to include the tallest possible buildings in it, the minimum required number of grid points is $\sim\pi r^2 h/(\lambda/2)^3 \sim 7 \times 10^{12}$. In the calculation, at each grid point we need to use one complex double precision variable for each Cartesian component of the electric field and one complex double precision variable for the complex dielectric constant of the medium. Since a complex double precision variable corresponds to 16 bytes of memory, the total amount of memory space required for a straightforward numerical solution to the problem is $\sim 4 \times 10^{14}$ bytes, or 400 Tbytes.

A similar consideration applied to the boundary data (i.e. the amount of information to be extracted from a building and terrain database) shows that the smallest database that can be used would have a size of at least 100 Gbytes.

Clearly, the problem of solving the full, vector wave equation exactly for meaningful problems in radiowave propagation is intractable for two reasons:

(i) the computational complexity of the problem is beyond the capability of any computer in the foreseeable future (even taking into account Moore's law);
(ii) the collection of the input data at this level of accuracy – even if automated fully – is a near practical impossibility.

10.1.2 General remarks on numerical methods

Numerically intensive radiowave propagation prediction methods share one common feature: they make judicious use of approximations and simplifications to render the problem at hand tractable. These approximations may be based on high-frequency asymptotic solutions to exact problems, *a priori* knowledge on the propagation direction, the paraxial approximation etc. Therefore, no single practical solution of Maxwell's equations exists that is applicable to all propagation problems.

10.1.3 Chapter outline

In this Chapter, we shall restrict our attention to tropospheric, clear air propagation in the VHF and UHF bands (30 MHz–3 GHz) only. Typical applications that would employ these numerically intensive methods include broadcasting, radar operation prediction and point-to-point radio link and mobile radio planning.

The deterministic propagation techniques considered here are:

(a) integral equations
(b) the parabolic equation
(c) ray tracing

In all of the above methods, the dominant consideration will be the terrain effects, whereas in the case of the parabolic equation method the effect of the atmosphere will also be included.

10.2 Integral equation methods

Here we shall consider only one integral equation formulation, based on Hufford's original work [1]. This formulation is based on the use of Green's integral theorem and approximate boundary conditions. For simplicity we shall consider only the case of an elevated Hertzian dipole source over an irregular, inhomogeneous earth.

10.2.1 Derivation of integral equation

Consider an irregular surface shown in Figure 10.1. Assuming that all fields have a exp($j\omega t$) dependence, Maxwell's equations in a homogeneous atmosphere reduce to the inhomogeneous scalar Helmholtz equation for the vertical electric and magnetic fields:

$$\nabla^2\psi + k^2\psi = -4\pi\tau \tag{10.1}$$

where ψ is the vertical component of the electric or magnetic field for a vertically or horizontally oriented Hertzian dipole, respectively, $k = 2\pi/\lambda$ is the propagation constant and τ is the source current density (zero everywhere except at the source point taken here to be the origin).

The field ψ is related to the path gain function W through:

$$\psi(R) = W(R)\frac{\exp(-jkr_0)}{r_0} \tag{10.2}$$

Furthermore, the free-space field at the receiver is given by:

$$\psi_0(R) = \mathrm{g}(R)\frac{\exp(-jkr_0)}{r_0} \tag{10.3}$$

where $g(R)$ is used to account for the transmitting antenna radiation pattern. We also define a function φ by:

$$\varphi(P) = \frac{\exp(-jkr_2)}{r_2} \tag{10.4}$$

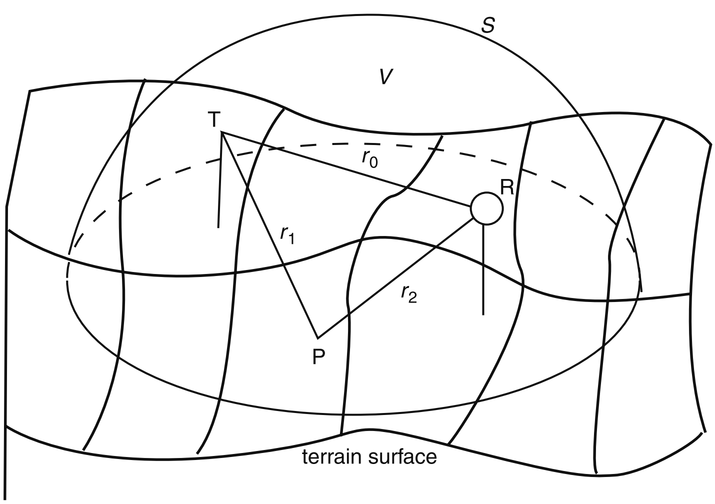


Figure 10.1 Geometry of terrain indicating the transmitter T, receiver R and point of integration P – hemispherical cap has infinite radius

related to the free-space Green's function and satisfying the homogeneous scalar Helmholtz equation:

$$\nabla^2\varphi + k^2\varphi = 0 \tag{10.5}$$

Finally, the field ψ satisfies the simple Leontovich [2] boundary condition:

$$\frac{\partial\psi}{\partial n}(P) = jk\delta\psi(P) \tag{10.6a}$$

where

$$\delta = \begin{cases} \sqrt{\varepsilon - 1}/\varepsilon & \text{for vertical polarisation} \\ \sqrt{\varepsilon - 1} & \text{for horizontal polarisation} \end{cases} \tag{10.6b}$$

and $\varepsilon = \varepsilon_r + j\sigma/\omega\varepsilon_0$ is the complex dielectric constant for the ground.

Green's theorem states that any two continuous differentiable scalar functions of position ψ and φ defined in a closed volume V bounded by a surface $S = \partial V$ satisfy:

$$\iiint_V (\psi\,\nabla^2\varphi - \varphi\,\nabla^2\psi)\,dV = \oiint_S \left(\varphi\frac{\partial\psi}{\partial n} - \psi\frac{\partial\varphi}{\partial n}\right)da \tag{10.7}$$

where $\partial/\partial n$ denotes a normal derivative. We define the volume V to be that of the hemisphere shown in Figure 10.1 minus the volume of an infinitesimally small sphere surrounding the receiving point. Using Eqns. 10.1, 10.4, 10.5 and 10.6 Green's theorem reduces to:

$$\iiint_V (\psi(-k^2\varphi) - \varphi(-k^2\psi - 4\pi\tau))dV$$

$$= \oiint_S \left(\frac{\exp(-jkr_2)}{r_2} \cdot jk\delta\psi - \psi \cdot \frac{\exp(-jkr_2)}{r_2} jk\left(-1 - \frac{1}{jkr_2}\right)\frac{\partial r_2}{\partial n}\right)da + 4\pi\psi(R) \tag{10.8}$$

where the last term arises from the surface integral around the infinitesimally small sphere centred at the receiving point. The left-hand side of Eqn. 10.8 identically simplifies to $4\pi\psi_0(R)$. Finally, using Eqns. 10.2 and 10.3 to derive an equation for the path gain function and after some rearranging Eqn. 10.8 can be reduced to:

$$W(R) = g(R) - \frac{jk}{4\pi}\iint_{\substack{\text{earth's}\\ \text{surface}}} W(P)\exp(-jk[r_1 + r_2 - r_0])\frac{r_0}{r_1 r_2}\left(\delta + \left(1 + \frac{1}{jkr_2}\right)\frac{\partial r_2}{\partial n}\right)da \tag{10.9}$$

where we have made use of the fact that as the radius of the hemisphere tends to infinity, the radiation condition ensures that we have no contribution to the integral from the hemispherical cap.

This equation is not yet amenable to direct computation because the term under the integral is oscillatory and the domain of integration is infinitely large. However, the oscillatory term $\exp(-jk[r_1 + r_2 - r_0])$ suggests that an approximation based on the method of stationary phase will significantly simplify the integral. Projecting the coordinates on the earth's surface onto a horizontal plane and introducing elliptical coordinates with foci at the transmitter and receiver, we find that the method of stationary phase enables us to reduce the two-dimensional to a one dimensional integral along the line joining the transmitter to the receiver (see Reference [1] for details).

In most practical applications the term $1/jkr_2$ is negligibly small which further simplifies the result to:

$$W(x) = g(x) - \left(\frac{jx}{\lambda}\right)^{1/2} \int_0^x \left(\delta(s) + \frac{dr_2}{dn}\right) W(s) \exp(-jk[r_1 + r_2 - r_0]) \frac{ds}{\sqrt{s(x-s)}} \quad (10.10)$$

and all the symbols are shown in Figure 10.2. This is a Volterra (or Fredholm) integral equation of the second kind.

10.2.2 Assumptions made in the derivation of the integral equation

A number of simplifying assumptions have been made implicitly in the derivation of Eqn. 10.10). These arise from the application of the Leontovich impedance boundary condition for the ground and are that the wavelength in free space, the distance to the source and the radii of curvature of the earth's surface must all be much larger than both the skin depth and the wavelength in the ground.

Further assumptions are that the two antennas must be in the far field of each

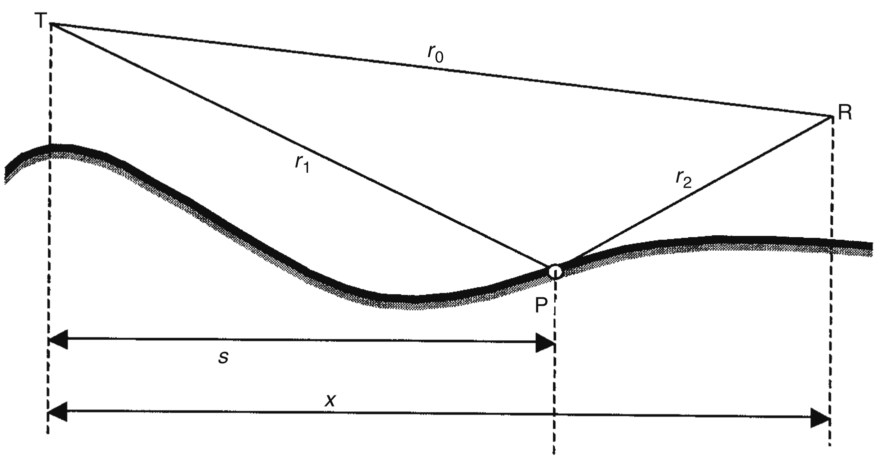


Figure 10.2 Two-dimensional path profile between the transmitter and receiver

other and that the transverse terrain variation to the path profile between the transmitter and the receiver (great circle path) is sufficiently slow in order not to invalidate the stationary phase analysis results. This latter approximation breaks down when there are strong reflected field components from laterally displaced terrain obstacles, as is often the case. Thus, the main approximation arises in going from Eqn. 10.9 to Eqn. 10.10.

10.2.3 Numerical evaluation of integral equations

The integral Eqn. 10.10 can only be easily solved at low to medium frequencies when the oscillating exponential $\exp(-jk\,[r_1 + r_2 - r_0])$ does not vary too rapidly. At high frequencies (typically in the UHF band and above) the numerical evaluation of the integral equation tends to become unstable resulting in significant loss of accuracy.

To evaluate Eqn. 10.10 we start by first solving the corresponding equation for a ground-based receiver. This is done by discretising the domain of integration into subintervals and solving this in an iterative manner as follows.

Dividing the interval $[0, x]$ into m steps each of length h then yields:

$$W(nh) = g(mh) - \left(\frac{jmh}{\lambda}\right)^{1/2} \times$$

$$\sum_{i=0}^{m-1} \int_0^h W(ih+y)\left(\delta(ih+y) + \frac{dr_2}{dn}\right)\exp(-jk[r_1 + r_2 - r_0])\frac{dy}{\sqrt{(ih+y)(mh-ih-y)}} \qquad (10.11)$$

Any discrete approximation to the integral in Eqn. 10.11 will contain only one $W(mh)$ term, which can then be moved to the left-hand side of the equation. Therefore, the integral equation becomes an iterative equation as the mth term only depends on the previous $m - 1$ terms. Care needs to be taken in the numerical integration scheme at the two endpoints as they both have an integrable singularity of the form $\int_0^h dx/\sqrt{x} = 2\sqrt{h}$.

Furthermore, the method usually requires the *a priori* knowledge of at least the first, if not the first few, terms in the sequence $\{W(0), W(h), W(2h), \ldots\}$. Estimating these first few values accurately usually relies on an appropriate analytical solution such as Sommerfeld's solution for a plane homogeneous earth, the two-ray model etc.

Finally, at high frequencies identification of the stationary points of the integrand in Eqn. 10.10 as a function of x can be exploited to assist the accurate numerical evaluation of the iterative Eqn. 10.11.

When the path gain function is found for all points on the ground along the path between the transmitter and receiver, deriving the attenuation at an elevated receiver is trivial, as we no longer identify the two $W(mh)$ terms on both sides of Eqn. 10.11. Then using values for r_0 and r_2 appropriate to the elevated antenna we evaluate this equation one last time.

As software capable of performing computations of (similar) integral equations is readily available [3], no numerical result examples will be presented here.

10.3 Parabolic equation methods

The parabolic equation is an example of a full-wave method, which attempts to directly solve the wave equation numerically, subject to a number of assumptions and simplifications. The method was derived as far back as the 1940s by Leontovich and Fock [4], but was deemed too computationally intensive to become a practical prediction tool.

With the advent of digital computers it found widespread use in underwater acoustic propagation prediction in the 1970s and 1980s [5, 6]. In underwater acoustics the wavelength tends to be long, thus easing the computational requirements somewhat.

By the 1990s, digital computers became fast enough and the parabolic equation method was applied to model propagation of radiowaves up to the UHF and SHF frequencies in the troposphere [7–9].

10.3.1 Derivation of the parabolic equation

Once again, we begin by reducing Maxwell's equations to the Helmholtz equation. However, in this instance we assume the presence of an atmosphere described by a complex refractive index:

$$n(\mathbf{r}) = \sqrt{\varepsilon(\mathbf{r})} = \sqrt{\varepsilon_r(\mathbf{r}) + j\sigma(\mathbf{r})\omega\varepsilon_0} \tag{10.12}$$

which is a continuously varying function of position. Provided that $\lambda_0 \nabla \ln n << 1$, the scalar Helmholtz equation describes accurately each of the Cartesian components of the electric and magnetic fields,

$$\nabla^2\psi + k_0^2 n^2 \psi = 0 \tag{10.13}$$

The wavenumber *in vacuo* is now given by $k_0 = 2\pi/\lambda_0$. Considering a two-dimensional propagation problem along the great circle path and making the approximation that the earth is flat[1] for simplicity, Eqn. 10.13 can be expanded in Cartesian co-ordinates as:

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial z^2} + k_0^2 n^2 \psi = 0 \tag{10.14}$$

We now introduce the assumption that an *a priori* preferred direction of propagation exists and identify this with the x axis. It is, therefore, reasonable that we can write the following form for the solution:

[1] The derivation of the parabolic equation in spherical co-ordinates appropriate to a spherical earth and the subsequent transformation to a (range, height) rectangular co-ordinate system is covered in Reference [7] and only changes the term $(n^2 - 1)$ in Eqn. 10.16 through to Eqn. 10.23 to $(n^2 - 1 - 2z/a_e)$ where a_e is the effective earth radius

$$\psi(x,z) = u(x,z)\exp(-jk_0x) \tag{10.15}$$

where the reduced wave amplitude $u(x,z)$ can now be assumed to vary slowly along the x direction on the scale of a free-space wavelength, λ. Substituting Eqn. 10.15 into Eqn. 10.14 and discarding the common factor $\exp(-jk_0x)$ after performing the differentiations, yields the following equation for the reduced wave amplitude:

$$\frac{\partial^2 u}{\partial x^2} - 2jk_0\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial z^2} + k_0^2(n^2 - 1)u = 0 \tag{10.16}$$

Eqn. 10.16 describes waves propagating both along the positive and negative x directions. By analogy with the one-dimensional wave equation

$$\frac{\partial^2 w}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 w}{\partial t^2} = 0 \qquad \Rightarrow \qquad \left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right)w = 0 \tag{10.17a}$$

which has linearly independent solutions

$$w = f(x - ct) \qquad \text{and} \qquad w = g(x + ct) \tag{10.17b}$$

which can be identified as forward and backward travelling waves corresponding to the two differential operators in Eqn. 10.17a. We factorise Eqn. 10.16 into forward and backward travelling wave operators:

$$\left(\frac{\partial}{\partial x} - jk_0 + jk_0\sqrt{1 + (n^2 - 1) + \frac{1}{k_0^2}\frac{\partial}{\partial z^2}}\right)\times$$
$$\left(\frac{\partial}{\partial x} - jk_0 - jk_0\sqrt{1 + (n^2 - 1) + \frac{1}{k_0^2}\frac{\partial^2}{\partial z^2}}\right)u = 0 \tag{10.18}$$

and discard the backward travelling wave for consistency with Eqn. 10.15, which finally gives:

$$\left(\frac{\partial}{\partial x} - jk_0 + jk_0\sqrt{1 + (n^2 - 1) + \frac{1}{k_0^2}\frac{\partial^2}{\partial z^2}}\right)u = 0 \tag{10.19}$$

It is to be understood that the differential operator under the square root sign in Eqn. 10.19 can only be interpreted in a formal sense. Its numerical evaluation can only be achieved by replacing the square root by a power series, or rational fractions, of operators, as we shall shortly see.

Thus, we rewrite Eqn. 10.19 as:

$$\frac{\partial u}{\partial x} = jk_0\left(1 - \sqrt{1 + (n^2 - 1) + \frac{1}{k_0^2}\frac{\partial^2}{\partial z^2}}\right)u \equiv jk_0\left(1 - \sqrt{1 + Q(x,z)}\right)u \tag{10.20}$$

The differential operator $Q(x,z)$ must give a significantly smaller answer than the unity operator when operated on $u(x,z)$ since, by assumption, the oscillatory variation of $i(x,z)$ is predominantly along the x direction, perpendicular to the z axis. Therefore, the z derivative on a scale of a wavelength $(1/k_0)$ is much smaller than the unity operator. For the atmosphere, we also know that $n(x,z) \approx 1$, giving:

$$Q(x,z)u(x,z) << u(x,z) \qquad \text{or, formally,} \qquad Q << 1 \tag{10.21}$$

The simplest approximation for the square root term is given by the first two terms in its Taylor expansion, namely,

$$\sqrt{1 + Q(x,z)} \cong 1 + Q(x,z)/2 \tag{10.22}$$

which finally yields the narrow-angle parabolic equation:

$$\frac{\partial u}{\partial x} = \frac{1}{2jk_0}\left(\frac{\partial^2 u}{\partial z^2} + k_0^2(n^2 - 1)u\right) \tag{10.23}$$

This is a first-order partial differential equation in x and can, therefore, be marched numerically if $u(z)$ is known on some initial plane x_0.

To investigate the range of angles of propagation θ, with respect to the x-axis, for which the prediction error arising from the approximation 10.22 remains bounded below any given threshold, we substitute the following plane wave solution into Eqn. 10.23 for $n(x,z) = 1$:

$$u(x,z) = A \exp\{jk_0 x(\cos\theta - 1) + jk_0 z \sin\theta\} \tag{10.24}$$

Subtracting the right-hand side of Eqn. 10.23 from the left-hand side and dividing by Eqn. 10.24 yields a relative error of:

$$e_{3DPE} = 1 - \cos\theta - \frac{\sin^2\theta}{2} \tag{10.25}$$

Numerical inspection of Eqn. 10.25 shows that for propagation angles in the range of ±15° the relative error at $\theta = 15°$ remains bounded below 0.00058. For most tropospheric propagation problems where propagation is predominantly along the surface of the earth over long distances, the above restriction tends not to be significant.

10.3.2 Summary of assumptions and approximations

A number of judicious approximations and assumptions have been made in deriving the parabolic equation method. These are summarised here for clarity:

1 The fast varying phase term has been factored out in order to make the equation less demanding to solve numerically (the marching steps can be in the range $10\lambda_0$–$100\lambda_0$, depending on the severity of the terrain). This implies that an *a priori* preferred direction of propagation must be chosen.

2 The reduced wave equation was factored into forward and backward travelling components and the backscattered component in the solution was discarded. Thus, we are making the implicit assumption that backscattering is not significant in the problems which we wish to study.

3 The angular domain of the reduced wave equation is restricted by

approximating the square root term. This assumes that high propagation angles and off great circle path scattering is not significant.

10.3.3 Parabolic equation marching (I) – the split step fast Fourier transform method

The split step Fourier transform method can be used to solve Eqn 10.23 as follows. We first rewrite Eqn. 10.23 in the form:

$$\frac{\partial u}{\partial x} = (a(x,z) + bD_z^2)u \tag{10.26a}$$

where

$$a(x,z) \equiv \frac{k_0}{2j}(n^2(x,z) - 1) \quad \text{and} \quad b \equiv \frac{1}{2jk_0} \tag{10.26b}$$

Fourier transforming $u(x,z)$ to $U(x,k_z)$ and using the theorem for the Fourier transform of derivatives, $F(D_z u) = jk_z U$, Eqn. 10.26*a* becomes:

$$\frac{\partial U}{\partial x} \cong aU - k_z^2 bU \tag{10.27}$$

where we have treated the term $a(x,z)$ as a constant due to the very weak spatial variation of the atmospheric refractive index. This of course introduces an error which we will quantify shortly. Eqn. 10.27 is a first-order ordinary differential equation under the above assumption, whose solution is:

$$U(x + \Delta x, k_z) = U(x,k_z)\exp(a\Delta x)\exp(-bk_z^2\Delta x) \tag{10.28}$$

Inverse Fourier transforming then yields:

$$u(x + \Delta x, z) = \exp(a\Delta x)F^{-1}(\exp(-bk_z^2\Delta x)F(u(x,z))) \tag{10.29}$$

The above equation can be used to march the solution forward along the x axis and can be implemented numerically using conventional fast Fourier transforms. However, the discretisation error arising from treating the refractive index as a constant now becomes range (Δx) dependent and is given by:

$$\{\exp(a\Delta x)\exp(bD_z^2\Delta x) - \exp((a + bD_z^2)\Delta x)\}u \tag{10.30}$$

The above equation can be further analysed in order to decide what maximum marching step size maintains the discretisation error below a certain desirable bound.

In order to implement the split step fast Fourier transform parabolic equation method we need to consider how to incorporate the terrain and its associated boundary condition given by Eqn. 10.6, as well as how to truncate the computational domain without introducing computational artefacts.

In the event that the terrain is completely flat, simply replacing the Fourier transform by the mixed Fourier transform pair [7,10], automatically satisfies the Leontovich boundary condition (10.6):

$$U(x,k_z) = \int_0^\infty u(x,z)\{-jk_0\delta \sin(k_z z) - k_2 \cos(k_z z)\}dz \tag{10.31a}$$

$$u(x,z) = \frac{2}{\pi}\int_0^\infty U(x,k_z)\frac{-jk_0\delta \sin(k_z z) - k_z \cos(k_z z)}{k_z^2 - k_0^2\delta^2}dk_z + K(x)\exp(jk_0\delta z) \tag{10.31b}$$

where

$$K(x) = K(x_0)\exp\{-jk_0\delta(n^2 - 1 - \delta^2)x/2\} \tag{10.31c}$$

and

$$K(x_0) = -2jk_0\delta\int_0^\infty u(x_0,z)\exp(jk_0\delta z)dz \quad \mathrm{Re}(-jk_0\delta) > 0 \tag{10.31d}$$

or

$$K(x_0) = 0 \qquad \mathrm{Re}(-jk_0\delta) < 0 \tag{10.31e}$$

However, as the terrain boundary is not flat, but irregular, an earth flattening transformation [9] can be introduced through a change of variables:

$$\chi = x, \quad \zeta = z - T(x) \quad \text{and} \quad T(x) = t(x) - \frac{x^2}{2a_e} \tag{10.32}$$

where $t(x)$ describes the terrain profile and a_e is the earth's radius. The transformed reduced wave equation amplitude is given by:

$$\hat{u}(\chi,\zeta) = u(x,z)\exp\left(j\,[k_0\zeta T'(\chi)] + \tfrac{3}{2}k_0\int_0^\chi [T'(a)]^2 da\right) \tag{10.33}$$

and the transformed parabolic Eqn. 10.23 is:

$$\frac{\partial \hat{u}}{\partial \chi} = \frac{1}{2jk_0}\left(\frac{\partial^2 \hat{u}}{\partial \zeta^2} + k_0^2[n^2(\chi,\zeta) - 1 - 2\zeta\{t''(\chi) - 1/a_e\}]\hat{u}\right) \tag{10.34}$$

As can be seen in Figure 10.3, the simplest manner in which the domain of computation can be truncated if the region of interest is $z \le H$ is to introduce a region of space at $H \le z \le 2H$ in which the refractive index and, therefore, the wavenumber have a small imaginary part which results in the dissipation of the wave amplitude through absorption. Care must be taken to ensure that the resulting impedance discontinuity at $z = H$ is negligible to avoid nonphysical reflections from this fictitious boundary. This method is computationally expensive as it doubles the solution domain size. Methods that are more efficient exist, and one of them will be presented in the next section.

10.3.4 Parabolic equation marching (II) – finite difference implementation

This method is based on a more direct discretisation of Eqn. 10.23, through the introduction of a rectangular grid and the evaluation of the various derivative terms using centred finite differences. Figure 10.4 depicts six field points on this grid, three on each marching plane. Hereafter we adopt the notation and $x = nk$ and $z = mh$, where n and m are integers and k and h are the grid spacing in the x and z direction, respectively.

The various terms appearing in Eqn. 10.23 are evaluated at the centre point through their finite difference discrete approximations to yield:

$$\frac{\partial u}{\partial x}(n+\tfrac{1}{2}) = \frac{u_m^{n+1} - u_m^n}{k} \tag{10.35a}$$

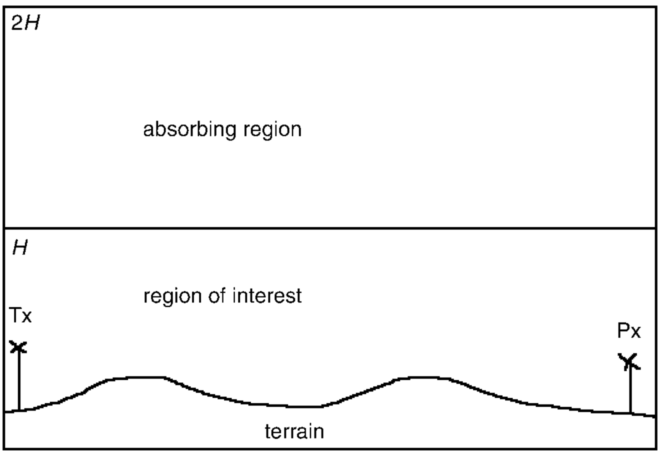


Figure 10.3 The various computation domains in the split step fast Fourier transform implementation of the parabolic equation

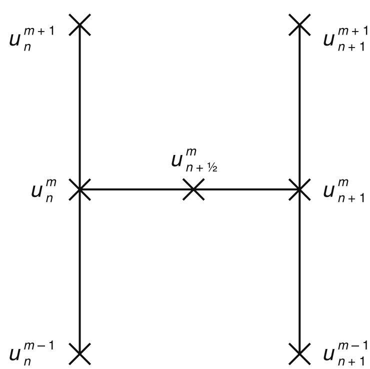


Figure 10.4 The discretisation grid including the centre point

$$\frac{\partial^2 u}{\partial z^2}(n+\tfrac{1}{2}) = \frac{u_{m-1}^{n} - 2u_m^n + u_{m+1}^{n} + u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}}{2h^2} \tag{10.35b}$$

$$[n^2-1]u(n+\tfrac{1}{2}) = [(n_m^{n+\frac{1}{2}})^2 - 1]\frac{u_m^n + u_m^{n+1}}{2} = a_m^{n+\frac{1}{2}}\frac{u_m^u + u_m^{n+1}}{2} \tag{10.35c}$$

Substituting Eqns 10.35 into Eqn. 10.23 and rearranging terms, we can separate the field points on the mth and $(m+1)$th x planes:

$$\begin{aligned}&\left[u_m^{n+1} - \tfrac{1}{2}k\left(a_m^{n+\frac{1}{2}}u_m^{n+1} + b\left(\frac{u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}}{h^2}\right)\right)\right]\\ &= \left[u_m^n + \tfrac{1}{2}k\left(a_m^{n+\frac{1}{2}}u_m^n + b\left(\frac{u_{m-1}^{n} - 2u_m^n + u_{m+1}^{n}}{h^2}\right)\right)\right]\end{aligned} \tag{10.36}$$

When all the grid points on both the mth and $(m+1)$th planes are taken into account, the above equation generalises to the matrix equation,

$$\begin{aligned}&\begin{bmatrix} X_1 & -\Gamma_1^{n+1} & 0 & \cdots & 0 & 0 \\ -\Gamma_2^{n+1} & X_2 & -\Gamma_2^{n+1} & \cdots & 0 & 0 \\ & \cdot & \cdot & & & \\ & \cdot & \cdot & & & \\ 0 & 0 & 0 & \cdots & X_{m-1} & -\Gamma_{m-1}^{n+1} \\ 0 & 0 & 0 & \cdots & -\Gamma_m^{n-1} & X_{\mathrm{m}} \end{bmatrix}\begin{bmatrix} u_1^{n+1} \\ u_2^{n+1} \\ \cdot \\ \cdot \\ u_{m-1}^{n+1} \\ u_m^{n+1} \end{bmatrix}\\ &= \begin{bmatrix} Y_1 & \Gamma_1^n & 0 & \cdots & 0 & 0 \\ \Gamma_2^n & Y_2 & \Gamma_2^n & \cdots & 0 & 0 \\ & \cdot & \cdot & & & \\ & \cdot & \cdot & & & \\ 0 & 0 & 0 & \cdots & Y_{m-1} & \Gamma_{m-1}^n \\ 0 & 0 & 0 & \cdots & \Gamma_m^n & Y_m \end{bmatrix}\begin{bmatrix} u_1^n \\ u_2^n \\ \cdot \\ \cdot \\ u_{m-1}^n \\ u_m^n \end{bmatrix}\end{aligned} \tag{10.37a}$$

where

$$\Gamma_m^n = \tfrac{1}{2}(k/h^2)b \tag{10.37b}$$

$$X_m = X_m^{n+1} = 1 - \tfrac{1}{2}ka_m^{n+\frac{1}{2}} + (k/h^2)b \tag{10.37c}$$

$$Y_m = Y_m^n = 1 + \tfrac{1}{2}ka_m^{n+\frac{1}{2}} - (k/h^2)b \tag{10.37d}$$

The terrain boundary condition 10.6 can similarly be discretised and rearranged to yield:

$$\begin{aligned}&u_1(2a(x,z)h^2 \sin\theta_g + 2b\sin\theta_g + 3h\cos\theta_g - 2jk_0h^2[\delta + \sin\theta_g])\\ &- u_2(4bh\sin\theta_g + 4h\cos\theta_g) + u_g(2b\sin\theta_g + h\cos\theta_g) = 0\end{aligned} \tag{10.38}$$

where θ_g is the angle between the tangent to the ground and the x axis. Eqn. 10.38 can be inserted into the matrix expression 10.37 to enforce the terrain boundary condition explicitly by replacing the appropriate matrix elements.

The incorporation of the terrain boundary condition also requires that the terrain profile is discretised as shown in Figure 10.5, so that the variable marching step size can ensure that one field point is added or subtracted at any one step to coincide with the terrain boundary.

A computationally efficient way in which to achieve the termination of the nonphysical upper boundary is to enforce mathematically a zero downward reflected wave. The method of achieving this is known as a nonreflecting boundary condition which, as we shall shortly see, is nonlocal. Thus, instead of the computational domain being described by Figure 10.3, it can now be described by Figure 10.6, which represents a much smaller domain.

For simplicity we shall assume that the nonreflecting boundary will be placed sufficiently high so that $n = 1$. The parabolic Eqn. 10.23 can then be Laplace transformed to yield:

$$\frac{\partial^2 U}{\partial z^2}(p,z) - 2jk_0 p\,U(p,z) = -2jk_0 u(x=0,z) \tag{10.39}$$

whose general solution expressed as an initial condition problem at $z = H$ is:

$$U(p,z) = A(p,H)e^{(1+j)\sqrt{k_0 p}(z-H)} + B(p,H)e^{(1+j)\sqrt{k_0 p}(z-H)} \tag{10.40}$$

Discarding the downward travelling wave component imposes a nonreflecting boundary condition, thus requiring that:

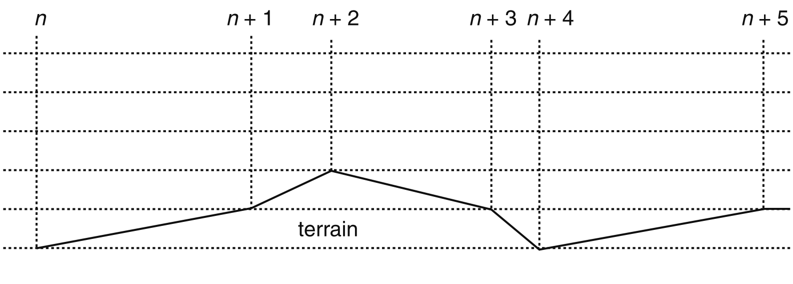


Figure 10.5 Variable marching step size terrain discretisation

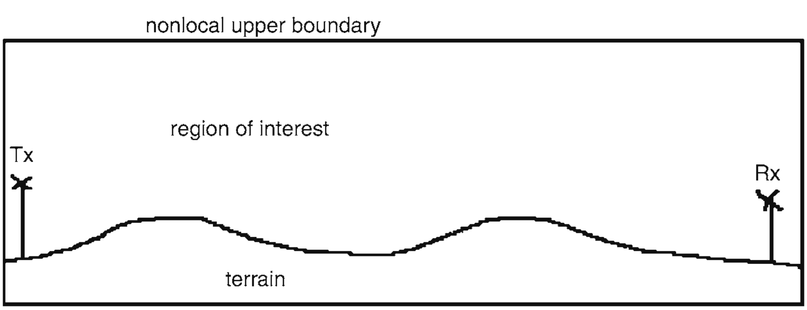


Figure 10.6 The computational domain for the parabolic equation with nonlocal nonreflecting boundary conditions

$$U(p,z) = U(p,h)e^{(1+j)\sqrt{k_0 p}(z-H)} \tag{10.41}$$

Taking the partial derivative of Eqn. 10.41 with respect to z at $z = H$ gives:

$$\frac{\partial U}{\partial z}(p,H) = (1+j)\sqrt{k_0}\{pU(p,H)\}\frac{1}{\sqrt{p}} \tag{10.42}$$

Finally, inverse Laplace transforming yields:

$$\frac{\partial u}{\partial z}(x,H) = (1+j)\sqrt{\frac{k_0}{\pi}}\int_0^x \frac{\partial u}{\partial z}(\kappa,H)\frac{d\kappa}{\sqrt{x-\kappa}} \tag{10.43}$$

Eqn. (10.43) is a nonlocal nonreflecting boundary condition, which can now be discretised directly and incorporated into the matrix expression 10.37 in the same manner as the Leontovich boundary condition for the terrain 10.38. The only difference is that the vertical field gradient at range x is determined in terms of a summation of field gradients at all previous ranges. Nevertheless, this does preserve the marching property of the method.

10.3.5 Parabolic equation conclusions

In summary, the parabolic equation method:

1 is numerically efficient (computation times are $T_{ssfft} \propto k_0\theta_{\max}x_{\max}z_{\max}$ and $T_{fd} \propto k_0^2\theta_{\max}^3x_{\max}z_{\max}$ for the SSFFT and FD methods, respectively);
2 is robust to errors and small modelling inaccuracies;
3 is ideal for long ranges due to large step size;
4 has a transmitting antenna radiation pattern which can be incorporated in the initial field aperture (since it is the field's angular spectrum of plane waves);
5 uses an accurate physical representation of both the terrain and troposphere.

The limitations of the parabolic equation method have already been summarised in Section 10.3.2 and need not be repeated here.

Various research challenges exist for long-range terrain modelling, from ways of constructing efficient three-dimensional parabolic equation solvers to extending existing short-range vector parabolic equation methods [11].

10.3.6 Sample applications of the parabolic equation method

Figure 10.7 shows the type of predictions that the parabolic equation is capable of. Not only does the method predict in a single sweep the field coverage over the entire domain, but also it facilitates through visualisation the best positioning of the receiving antenna.

In Figure 10.8 a contour terrain map of the Bromyard Downs area near

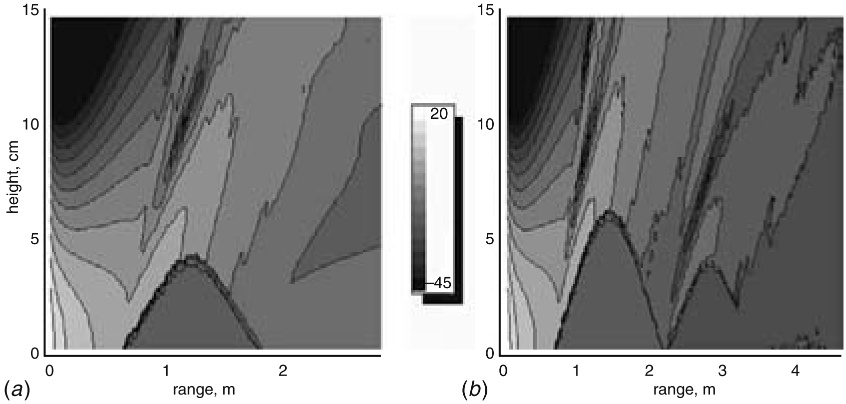


Figure 10.7 *Illustrative coverage diagrams generated by the parabolic equation method – propagation over perfectly conducting laboratory-sized obstacles at 30 GHz (computation time is a few seconds on a PC)*
(a) rounded hill
(b) two rounded hills

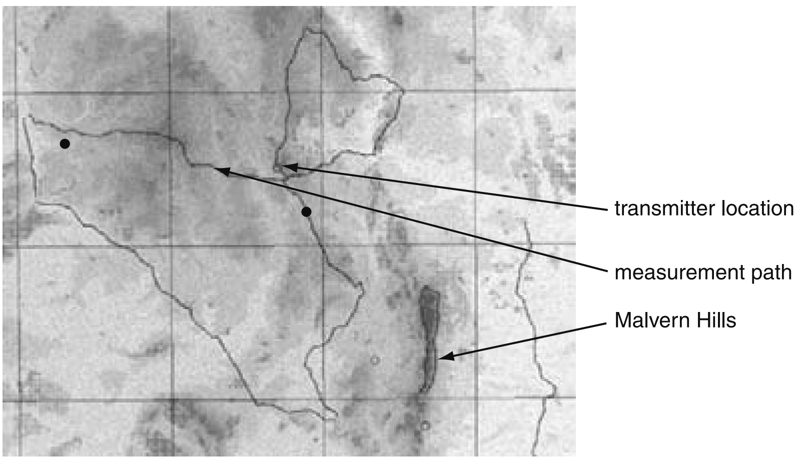


Figure 10.8 *Measurement area in rolling countryside (Bromyard Downs in Worcestershire) – measurements were conducted at 150 MHz; measurement path endpoints indicated by black dots*

Worcester is given. The predictions of the parabolic equation method are shown alongside measurements in Figure 10.9, with excellent agreement between the two. However, it should be borne in mind that the comparison was done under standard atmospheric conditions in a gently rolling terrain, arguably an easy test given the capabilities of the parabolic equation method.

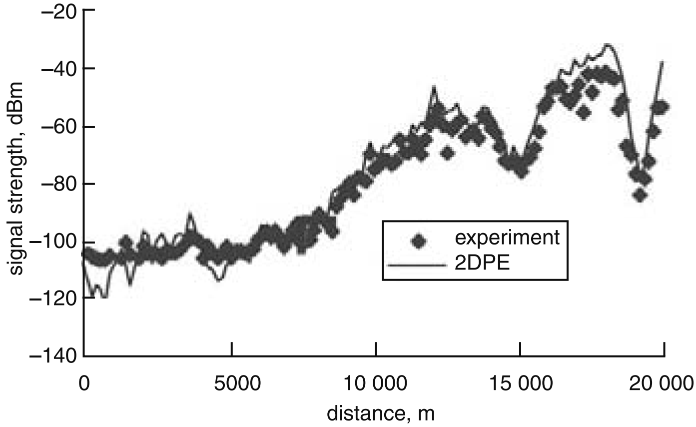


Figure 10.9 Comparison of the predictions of the parabolic equation method with measurements at 150 MHz

10.4 Ray-tracing methods

The discussion here concerns ray-tracing propagation prediction methods only used in mobile radio applications in the UHF and SHF bands. Ray-tracing methods are based on geometrical optics and its extensions and are, therefore, intrinsically high-frequency techniques. The rule of thumb is that the smallest terrain or environment feature must be about an order of magnitude greater than the wavelength in order for ray tracing propagation methods to be valid.

Ray-optical tools tend to be used to predict wideband channel characteristics (e.g. an average channel impulse response, angles of arrival etc.) both indoors and outdoors. As ray-tracing methods are site specific they require as an input a building database (usually three dimensional) which contains at least building (or room) outlines and heights.

In general, these methods are exceptionally computationally intensive and this restricts the range at which they can be used to at most a few kilometres. To appreciate the computational complexity of the methods we shall briefly consider the way in which ray-tracing programs work.

10.4.1 Ray-tracing elements

Ray-tracing attempts to model all possible direct (free-space) fields, specularly reflected fields using Fresnel reflection coefficients, diffracted fields using the uniform theory of diffraction and (sometimes) transmitted fields using Fresnel transmission coefficients.

A natural consequence is that as many ray paths as possible connecting a transmitter to a receiving location must be traced in order to ensure that accurate predictions can be made. In general, there are two ray-tracing methods.

The first of the two methods is the imaging method [12,13]. Here, we locate all the images of the transmitter into all the facets in the environment in order to model single and multiple reflections, as shown in Figure 10.10. All the illuminated edges are also located to model single and multiple diffractions. Subsequently, all the illuminated edges, which act as secondary sources, are imaged in order to model mixed reflections and diffractions. Finally, if transmission through walls is known to be significant, transmitted rays are traced out. In general, mixed transmissions and reflections or diffractions are not modelled because the number of rays needing to be traced rapidly becomes intractably large.

The second method is called the point and shoot, or ray-launching, method [14]. This involves the launching of rays spaced apart with a small angular increment at the transmitter, as shown in Figures 10.10 and 10.11. When a ray

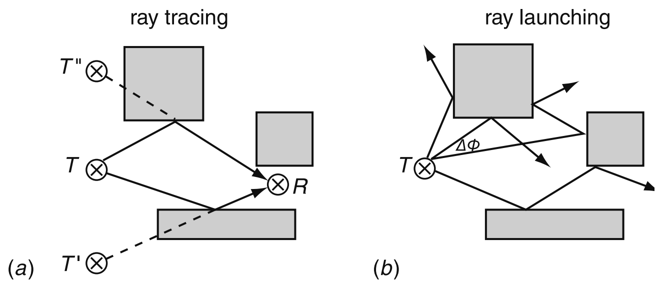


Figure 10.10 Ray-tracing methods
(a) imaging technique, individual determination of rays, point-to-point prediction
(b) angle increment, launching of rays, area prediction
Source: http://www.ihf.uni-Stuttgart.de/Winprop/winprop_e.html

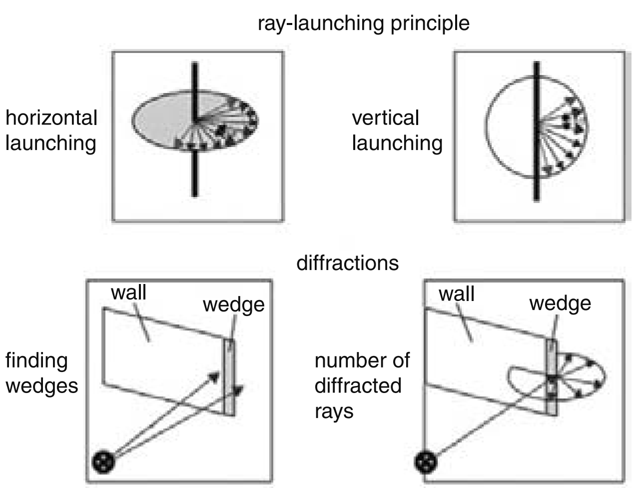


Figure 10.11 Ray-launching principle and secondary source illumination
Source: http://www.ihf.uni-Stuttgart.de/Winprop/winprop_e.html

intercepts a surface, the law of reflection is applied, whereas when the ray passes through a predetermined small distance from either an edge or the receiver, it is deemed to illuminate them. The assumption here is that small perturbations in the launching angle result in a ray path that both exists and exactly crosses the edge or receiver. Edges are then treated as secondary sources. When the spacing between rays becomes too large, secondary rays can be launched from the same wavefront in the vicinity of the original ray to extend the range at which the ray launching maintains its accuracy. In general, ray launching tends to be much faster than the imaging method, but can occasionally miss important rays thus compromising its prediction accuracy.

Preprocessing of the environment can reduce significantly the number of edges and facets considered for possible mutual illumination, thus cutting down computational time. The imaging method is the one that naturally benefits the most from this approach. The way this is conventionally done is that logic rules are applied to construct a visibility, or connectivity, matrix between objects in the environment, and only rays bouncing between allowable objects are traced.

Termination of the tracing of rays is usually achieved by both monitoring the attenuation experienced by a ray and discarding it if this exceeds a preset threshold value, and by permitting a maximum number and type of interactions per ray path (e.g. maximum six reflections and one diffraction per ray path).

10.4.2 Field strength calculation

Once the paths have been traced, the field strength associated with each path is computed using geometrical optics and its extensions. Typically, a mixed path with n reflections and m diffractions can be written as:

$$\mathbf{E}(Rx) = \mathbf{E}(ref)\,.\prod_{\nu=1}^{n}\mathsf{R}\,(\theta_i)\,.\prod_{\mu=1}^{m}\mathsf{D}\,(\varphi,\varphi')A(\{\mathbf{r}_i\})\exp\left\{-jk\sum_i|\mathbf{r}_i|\right\} \qquad (10.44)$$

where the reflection, R, and diffraction, D, coefficients are dyads (rank two tensors) derived through high-frequency asymptotic analysis of canonical electromagnetic scattering problems [15]. $\mathbf{E}(ref)$ is a vector representing the reference electric field at a unit distance from the transmitter along the direction of the traced path, A is an appropriate amplitude spreading factor to account for astigmatism in the ray and $\{\mathbf{r}_i\}$ are the lengths of the segments comprising this path. The products in (10.44) are all scalar vector products.

Clearly, ray-tracing models are fully three dimensional and capable of polarimetric predictions, a property which can be exploited to inform advanced antenna design work.

A typical output result from a ray-tracing tool is shown in Figure 10.12. In this instance only the local mean power coverage is shown, but such tools can typically also generate power delay profiles, angle-of-arrival information, delay spreads and coherence bandwidths.

Ray-tracing tools possess disadvantages other than their sheer computational

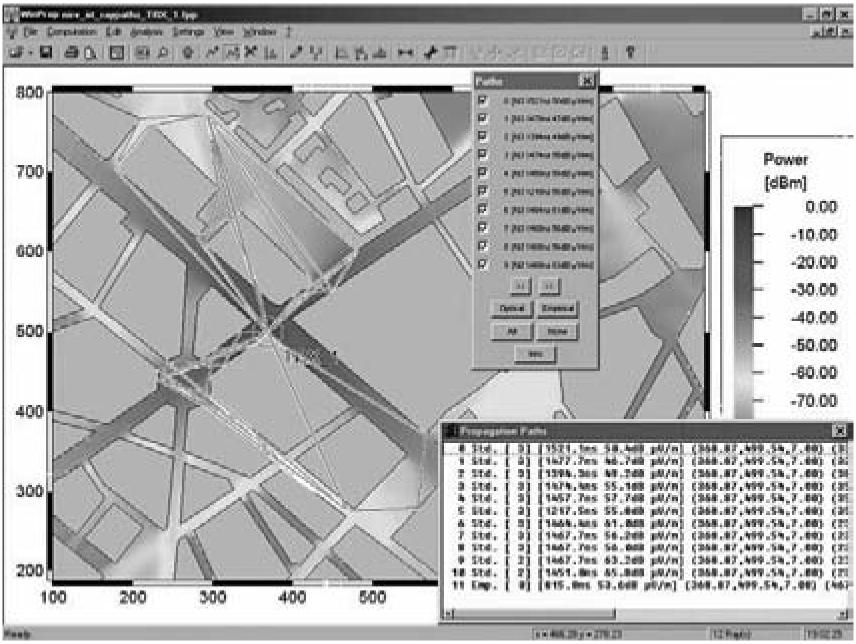


Figure 10.12 A ray-tracing tool output showing traced paths between the transmitter and one receiver location
Source: http://www.ihf.uni-stuttgart.de/Winprop/winprop_e.html

complexity. In particular, they are not capable of modelling nongeometrical optic processes such as diffuse scattering from rough surfaces and propagation through disordered media such as vegetation. Furthermore, their prediction capabilities are often compromised through our lack of knowledge of the interior structure of buildings, which often causes strong scattering that cannot be modelled through canonical electromagnetic scattering problems. Finally, in certain environments the *a priori* selection of propagation mechanisms to be modelled may fail to take into account important propagation mechanisms, thus compromising the ray-tracing tool's prediction capabilities significantly.

In spite of such limitations, ray-tracing tools are some of the most promising medium to short-range propagation models available for built-up areas.

10.5 References

1 HUFFORD, G.A.: 'An integral equation approach to the problem of wave propagation over an irregular terrain', *Q. J. Appl. Math.*, 1952, **9** (4), pp.391–404
2 LEONTOVICH, M.A.: 'On a method of solving the problem of electromagnetic waves near the surface of the earth', *Bull. Ac. Sci. USSR, Ser. Phys.*, 1944, **8**, pp.16–22

3 WU, Z., and MACLEAN, T.S.M.: 'Radiowave propagation over ground software' (Chapman and Hall, London, 1998)
4 LEONTOVICH, M.A. and FOCK, V.: 'Solution of the problem of electromagnetic waves along the earth's surface by the method of parabolic equation', *J. Physics (USSR)*, 1946, **10**, pp.13–24
5 TAPPERT, F.D.: 'The parabolic equation method', *in* KELLER, J.B., and PAPADAKIS, J. (Eds.): 'Wave propagation and underwater acoustics, Lecture notes in physics, **70**' (Springer, Berlin, 1977)
6 LEE, D., and McDANIEL, S.T.: 'Ocean acoustic propagation by finite difference methods', *J. Comp. & Math. With Appl.*, 1987, **14**, pp.305–423
7 KUTTLER, J.R., and DOCKERY, G.D.: 'Theoretical description of the parabolic approximation/Fourier split-step method of representing electromagnetic propagation in the troposphere', *Radio Sci.*, 1991, **26**, pp.381–393
8 CRAIG, K.H., and LEVY, M.F.: 'Parabolic equation modelling of the effects of multipath and ducting on radar systems', *IEE Proc. F.*, 1991, **138**, pp.153–162
9 BARRIOS, A.E.: 'A terrain parabolic equation model for propagation in the troposphere', *IEEE Trans. Antennas Propag.*, 1994, **42**, pp.90–98
10 DOCKERY, G.D., and KUTTLER, J.R.: 'An improved impedance-boundary algorithm for Fourier split-step solutions of the parabolic equation', *IEEE Trans. Antennas Propag.*, 1996, **44**, pp.1592–1599
11 ZAPOROZHETS, A.A.: 'Application of the vector parabolic equation method to urban radiowave propagation problems', *IEE Proc. H*, 1999, **146**, pp.253–256
12 IKEGAMI, F., TAKEUCHI, T., and YOSHIDA, S.: 'Theoretical prediction of mean field strength for urban mobile radio', *IEEE Trans. Antennas Propag.*, 1991, **39**, pp.299–302
13 ROSSI, J.-P., and LEVY, A.J.: 'A ray model for decimetric radiowave propagation in an urban area', *Radio Sci.*, 1992, **27**, pp.971–979
14 LUEBBERS, R., and SCHUSTER, J.: 'Site-specific propagation prediction methods for urban environments'. AGARD CP-574, 1995, pp.41.1–12
15 HANSEN, R.C. (Ed.): 'Geometric theory of diffraction' (IEEE Press, New York, 1981)

Chapter 11

Outdoor mobile propagation

Simon R. Saunders

11.1 Introduction

This Chapter describes key modelling concepts for mobile communication systems operated in outdoor environments. Much of the content is adapted from Saunders [1], to which the reader is referred for further detail.

Figure 11.1 shows a few of the many interactions between electromagnetic waves, the antennas which launch and receive them and the environment through which they propagate, in order to understand outdoor mobile propagation. Sometimes these effects are treated using detailed physical models, but more usually they are considered too complex and are treated in an empirical or statistical manner. This Chapter examines how these effects are modelled for macrocells and microcells, which together comprise the main system types used in cellular mobile communications.

11.1.1 The outdoor mobile radio channel

The outdoor mobile channel is a special case of the channel defined for generic communication systems by Shannon [2]. The transmitted information, encoded as suitable waveforms by the transmitter, is modified by channel noise in ways which may be more or less unpredictable to the receiver, so the receiver must be designed to overcome these modifications and hence to deliver the information to its final destination with as few errors or distortions as possible.

In the wireless channel, the noise sources can be subdivided into multiplicative and additive effects. The additive noise arises from noise generated within the receiver itself, such as thermal and shot noise in passive and active devices and also from external sources particularly interference from other transmitters and electrical systems. Some of this interference may be intentionally introduced, but carefully controlled, such as when channels are reused in order to maximise the capacity of a cellular radio system.

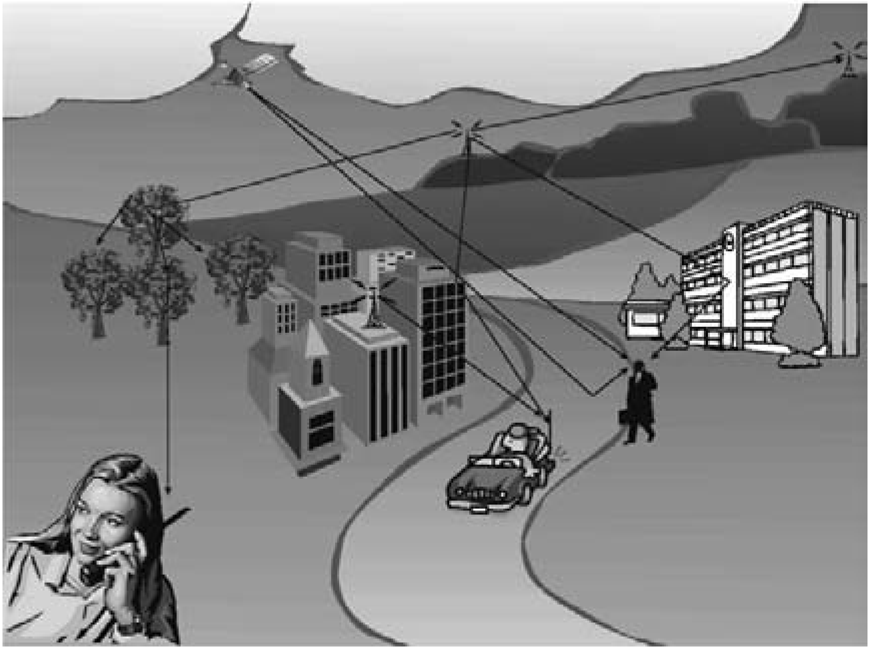


Figure 11.1 The outdoor wireless propagation landscape

Multiplicative noise arises from the various processes encountered by transmitted waves on their way from the transmitter antenna to the receiver antenna. Here are some of them:

- the directional characteristics of both the transmitter and receiver antennas
- reflection (from the smooth surfaces of walls and hills)
- absorption (by walls, trees and by the atmosphere)
- scattering (from rough surfaces such as the sea, rough ground and the leaves and branches of trees)
- diffraction (from edges, such as building rooftops and hilltops)
- refraction (due to atmospheric layers and layered or graded materials)

It is conventional to subdivide the multiplicative processes in the channel into three types of fading, path loss, shadowing (or slow fading) and fast fading (or multipath fading), which appear as time-varying processes between the antennas, as shown in Figure 11.2. All of these processes vary as the relative positions of the transmitter and receiver change and as any contributing objects between the antennas are moved.

An example of the three fading processes is illustrated in Figure 11.3; it shows the signal received by a mobile receiver moving away from a transmitting base station. The path loss is an overall decrease in field strength, as the distance between the transmitter and the receiver increases. The physical processes which cause it are the outward spreading of waves from the transmit antenna and the

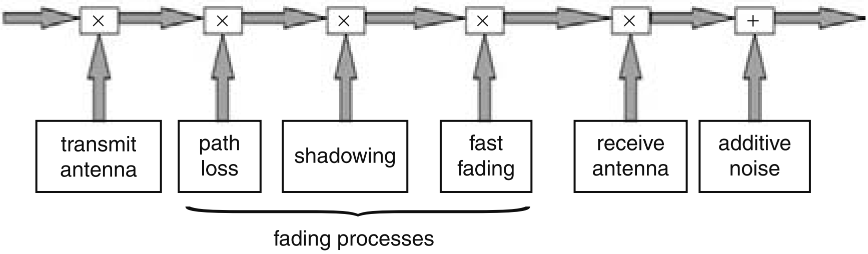


Figure 11.2 Contributions to noise in the wireless channel

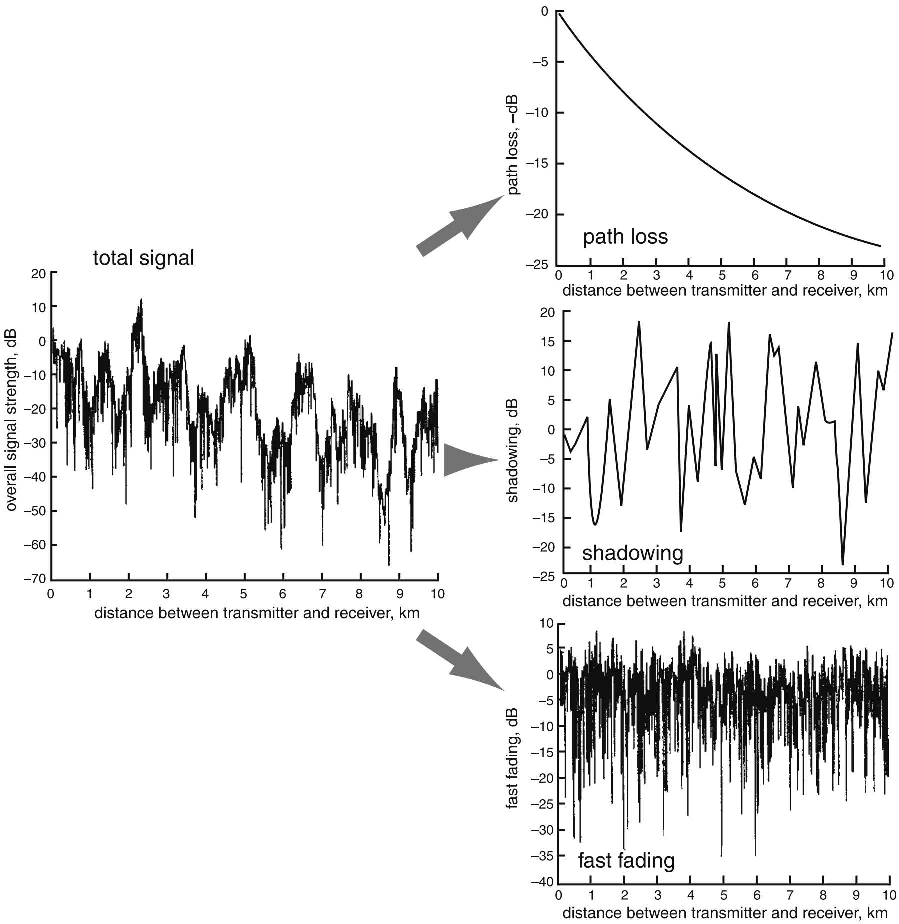


Figure 11.3 The three scales of mobile signal variation

obstructing effects of trees and buildings; a typical system may involve variations in path loss of around 150 decibels (dB) over its designed coverage area. Superimposed on the path loss is the shadowing, which changes more rapidly with significant variations over distances of hundreds of metres and generally involves variations of up to around 20 dB. Shadowing arises due to the varying nature of the particular obstructions between the base and the mobile, such as particular tall buildings or dense woods. Fast fading involves variations on the scale of a half wavelength (50 cm at 300 MHz, 17 cm at 900 MHz) and frequently introduces variations as large as 35 to 40 dB. It results from the constructive and destructive interference between multiple waves reaching the mobile from the base station. In this Chapter only the path loss and shadowing for outdoor terrestrial mobile systems will be examined. These are the key variations for predicting area coverage for practical systems.

Most outdoor mobile propagation occurs at VHF and above. The corresponding wavelengths, of the order of metres or less, are shorter than the critical dimensions of most objects encountered, such as hills or buildings. This allows the application of various high-frequency approximations to simplify the analysis of propagation effects.

11.1.2 System types

This Chapter covers two main outdoor system types, namely macrocells and microcells.

(i) *Macrocells* (Section 11.2): designed to provide mobile services (including both voice and paging), particularly outdoors, to rural, suburban and urban environments with medium traffic densities. Base station antenna heights are greater than the surrounding buildings, providing a cell radius from around 1 km to many tens of kilometres. Mostly operated at VHF and UHF.

(ii) *Microcells* (Section 11.4): designed for high traffic densities in urban and suburban areas to users both outdoors and within buildings. Base station antennas are lower than nearby building rooftops, so coverage area is defined by street layout. Cell length up to around 500 m. Again, mostly operated at VHF and UHF, but services as high as 60 GHz have been proposed.

Several other system types are examined in depth in Saunders [1].

11.2 Macrocells

11.2.1 Introduction

This Section introduces methods for predicting the path loss encountered in macrocells. In principle, wave propagation could be analysed in detail to predict the loss over every path between the base station and every possible mobile location. However, the data describing the terrain and clutter would be very large and the computational effort involved would often be excessive. Even if such resources were available, the important parameter for the macrocell designer is the overall area covered, rather than the specific field strength at particular locations.

The models presented in this section treat the path loss associated with a given macrocell as being dependent on distance, provided that the environment surrounding the base station is fairly uniform. In consequence, the coverage area predicted by these models for an isolated base station will be approximated as circular. Although this is clearly inaccurate, it is useful for system dimensioning purposes. Methods will be indicated in Section 11.3 for improving the reality of this picture.

11.2.2 Definition of parameters

The following terms will be used here in defining path loss models and are illustrated in Figure 11.4:

h_m	mobile station antenna height above local terrain height (m), often taken as 1.5 m
d_m	distance between the mobile and the nearest building (m)
h_o	typical (usually the mean) height of a building above local terrain height (m)
h_b	base station antenna height above local terrain height (m)
r	great circle distance between base station and mobile (m)
$R = r \times 10^{-3}$	great circle distance between base station and mobile (km)
f	carrier frequency (Hz)
$f_c = f \times 10^{-6}$	carrier frequency (MHz)
λ	free space wavelength (m)

The basic definition of a macrocell is that $h_b > h_0$. Although buildings are not the only obstructions in practice, they are usually by far the most significant obstructions at typical macrocellular frequencies. Typical base station heights in practice are around 15 m if a mast is used, or around 20 m upwards if deployed

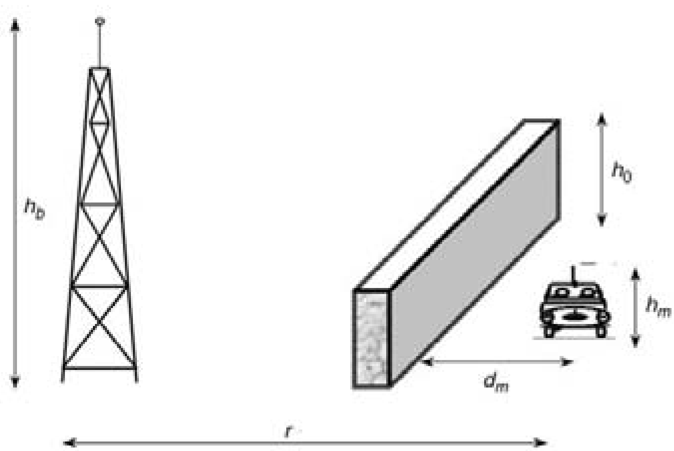


Figure 11.4 Definition of parameters for macrocell propagation models

on a building rooftop. The effective base station height may be increased dramatically by locating it on a hill overlooking the region to be covered.

11.2.3 Empirical path loss models

In order to create practical predictions of macrocell coverage given the wide variety of propagation mechanisms illustrated in Figure 11.1, simplifications are needed. One appropriate way of accounting for these complex effects is *via* an empirical model. To create such a model, an extensive set of actual path loss measurements is made, and an appropriate function is fitted to the measurements, with parameters derived for the particular environment, frequency and antenna heights so as to minimise the error between the model and the measurements. Note that each measurement represents an average of a set of samples, the local mean, taken over a small area (around 10–50 m), in order to remove the effects of fast fading, as originally suggested by Clarke [3]. See Parsons [4] and Lee [5] for further details of measurement procedures. The model can then be used to design systems operated in similar environments to the original measurements. A real example of an empirical model fitted to measurements is shown in Figure 11.5. Methods of accounting for the very large spread of the measurements at a given distance are the subject of Section 11.3.

The simplest useful form for an empirical path loss model is as follows:

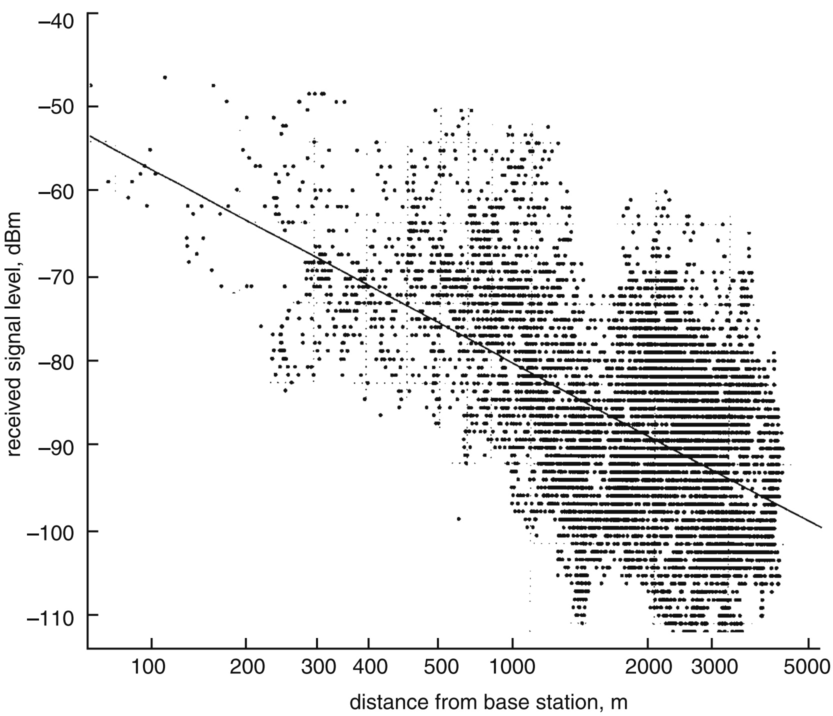


Figure 11.5 Empirical model of macrocell propagation: the dots are measurements taken in a suburban area and the line represents a best-fit empirical model

$$\frac{P_R}{P_T} = \frac{1}{L} = \frac{k}{r^n} \text{ or in decibels } L = 10n \log r + K \tag{11.1}$$

where P_T and P_R are the effective transmitted and predicted isotropic received powers, L is the path loss, r is the distance between the base station and the mobile and $K = 10 \log_{10} k$ and n are constants of the model. This will be referred to as a power law model. A more convenient form (in decibels) is

$$L = 10n \log(r/r_{ref}) + L_{ref} \tag{11.2}$$

where L_{ref} is the predicted loss at a reference distance r_{ref}.

Both the free-space loss and the plane earth loss can be expressed in this form. The parameter n is known as the path loss exponent and is found by measurement to depend on the system parameters, such as antenna heights and the environment. The path loss exponent is critical in establishing the coverage and capacity of a cellular system. Parameter k can be considered as the reciprocal of the propagation loss that would be experienced at one metre range ($r = 1$ m).

11.2.3.1 Clutter factor models

Measurements taken in urban and suburban areas usually find a path loss exponent close to 4, just as in the plane earth loss but with a greater absolute loss value – smaller K in Eqn. 11.1. This has led to some models being proposed which consist of the plane earth loss (see Chapter 6), plus (in decibels) an extra loss component called the clutter factor, as shown in Figure 11.6. The various models differ basically in the values which they assign to k and n for different frequencies and environments.

A good example of a clutter factor model is the method due to Egli [6], which is based upon a large number of measurements taken around American cities. Egli's overall results were originally presented in nomograph form, but Delisle [7] has given an approximation to these results for easier computation:

$$L = 40 \log R + 20 \log f_c - 20 \log h_b + L_{\mathrm{m}} \tag{11.3}$$

where

$$L_m = \begin{cases} 76.3 - 10 \log h_m & \text{for } h_m < 10 \\ 76.3 - 20 \log h_m & \text{for } h_m \geq 10 \end{cases} \tag{11.4}$$

Note that this approximation (Eqn. 11.4) involves a small discontinuity at $h_m = 10$ m. Although plane earth loss is frequency independent, this model introduces an additional f_c^{-2} received power dependence, which is more representative of the results of real measurements. For very large antenna heights, the loss predicted by Eqn. 11.3 may be less than the free-space value, in which case the free-space value is used.

The mobile antenna height-gain characteristic is approximately linear for antennas which clear the surrounding terrain features, assumed to be above 10 m. Elsewhere there is a square root variation for heights in the range

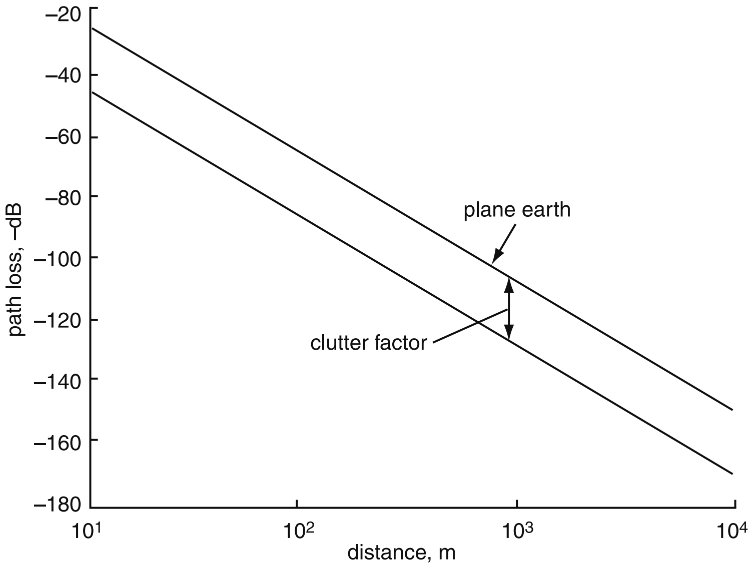


Figure 11.6 *Clutter factor model, note that the* y *axis in this Figure and in several to follow is the negative of the propagation loss in decibels; this serves to make clear the way in which the received power diminishes with distance*

2–10 m. The transition value (10 m) presumably corresponds to the mean obstruction height, although no correction is made for other heights. The average effect of polarisation is considered negligible.

11.2.3.2 The Okumura–Hata model

The Okumura–Hata model is a fully empirical prediction method [8], based entirely upon an extensive series of measurements made in and around Tokyo city between 200 MHz and 2 GHz. There is no attempt to base the predictions on a physical model such as the plane earth loss. Predictions are made *via* a series of graphs, the most important of which have since been approximated in a set of formulae by Hata [9]. The thoroughness of these two works taken together has made them the most widely quoted macrocell prediction model, often regarded as a standard against which to judge new approaches. The urban values in the model presented below have been standardised for international use by the ITU [10].

The method involves dividing the prediction area into a series of clutter and terrain categories, namely open, suburban and urban. These are summarised as follows:

(i) *Open area*: open space, no tall trees or buildings in path, plot of land cleared for 300–400 m ahead, e.g. farmland, rice fields, open fields.
(ii) *Suburban area*: village or highway scattered with trees and houses, some obstacles near the mobile but not very congested.

(iii) *Urban area*: built-up city or large town with large buildings and houses with two or more storeys, or larger villages with close houses and tall, thickly grown trees.

Okumura takes urban areas as a reference and applies correction factors for conversion to the other classifications. This is a sensible choice, since such areas avoid the large variability present in suburban areas (see Chapter 9) and yet include the effects of obstructions better than could be done with open areas. A series of terrain types is also defined for when such information is available. Quasi-smooth terrain is taken as the reference and correction factors are added for the other types of terrain.

Okumura's predictions of median path loss are usually calculated using Hata's approximations as follows [9]:

$$\begin{array}{ll} \text{urban areas} & L_{dB} = A + B \log R - E \\ \text{suburban areas} & L_{dB} = A + B \log R - C \\ \text{open areas} & L_{dB} = A + B \log R - D \end{array} \qquad (11.5)$$

where

$$\begin{array}{ll} A = 69.55 + 26.16 \log f_c - 13.82 \log h_b & \\ B = 44.9 - 6.55 \log h_b & \\ C = 2(\log(f_c/28))^2 + 5.4 & \\ D = 4.78(\log f_c)^2 + 18.33 \log f_c + 40.94 & \\ E = 3.2 (\log(11.75\, h_m))^2 - 4.97 & \text{for large cities, } f_c \geq 300 \text{ MHz} \\ E = 8.29 (\log(1.54\, h_m))^2 - 1.1 & \text{for large cities, } f_c < 300 \text{ MHz} \\ E = (1.1 \log f_c - 0.7)h_m - (1.56 \log f_c - 0.8) & \text{for medium to small cities} \end{array} \qquad (11.6)$$

The model is valid only for 150 MHz $\leq f_c \leq$ 1500 MHz, 30 m $\leq h_b \leq$ 200 m, 1 m $< h_m <$ 10 m and $R >$ 1 km. The path loss exponent is given by $B/10$, which is a little less than 4, decreasing with increasing base station antenna height.

Base station antenna height h_b is defined as the height above the average ground level in the range 3–10 km from the base station; h_b may therefore vary slightly with the direction of the mobile from the base. The height gain factor varies between 6 dB octave^{-1} and 9 dB octave^{-1} as the height increases from 30 m to 1 km. Measurements also suggest that this factor depends upon range.

Okumura found that mobile antenna height gain is 3 dB octave^{-1} up to $h_m = 3$ m and 8 dB octave^{-1} beyond. It depends partially upon urban density, apparently as a result of the effect of building heights on the angle-of-arrival of wave energy at the mobile and the consequent shadow loss variation. Urban areas are therefore subdivided into large cities and medium/small cities, where an area having an average building height in excess of 15 m is defined as a large city.

Other correction factors are included in Okumura's original work for the effects of street orientation (if an area has a large proportion of streets which are either radial or tangential to the propagation direction) and a fine correction for rolling hilly terrain (used if a large proportion of streets are placed at either the peaks or valleys of the terrain undulations). Application of the method involves first finding the basic median field strength in concentric circles around the base station, then amending them according to the terrain and clutter correction graphs.

Okumura's predictions have been found useful in many cases [11], particularly in suburban areas. However, other measurements have been in disagreement with these predictions; the reasons for error are often cited as the difference in the characteristics of the area under test with Tokyo. Other authors such as Kozono [12] have attempted to modify Okumura's method to include a measure of building density, but such approaches have not found common acceptance.

The Okumura–Hata model, together with related corrections, is probably the single most common model used in designing real systems. Several commercial prediction tools essentially rely on variations of this model, optimised for the particular environments they are catering for, as the basis of their predictions.

11.2.3.3 The COST 231–Hata model

The Okumura–Hata model for medium to small cities has been extended to cover the band 1500 MHz < f_c < 2000 MHz [13]:

$$L_{dB} = F + B \log R - E + G \tag{11.7}$$

where

$$F = 46.3 + 33.9 \log f_c - 13.82 \log h_b \tag{11.8}$$

E is as defined in Eqn. 11.6 for medium to small cities and

$$G = \begin{cases} 0 \text{ dB medium-sized cities and suburban areas} \\ 3 \text{ dB metropolitan areas} \end{cases} \tag{11.9}$$

11.2.3.4 Environment categories

In an empirical model, it is crucial to correctly classify the environment in which the system is operating. The models assume that the characteristics of the environment to be predicted are sufficiently similar to those where the original measurements were taken that the propagation loss at a given distance will be similar. Good results will therefore only be obtained if the correct classification is chosen. The categories of environment should also be sufficiently numerous that the properties of different locations classed within the same category are not too variable. The decision as to which category an environment fits into is usually purely subjective and may vary between individuals and countries. For example, the Okumura–Hata model uses four categories: large cities, medium-small cities, suburban areas and open areas. Since the original measurements were made in Tokyo, the model relies on other parts of the world having characteristics which are somehow similar to those in Tokyo. Although this is an extremely questionable assumption, it is nevertheless true that the model has been applied to many successful system designs.

Many more detailed schemes exist for qualitative classification of land usage. Schemes often correspond to sources of data, such as satellite remote-sensing data which classifies land according to the degree of scattering experienced at various wavelengths. This at least avoids the need for ambiguous judgements to be made. Nevertheless, there is no guarantee that there is any one-to-one mapping between the propagation characteristics and such measures of land usage. In order to find more appropriate parameters, the growing tendency in macrocellular propagation is towards models which have a physical basis, and these are examined in the next section.

11.2.4 Physical models

Although empirical models have been extensively applied with good results, they suffer from a number of disadvantages:

- they can only be used over parameter ranges included in the original measurement set;
- environments must be classified subjectively according to categories such as urban, which have different meanings in different countries;
- they provide no physical insight into the mechanisms by which propagation occurs.

The last point is particularly significant, as empirical models are unable to account for factors such as an unusually large building or hill which may greatly modify propagation in particular locations.

Although the plane earth model has a path loss exponent close to that observed in actual measurements (i.e. 4), the simple physical situation it describes is rarely applicable in practice. The mobile is almost always operated (at least in macrocells) in situations where it does not have a line-of-sight path to either the base station or to the ground reflection point, so the two-ray situation which the plane earth model relies upon is hardly ever applicable. In order to find a more satisfactory physical propagation model, we examine diffraction as a potential mechanism.

11.2.4.1 The Ikegami model

This model attempts to produce an entirely deterministic prediction of field strengths at specified points [14]. Using a detailed map of building heights, shapes and positions, ray paths between the transmitter and receiver are traced, with the restriction that only single reflections from walls are accounted for. Diffraction is calculated using a single edge approximation at the building nearest the mobile, and wall reflection loss is assumed to be fixed at a constant value. The two rays (reflected and diffracted) are power summed, resulting in the following approximate model:

$$L_E = 10 \log f_c + 10 \log(\sin \phi) + 20 \log(h_0 - h_{\mathrm{m}}) \\ - 10 \log w - 10 \log\left(1 + \frac{3}{L_r^2}\right) - 5.8 \tag{11.10}$$

where ϕ is the angle between the street and the direct line from base to mobile and $L_r = 0.25$ is the reflection loss. The analysis assumes that the mobile is in the centre of the street. The model therefore represents the situation illustrated in Figure 11.7. It further assumes that the elevation angle of the base station from the top of the knife edge is negligible in comparison with the diffraction angle down to the mobile level.

A comparison of the results of this model with measurements at 200, 400 and 600 MHz shows that the general trend of variations along a street is accounted for successfully. The predictions suggest that field strength is broadly independent of a mobile's position across the street. This is confirmed by the mean values of a large number of measurements, although the spread of values is rather high. Acceptable agreement is also obtained for variations with street angle and width.

Although it accounts reasonably well for close-in variations in field strength, it is a flawed assumption that base station antenna height does not affect propagation. The same assumption means that the free-space path loss exponent is assumed, so the model tends to underestimate loss at large distances. Similarly, the variation with frequency is underestimated compared with measurements.

11.2.4.2 Rooftop diffraction

When a macrocell system is operated in a built-up area with reasonably flat terrain, the dominant mode of propagation is multiple diffraction over the building rooftops. Diffraction can occur around the sides of individual buildings, but this tends to become highly attenuated over reasonable distances since many interactions with individual buildings are involved.

The diffraction angle over most of the rooftops is small for typical base

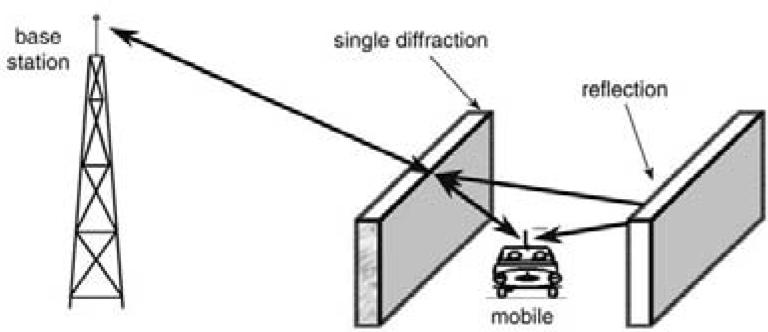


Figure 11.7 Physical interpretation of Ikegami model

station heights and distances, usually less than 1°. In these cases the diffraction is largely unaffected by the particular shape of the obstacles, so it is appropriate to represent the buildings by equivalent knife edges. The one exception to this is diffraction from the final building at which the wave is diffracted from rooftop level down to the street-level antenna of the mobile (Figure 11.8). It is usual to separate these processes into multiple diffraction across the first $(n - 1)$ buildings, treated as knife edges, and a final building which can be treated either as a knife edge or as some more complex shape for which the diffraction coefficient is known.

Special methods have been developed to enable reasonably rapid calculation of the multiple diffraction integral for cases where accurate results are required and where the necessary data on the building positions and heights is available [15]. Such data is usually too expensive for general use in macrocells. Two simplified solutions with reduced data and computational requirements are therefore examined here.

11.2.4.3 The flat edge model

In this model [16], the situation is simplified by assuming all of the buildings to be of equal height and spacing. The values used can be average values for the area under consideration, or can be calculated individually for each direction from the base station if the degree of urbanisation varies significantly. The geometry is shown in Figure 11.9, illustrating the following parameters additional to the definitions in Section 11.2.2, distance r_1 from the base station to the first building (m) and elevation angle a of the base station antenna from the top of the final building (radians). In Figure 11.9, buildings are arranged normal to the great circle path. Since this will not normally be the case in practice, the value of w used should be an effective one to account for the longer paths between the buildings for oblique incidence.

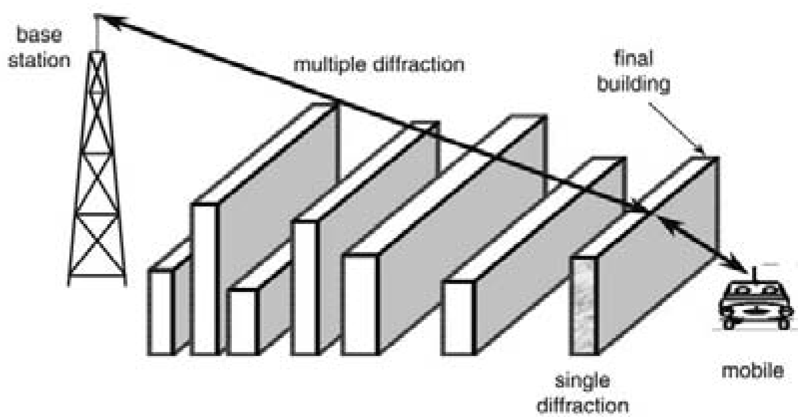


Figure 11.8 Multiple diffraction over building rooftops

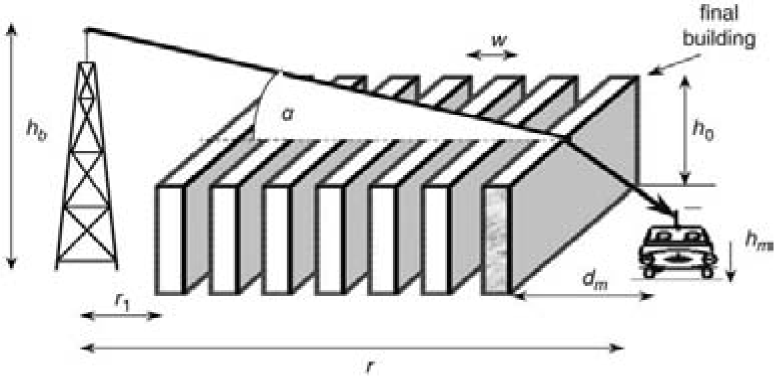


Figure 11.9 Geometry for the flat edge model

The excess path loss is then expressed as:

$$L_{ex} = L_{n-1}(t)L_{ke} \tag{11.11}$$

where L_{ke} accounts for single edge diffraction over the final building and L_{n-1} accounts for multiple diffraction over the remaining $(n - 1)$ buildings. Provided $r_1 >> nw$ (i.e. the base station is relatively distant from the first building), the multiple diffraction integral can be completely solved in this special case.

The result is that L_{n-1} is a function of a parameter t only, where t is given by:

$$t = -\alpha\sqrt{\frac{\pi w}{\lambda}} \tag{11.12}$$

It is given by the following formula:

$$L_n(t) = \frac{1}{n}\sum_{m=0}^{n-1} L_m(t)F_s(-jt\sqrt{n-m}) \text{ for } n \geq 1 \quad L_0(t) = 1 \tag{11.13}$$

where

$$F_s(jx) = \frac{e^{-jx^2}}{\sqrt{2j}}\left\{\left[S\left(x\sqrt{\frac{2}{\pi}}\right)+\frac{1}{2}\right]+j\left[C\left(x\sqrt{\frac{2}{\pi}}\right)+\frac{1}{2}\right]\right\} \tag{11.14}$$

and $S(.)$ and $C(.)$ are the standard Fresnel sine and cosine integrals (see Saunders [1] Appendix B). This formulation is extremely quick and simple to compute and it applies for any values of α, even when the base station antenna height is below the rooftop level. The number of buildings can be increased to extremely high values with no difficulties.

The flat edge model may be calculated either directly from Eqn. 11.13, or the results may be estimated from the prediction curves in Figure 11.10, which show the cases where $h_b \geq h_0$. An alternative approach is to use the approximate formula

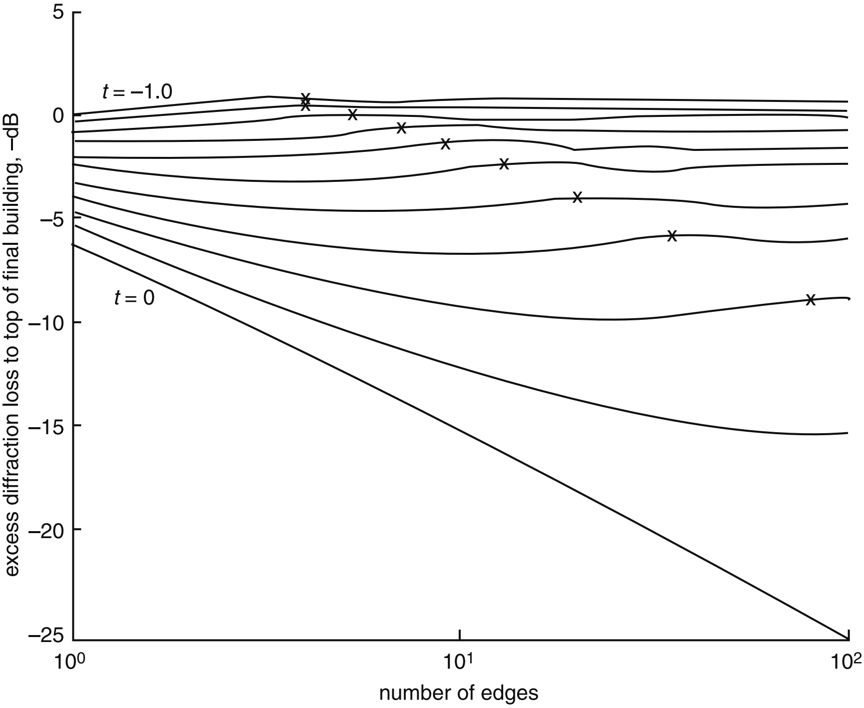


Figure 11.10 Flat edge model prediction curves for elevated base antennas: curves relate to t varying from 0 to –1 in steps of 0.1; the crosses indicate the number of edges required for a settled field according to Eqn. 11.20

$$L_n(t) = -20 \log A_n(t)$$
$$= -(c_1 + c_2 \log n)\log(-t) - (c_3 + c_4 \log n) \text{ dB} \qquad (11.15)$$

where $c_1 = 3.29$, $c_2 = 9.90$, $c_3 = 0.77$, $c_4 = 0.26$. This approximates the value of Eqn. 11.13 with an accuracy of better than ±1.5 dB for $1 \le n \le 100$ and $-1 \le t < 0$. It also enables us to investigate the behaviour of the effective path loss exponent for the flat edge model, since for fixed n, we can rewrite Eqn. 11.15 with L being the path loss as a power ratio as:

$$L \propto (-t)^{-(c_2/10)\log n} = \left(\alpha\sqrt{\frac{\pi w}{\lambda}}\right)^{-(c_2/10)\log n} \approx \left(\frac{h_b - h_0}{r}\sqrt{\frac{\pi w}{\lambda}}\right)^{-(c_2/10)\log n} \qquad (11.16)$$

where the approximation holds if $(h_b - h_0) \ll r$. This is the excess field strength, so the overall path loss exponent, including an extra 2 from the free-space part of the loss, is as follows:

$$\text{path loss exponent} = 2 + (c_2/10)\log n \qquad (11.17)$$

This expression is illustrated in Figure 11.11, where it is apparent that, for reasonably large numbers of buildings, the path loss exponent for the flat edge

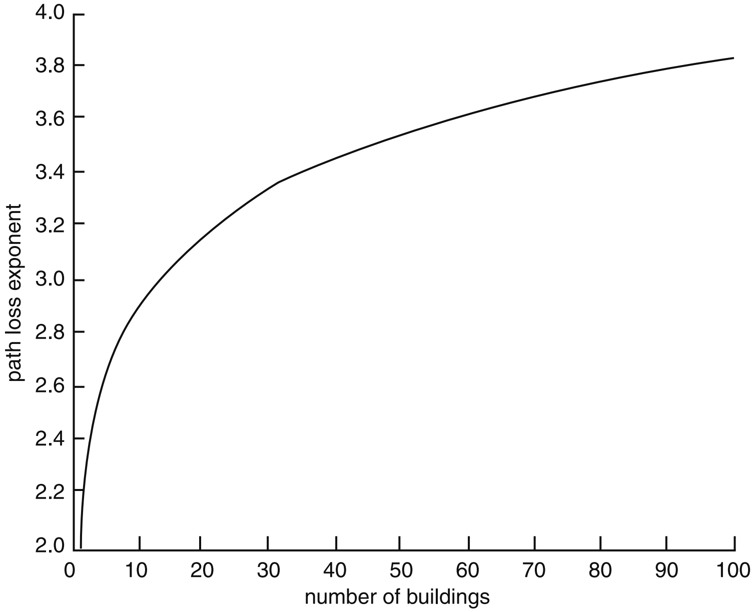


Figure 11.11 Path loss exponent for the flat edge model

model is close to 4, just as observed in practical measurements. More generally, we can state: *Multiple building diffraction accounts for the variation of path loss with range which is observed in measurements.*

Figure 11.10 shows that, for $h_b > h_0$, i.e. $t < 0$, the field at the top of the final building eventually settles to a constant value as the number of edges increases. This number, n_s, corresponds to the number required to fill the first Fresnel zone around the ray from the base station to the final building. The first Fresnel zone radius r_1 is given approximately by:

$$r_1 = \sqrt{\lambda s} \tag{11.18}$$

where s is the distance along the ray from the field point. Hence, for small α:

$$\alpha = \tan^{-1}\frac{r_1}{n_s w} \approx \frac{\sqrt{\lambda n_s w}}{n_s w} \tag{11.19}$$

So

$$n_s \approx \frac{\lambda}{\alpha^2 w} = \frac{\pi}{t^2} \tag{11.20}$$

This is marked in Figure 11.10. Note that the number of edges required for settling rises very rapidly with decreasing α. Whenever $\alpha \le 0$ the field does not settle at all, but decreases monotonically for all n.

The flat edge model is completed by modelling the final building diffraction loss and the reflections from the buildings across the street using the Ikegami model from Section 11.2.4.1. Thus the total path loss is given by:

$$L_T = L_n(t) + L_F + L_E \tag{11.21}$$

where $L_n(t)$ can be found from Eqn. 11.13, Figure 11.10 or Eqn. 11.15, L_F is the free space loss (2.7) and L_E is given in Eqn. 11.10.

11.2.4.4 The Walfisch–Bertoni model

This model can be considered as the limiting case of the flat edge model when the number of buildings is sufficient for the field to settle, i.e. $n \geq n_s$. The multiple diffraction process was investigated by Walfisch and Bertoni [17] using a numerical evaluation of the Kirchhoff–Huygens integral and a power law formula is fitted to the results for the settled field. The Walfisch–Bertoni model was the first to actually demonstrate that multiple building diffraction accounts for the variation of distance with range which is observed in measurements.

The settled field approximation is as follows:

$$A_{\text{settled}}(t) \approx 0.1\left(\frac{\alpha}{0.03}\sqrt{\frac{w}{\lambda}}\right)^{0.9} = 0.1\left(\frac{-t}{0.03}\right)^{0.9} \tag{11.22}$$

This is only valid for $0.03 \leq t \leq 0.4$. For large ranges, we can again put:

$$t \approx -\alpha\sqrt{\frac{\pi w}{\lambda}} \approx -\frac{h_b - h_m}{r}\sqrt{\frac{\pi w}{\lambda}} \tag{11.23}$$

Hence $L_{\text{settled}} \propto r^{-1.8}$. The free-space loss is proportional to r^{-2}, so this model predicts that the total propagation loss is proportional to $r^{-3.8}$, which is close to the r^{-4} law which is commonly assumed in empirical models and found in measurements. A single knife-edge approximation with a reflection from the building opposite is again used, just as in the Ikegami model, to account for the diffraction from the final building. The complete model is expressed as:

$$L_{ex} = 57.1 + L_A + \log f_c + 18 \log R - 18 \log(h_b - h_0) - 18 \log\left[1 - \frac{R^2}{17(h_b - h_0)}\right] \tag{11.24}$$

where

$$L_A = 5 \log\left[\left(\frac{w}{2}\right) + (h_0 - h_m)^2\right] - 9 \log w + 20 \log\left\{\tan^{-1}\left[\frac{2(h_0 - h_m)}{w}\right]\right\} \tag{11.25}$$

The use of the settled field approximation requires that large numbers of buildings are present, particularly when α is small. Despite this limitation, the Walfisch–Bertoni model is the first to have accounted for observed path loss variation using realistic physical assumptions rather than relying upon forcing agreement using propagation models of entirely different situations.

11.2.4.5 COST 231 / Walfisch–Ikegami model

The Walfisch–Bertoni model for the settled field has been combined with the Ikegami model for diffraction down to street level and some empirical correction factors to improve agreement with measurements in a single integrated model by the COST 231 project [13].

For nonline-of-sight conditions the total loss is given by:

$$L = L_F + L_{msd} + L_{sd} \tag{11.26}$$

where L_F is the free space loss, L_{msd} accounts for multiple knife-edge diffraction to the top of the final building and L_{sd} accounts for the single diffraction and scattering process down to street level. L is given a minimum value of L_F in case the other terms become negative. The individual terms are:

$$L_{sd} = -16.9 + 10\log f_c + 10\log\frac{(h_0 - h_m)^2}{w_m} + L(\phi) \tag{11.27}$$

where w_m is the distance between the building faces on either side of the street containing the mobile (typically $w_m = w/2$), and the final term accounts for street orientation at an angle to the great circle path:

$$L(\phi) = \begin{cases} -10 + 0.354\phi & \text{for } 0° < \phi < 35° \\ 2.5 + 0.075(\phi - 35°) & \text{for } 35° \le \phi < 55° \\ 4.0 + 0.114(\phi - 55°) & \text{for } 55° \le \phi < 90° \end{cases} \tag{11.28}$$

Finally, the rooftop diffraction term is given by

$$L_{msd} = L_{bsh} + k_a + k_d \log R + k_f \log f_c - 9\log w \tag{11.29}$$

where

$$L_{bsh} = \begin{cases} -18\log[1 + (h_b - h_0)] & \text{for } h_b > h_0 \\ 0 & \text{for } h_b \le h_0 \end{cases} \tag{11.30}$$

$$k_a = \begin{cases} 54 & \text{for } h_b > h_0 \\ 54 - 0.8(h_b - h_0) & \text{for } R \ge 0.5\text{ km and } h_b \le h_0 \\ 54 - 0.8\dfrac{(h_b - h_0)R}{0.5} & \text{for } R < 0.5\text{ km and } h_b \le h_0 \end{cases} \tag{11.31}$$

$$k_d = \begin{cases} 18 & \text{for } h_b > h_0 \\ 18 - 15\dfrac{(h_b - h_0)}{h_0} & \text{for } h_b \le h_0 \end{cases} \tag{11.32}$$

$$k_f = -4 + 0.7\left(\frac{f_c}{925} - 1\right) \quad \text{for medium-sized city and}$$

$$\text{suburban areas with medium tree density} \qquad (11.33)$$

$$k_f = -4 + 1.5\left(\frac{f_c}{925} - 1\right) \quad \text{for metropolitan centres}$$

For approximate work, the following parameter values can be used:

$$h_0 = \begin{cases} 3n_{\text{floors}} & \text{for flat roofs} \\ 3n_{\text{floors}} + 3 & \text{for pitched roofs} \end{cases} \qquad (11.34)$$

$$w = 20 - 50\ \text{m}, \quad d_m = w/2, \quad \phi = 90°$$

where n_{floors} is the number of floors in the building. The model is applicable for 800 MHz $\le f_c \le$ 2000 MHz, 4 m $\le h_b \le$ 50 m, 1 m $\le h_m \le$ 3 m and 0.02 km $\le R \le$ 5 km.

An alternative approach is to replace the L_{msd} term by $L_n(t)$ from the flat edge model. This would enable the path loss exponent to vary according to the number of buildings and to be uniformly valid for $h_b \le h_0$. Note, however, that for very low base station antennas other propagation mechanisms, such as diffraction around vertical building edges and multiple reflections from building walls, are likely to be significant.

11.2.5 Computerised planning tools

The methods described in this Chapter are most often implemented for practical planning within computer software. The development of such software has been motivated and enabled by a number of factors:

- the enormous increase in the need to plan cellular systems accurately and quickly
- the development of fast, affordable computing resources
- the development of geographical information systems, which index data on terrain, clutter and land usage in an easily accessible and manipulable form.

Such techniques have been implemented in a wide range of commercially available and company-specific planning tools. Although most are based on combined empirical and simple physical models, it is anticipated that there will be progressive evolution in the future towards more physical or physical–statistical methods as computing resources continue to cheapen, clutter data improves in resolution and cost and as research develops into numerically efficient path loss prediction algorithms

11.2.6 Conclusions

The accuracy of the path loss predictions is crucial in determining whether a particular mobile system design will be viable. In macrocells, empirical models have been used with great success, but deterministic physical models are being increasingly investigated as a means of improving accuracy, based on the use of multiple rooftop diffraction as the key propagation mechanism. This accuracy comes at the expense of increased input data requirements and computational complexity. Another generation of models is expected to appear which combines sound physical principles with statistical parameters, which can economically be obtained in order to provide the optimum balance between accuracy and complexity.

11.3 Shadowing

11.3.1 Introduction

The models of macrocellular path loss described in Section 11.2 assume that path loss is a function only of parameters such as antenna heights, environment and distance. The predicted path loss for a system operated in a particular environment will therefore be constant for a given base-to-mobile distance. In practice, however, the particular clutter (buildings, trees) along a path at a given distance will be different for every path, causing variations with respect to the nominal value given by the path loss models, as shown by the large scatter evident in the measurements in Figure 11.5. Some paths will suffer increased loss, while others will be less obstructed and have an increased signal strength, as illustrated in Figure 11.12. This phenomenon is called shadowing or slow fading. It is crucial to account for this in order to predict the reliability of coverage provided by any mobile cellular system.

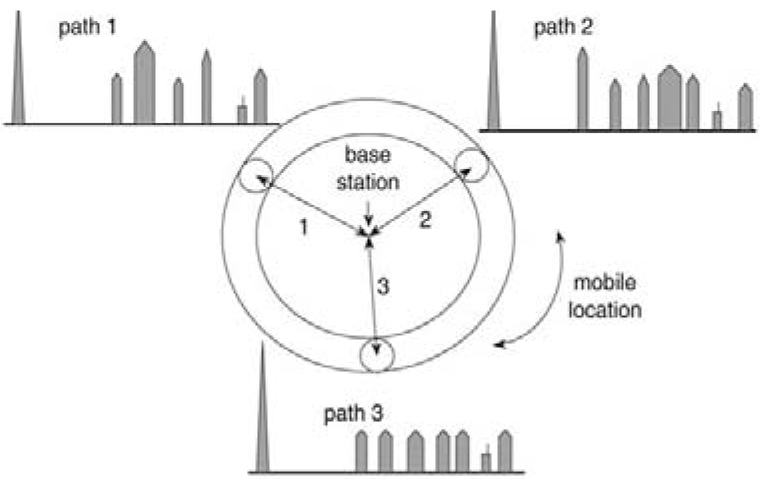


Figure 11.12 Variation of path profiles encountered at a fixed range from a base station

11.3.2 Statistical characterisation

If a mobile is driven around a base station (BS) at a constant distance, then the local mean signal level will typically appear similar to Figure 11.13, after subtracting the median (50 per cent) level in decibels. If the probability density function of the signal is then plotted, a typical result is Figure 11.14. The distribution of the underlying signal powers is log-normal; that is, the signal measured in decibels has a normal distribution. The process by which this distribution comes about is known as shadowing or slow fading. The variation occurs over distances comparable to the widths of buildings and hills in the region of the mobile, usually tens or hundreds of metres.

The standard deviation of the shadowing distribution (in decibels) is known as the location variability, σ_L. The location variability varies with frequency, antenna heights and the environment; it is greatest in suburban areas and smallest in open areas, and is usually in the range 5–12 dB (Section 11.3.4); the value in Figures 11.2 and 11.3 is 8 dB.

In practice, not all of the losses will contribute equally, with those nearest the mobile end being most likely to have an effect in macrocells. Moreover, as shown in Chapter 8, the contributions of individual diffracting obstacles cannot simply be added, so the assumption of independence is not strictly valid. Nevertheless, when the different building heights, spacings and construction methods are taken into account, along with the attenuation due to trees, the resultant distribution function is indeed very close to log-normal [18, 19].

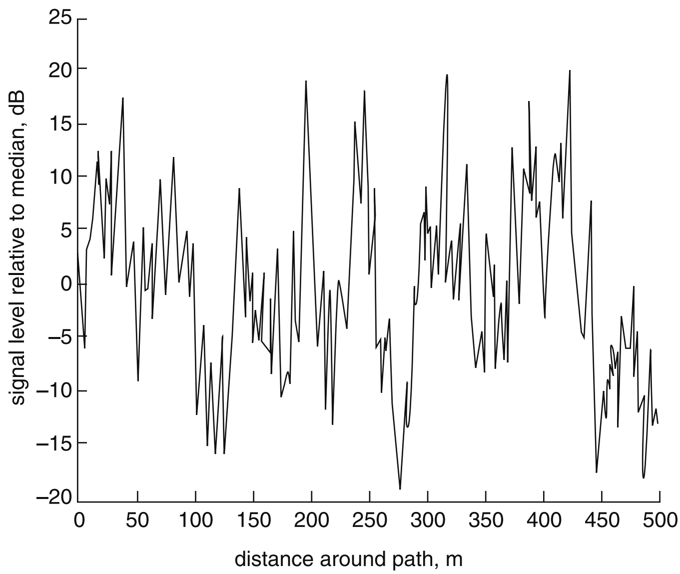


Figure 11.13 Typical variation of shadowing with mobile position at fixed BS distance

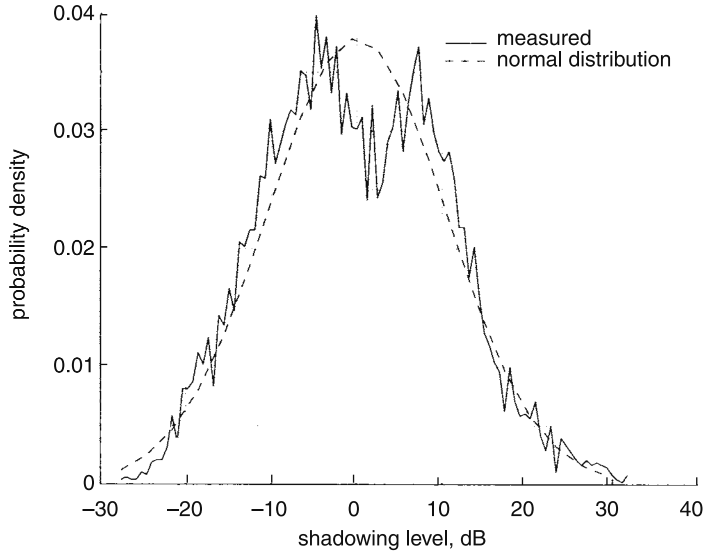


Figure 11.14 Probability density function of shadowing; measured values are produced by subtracting the empirical model shown in Figure 11.5 from the total path loss measurements; theoretical values come from the log-normal distribution

11.3.3 Impact on coverage

When shadowing is included, the total path loss becomes a random variable, given by:

$$L = L_{50} + L_S \tag{11.35}$$

where L_{50} is the level not exceeded at 50 per cent of locations at a given distance, as predicted by any standard path loss model (the local median path loss). L_S is the shadowing component, a zero-mean Gaussian random variable with standard deviation σ_L. The probability density function of L_S is therefore given by the standard Gaussian formula:

$$p(L_S) = \frac{1}{\sigma_L\sqrt{2\pi}} \exp\left[-\frac{L_S^2}{2\sigma_L^2}\right] \tag{11.36}$$

In order to provide reliable communications at a given distance, therefore, an extra fade margin has to be added in to the link budget according to the reliability required from the system. In Figure 11.15 the cell range would be around 9.5 km if shadowing were neglected, then only 50 per cent of locations at the edge of the cell would be properly covered. By adding the fade margin, the cell radius is reduced to around 5.5 km but the reliability is greatly increased.

The probability that the shadowing increases the median path loss by at least z dB is then given by:

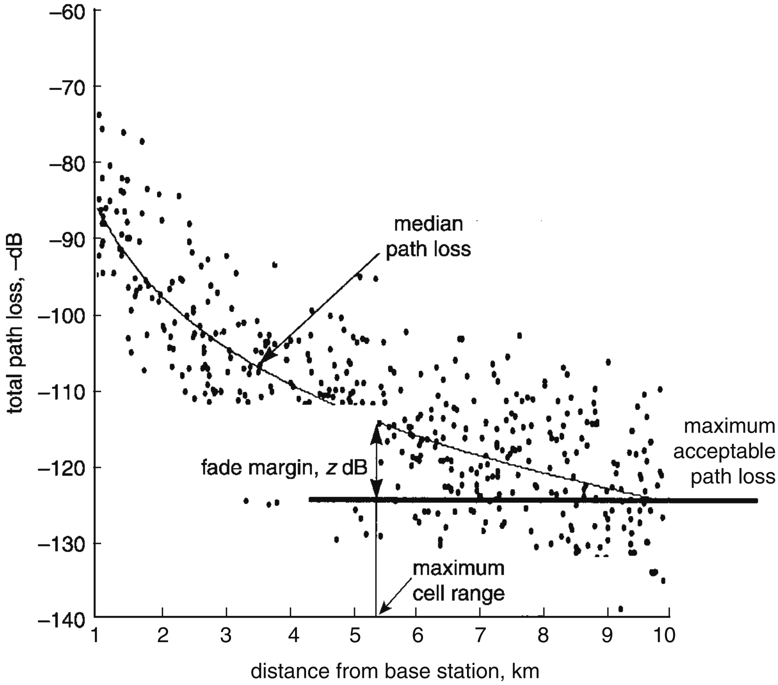


Figure 11.15 Effect of shadowing margin on cell range

$$\Pr[L_S > z] = \int_{L_S=z}^{\infty} p(L_S) dL_S = \int_{L_S=z}^{\infty} \frac{1}{\sigma_L\sqrt{2\pi}} \exp\left[-\frac{L_S^2}{2\sigma_L^2}\right] dL_S \qquad (11.37)$$

It is then convenient to normalise the variable z by the location variability:

$$\Pr[L_S > z] = \int_{x=z/\sigma_L}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] dx = Q\left(\frac{z}{\sigma_L}\right) \qquad (11.38)$$

where the $Q(.)$ function is the complementary cumulative normal distribution. Values for Q are widely tabulated, or they can be calculated from erfc(.), the standard cumulative error function, using:

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_{x=1}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{2}\,\mathrm{erfc}\left(\frac{t}{\sqrt{2}}\right) \qquad (11.39)$$

$Q(t)$ can then be used to evaluate the shadowing margin needed for any location variability in accordance with Eqn. 11.39 by putting $t = z/\sigma_L$, as described in the following example.

Example

A mobile communications system is to provide 90 per cent successful communications at the fringe of coverage. The system operates in an environment where propagation can be described by a plane earth model plus a 20 dB clutter factor, with shadowing location variability of 6 dB. The maximum acceptable path loss for the system is 140 dB. Antenna heights for the system are $h_m = 1.5$ m and $h_b = 30$ m. Determine the range of the system. How is this range modified if the location variability increases to 8 dB?

Solution

The total path loss is given by the sum of the plane earth loss, the clutter factor and the shadowing loss:

$$\begin{aligned} L_{total} &= L_{PEL} + L_{clutter} + L_S \\ &= 40 \log r - 20 \log h_m - 20 \log h_b + 20 + L_S \end{aligned}$$

To find L_S, we take the value of $t = z/\sigma_L$ for which the path loss is less than the maximum acceptable value for at least 90 per cent of locations, or when $Q(t) = 10$ per cent $= 0.1$. From tables this occurs when $t \approx 1.25$. Multiplying this by the location variability gives

$$L_S = z = t\sigma_L = 1.25 \times 6 = 7.5 \text{ dB}$$

Hence

$$\log r = \frac{140 + 20 \log 1.5 + 20 \log 30 - 20 - 7.5}{40} = 3.64$$

So the range of the system is $r = 10^{3.64} = 4.4$ km. If σ_L rises to 8 dB, the shadowing margin $L_s = 10$ dB and $d = 3.8$ km. Thus shadowing has a decisive effect on system range.

In the example above, the system was designed so that 90 per cent of locations at the edge of the cell have acceptable coverage. Within the cell, although the value of shadowing exceeded for 90 per cent of locations is the same, the value of the total path loss will be less, so a greater percentage of locations will have acceptable coverage. Most mobiles will therefore experience considerably better coverage. It is perhaps more appropriate to design the system in terms of the coverage probability experienced over the whole cell. Such analysis may be found in Reference [1] or in Reference [20].

11.3.4 Location variability

Figure 11.16 shows the variation of the location variability with frequency, as measured by several studies. It is clear that there is a tendency for σ_L to increase with frequency and that it depends upon the environment. Suburban cases tend to provide the largest variability, due to the large variation in the characteristics of local clutter. Urban situations have rather lower variability, although the overall path loss would be higher. No consistent variation with range has been

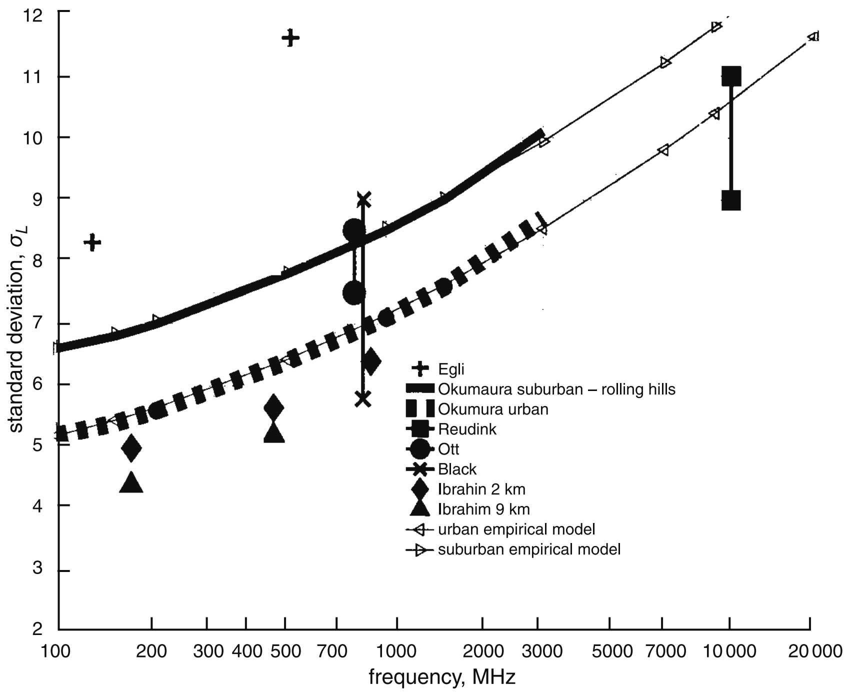


Figure 11.16 Location variability versus frequency; measured values from Okumura [8], *Egli* [6], *Reudink* [22], *Ott* [23], *Black* [24] *and Ibrahim* [21]*; after Jakes* [20]

reported; the variations in the measurements at 2 km and 9 km [21] are due to differences in the local environment.

Figure 11.16 also includes plots of an empirical relationship fitted to the Okumura curves [22] and chosen to vary smoothly up to 20 GHz. This is given by

$$\sigma_L = 0.65(\log f_c)^2 - 1.3 \log f_c + A \tag{11.40}$$

where A = 5.2 in the urban case and 6.6 in the suburban case. Note that these values only apply to macrocells.

11.3.5 Correlated shadowing

So far in this Section, the shadowing on each propagation path from base station to mobile has been considered independently, but this approach must be modified for accurate calculations. Consider the situation illustrated in Figure 11.17. Two mobiles are separated by a small distance r_m and each can receive signals from two base stations. Alternatively, the two mobile locations may represent two positions of a single mobile, separated by some time interval. Each of the paths between the base and mobile locations is marked with the

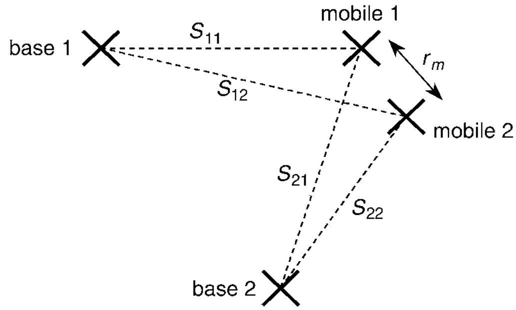


Figure 11.17 Definitions of shadowing correlations

value of the shadowing associated with that path. Each of the four shadowing paths can be assumed to be log-normal, so the shadowing values S_{11}, S_{12}, S_{21} and S_{22} are zero-mean Gaussian random variables when expressed in decibels. However, they are not independent of each other, since the four paths may include many of the same obstructions in the path profiles. There are two types of correlation to distinguish:

(i) Correlations between two mobile locations, receiving signals from a single base station, such as between S_{11} and S_{12} or between S_{21} and S_{22}. These are serial correlations, or simply the autocorrelation of the shadowing experienced by a single mobile as it moves.
(ii) Correlations between two base station locations as received at a single mobile location, such as between S_{11} and S_{21} or between S_{12} and S_{22}. These are site-to-site correlations or simply cross correlations.

Both of these correlations have a profound impact on system performance, especially coverage, handover and interference statistics, power control and capacity. A detailed treatment can be found in Saunders [1].

11.3.6 Conclusions

The inclusion of shadowing into propagation models transforms the coverage radius of a cell from a fixed, predictable value into a statistical quantity. The properties of this quantity affect the coverage and capacity of a system in ways which can be predicted using the techniques introduced in this Chapter. In particular, the shadowing affects the dynamics of signal variation at the mobile, the percentage of locations which receive sufficient power and the percentage which receive sufficient signal-to-interference ratio.

11.4 Microcells

11.4.1 Introduction

The deployment of microcells is motivated by a desire to reduce cell sizes in areas where large numbers of users require access to the system. Serving these users with limited radio spectrum requires frequencies to be reused over very short distances, with each cell containing only a reduced number of users. This could in principle be achieved with base station antennas at the same heights as in macrocells, but this would increase the costs and planning difficulties substantially. In a microcell the base station antenna is typically at about the same height as lamp posts in a street (3–6 m above ground level), although the antenna is more usually mounted onto the side of a building, Figure 11.18. Coverage, typically over a few hundred metres, is then determined mostly by the specific locations and electrical characteristics of the surrounding buildings, with cell shapes being far from circular. Pattern shaping of the base station antenna can yield benefits in controlling interference, but it is not the dominant factor in determining the cell shape. The dominant propagation mechanisms are free-space propagation plus multiple reflection and scattering within the cell's desired coverage area, together with diffraction around the vertical edges of buildings and over rooftops, which becomes significant when determining interference between co-channel cells. Microcells thus make increased use of the potential of the environment surrounding the base station to carefully control the coverage area and hence to manage the interference between sites. More general information on microcell systems is available in Greenstein [25], Sarnecki [26] and Madfors [27].

11.4.2 Dual-slope empirical models

In order to model the path loss in microcells, empirical models could be used in principle. However, measurements, e.g. Green [28], indicate that a simple power

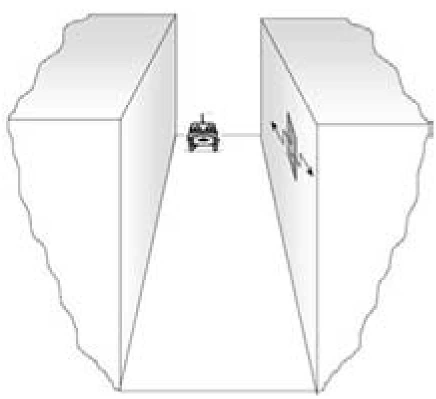


Figure 11.18 A microcell in a built-up area

law path loss model cannot usually fit measurements with good accuracy. A better empirical model in this case is a dual-slope model (see also Chapter 9). Two separate path loss exponents are used to characterise the propagation, together with a breakpoint distance of a few hundred metres between them where propagation changes from one regime to the other. In this case the path loss is modelled as:

$$\frac{1}{L} = \begin{cases} \dfrac{k}{r^{n_1}} & \text{for } r \le r_b \\ \dfrac{k}{(r/r_b)^{n_2} r_b^{n_1}} & \text{for } r > r_b \end{cases} \tag{11.41}$$

or, in decibels:

$$L = \begin{cases} 10n_1 \log r + L_1 & \text{for } r \le r_b \\ 10n_2 \log \dfrac{r}{r_b} + 10n_1 \log r_b + L_1 & \text{for } r > r_b \end{cases} \tag{11.42}$$

where L_1 is the reference path loss at $r = 1$ m, r_b is the breakpoint distance, n_1 is the path loss exponent for $r \le r_b$ and n_2 is the path loss exponent for $r > r_b$. In order to avoid the sharp transition between the two regions of a dual-slope model, it can also be formulated according to an approach suggested by Harley [29]:

$$\frac{1}{L} = \frac{k}{r^{n_1}(1 + (r/r_b))^{n_2 - n_1}} \tag{11.43}$$

This can be considered in two regions: for $r << r_b$, $(1/L) \approx kr^{-n_1}$, and for $r >> r_b$, $(1/L) \approx k(r/r_b)^{-n_2}$. Hence the path loss exponent is again n_1 for short distances and n_2 for larger distances. The model is conveniently expressed in decibels as:

$$L = L_1 + 10n_1 \log r + 10(n_2 - n_1)\, log\left(1 + \frac{r}{r_b}\right) \tag{11.44}$$

where L_1 is a reference value for the loss at 1 m. Figure 11.19 compares Eqns. 11.42 and 11.44.

Typical values for the path loss exponents are found by measurement to be around $n_1 = 2$ and $n_2 = 4$, with breakpoint distances of 200–500 m, but it should be emphasised that these values vary greatly between individual measurements. See for example Chia [30], Green [28], Xia [31] and Bultitude [32]. In order to plan the locations of microcells effectively, it is important to ensure that co-channel cells have coverage areas which do not overlap within the breakpoint distance. The rapid reduction of signal level beyond the breakpoint then produces a large carrier-to-interference ratio, which can be exploited to maximise system capacity.

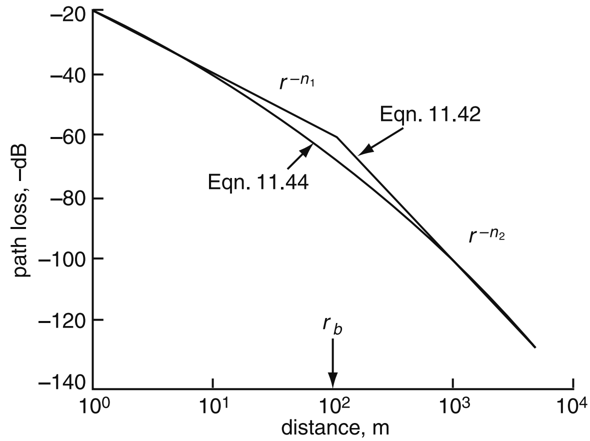


Figure 11.19 Dual-slope empirical loss models, $n_1 = 2$, $n_2 = 4$, $r_b = 100$ m and $L_1 = 20$ dB

11.4.3 Physical models

In creating physical models for microcell propagation, it is useful to distinguish between line-of-sight and non line-of-sight situations. We shall see that it is possible to make some reasonable generalisations about the LOS cases; NLOS cases require more site-specific information. Obstructed paths tend to suffer far greater variability at a given range than do unobstructed ones. Such effects must be accounted for explicitly in the models.

11.4.4 Line-of-sight models

11.4.4.1 Two-ray model

In a line-of-sight situation, at least one direct ray and one reflected ray will usually exist (Figure 11.20). Analysis of this situation follows a similar approach to the derivation of the plane earth loss in Chapter 6, except it is no longer appropriate to assume that the direct and reflected path lengths are necessarily similar, or to assume that the reflection coefficient necessarily has a magnitude of unity. The loss is then:

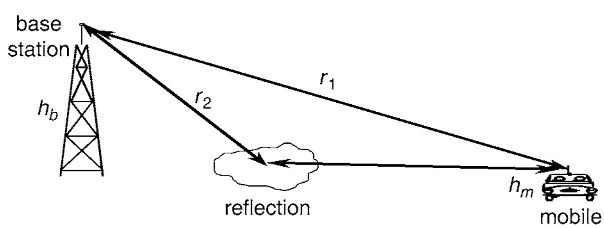


Figure 11.20 Two-ray model of line-of-sight propagation

$$\frac{1}{L} = \left(\frac{\lambda}{4\pi}\right)^2 \left| \frac{e^{-jkr_1}}{r_1} + R\frac{e^{-jkr_2}}{r_2} \right| \tag{11.45}$$

where R is the Fresnel reflection coefficient for the relevant polarisation. In the horizontally polarised case the reflection coefficient is very close to −1, so the path loss exponent tends toward 4 at long distances as in the plane earth loss. For the vertically polarised case the path loss exponent is essentially 2 at all distances, but the large fluctuations present at short ranges disappear at longer distances. Hence both cases produce two regimes of propagation.

Since the reflection coefficient for vertical polarisation is approximately +1 for large distances, the distance at which the rays are in antiphase is closely approximated by the distance at which $r_2 = (r_1 + \lambda/2)$, and this gives the position of the last dip in the vertically polarised signal. This is exactly the definition of the first Fresnel zone (Chapter 8). For high frequencies the distance at which the first Fresnel zone first touches the ground is given approximately by:

$$r_b = \frac{4h_b h_m}{\lambda} \tag{11.46}$$

It has been suggested that this forms a physical method for calculating the breakpoint distance for use in empirical models such as Eqns. 11.42 and 11.44 [33].

The two-ray model forms a useful idealisation for microcells operated in fairly open, uncluttered situations, such as on long straight motorways where a line-of-sight path is always present and little scattering from other clutter occurs.

11.4.4.2 Street canyon models

Although a line-of-sight path frequently exists within microcells, such cells are most usually situated within built-up areas. The buildings around the mobile can all interact with the transmitted signal to modify the simple two-ray regime described in Section 11.4.4.1. A representative case is illustrated in Figure 11.21. It assumes that the mobile and base station are both located in a long straight street, lined on both sides by buildings with plane walls. Models which use this canonical geometry are street canyon models.

Six possible ray paths are also illustrated. Many more are possible, but they tend to include reflections from more than two surfaces. These reflections are typically attenuated to a much greater extent, so the main signal contributions are accounted for by those illustrated.

In comparison with a two-ray model, street canyon models produce more rapid multipath fading, and the differences between vertical and horizontal polarisation are less. Eventually the vertically polarised component diminishes with an average path loss exponent of 4, and the horizontally polarised case tends to 2. However, real streets are rarely straight for long enough to observe this distance range.

In general, for line-of-sight microcells, the base station height has only a weak effect on the cell range. There is some effect due to obstruction from clutter (in

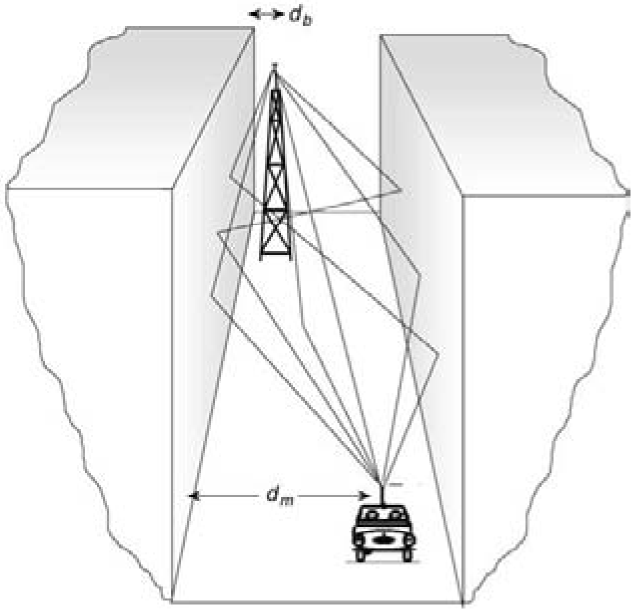


Figure 11.21 Street canyon model of line-of-sight microcellular propagation

this case vehicles, street furniture and pedestrians), but we will see in later sections that increasing the base station height does have a significant effect on interference distance. Thus the antenna should be maintained as low as possible, consistent with providing a line of sight to locations to be covered.

11.4.5 Nonline-of-sight models

11.4.5.1 Propagation mechanisms

When the line-of-sight path in a microcell is blocked, signal energy can propagate from the base to the mobile *via* several alternative mechanisms:

- diffraction over building rooftops
- diffraction around vertical building edges
- reflection and scattering from walls and the ground.

These mechanisms are described further in Dersch [34]. At relatively small distances from the base station and low base antenna heights, there is a large angle through which the signal must diffract over rooftops in order to propagate and the diffraction loss is correspondingly big. Then propagation is dominated by the other two mechanisms in the list, where the balance between the diffraction and reflection depends on the specific geometry of the buildings, in particular whether a strong wall reflection is supported.

At larger distances, particularly those involved in interference between

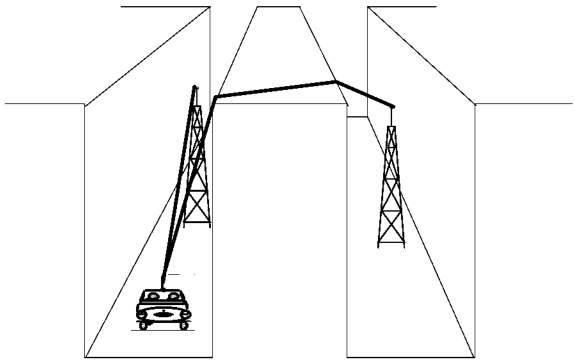


Figure 11.22 Rooftop diffraction as an interference mechanism

co-channel microcells, the rooftop-diffracted signal (Figure 11.22) again begins to dominate due to the large number of diffractions and reflections required for propagation over long distances. Figure 11.23 shows the plan view of buildings arranged in a regular Manhattan grid structure. The short paths *A* and *B* involve only a single reflection or diffraction and are likely to be dominant sources of signal energy. By contrast, the long path *C* is likely to be very weak as four individual reflection losses are involved, and the rooftop-diffracted path *D* is then likely to dominate. This variation in propagation mechanism with distance is another source of the two slopes in the empirical models of Section 11.4.2.

System range is greatest along the street containing the base site. When the mobile turns a corner into a side street, the signal drops rapidly, often by 20–30 dB. The resultant coverage area is therefore broadly diamond shaped, as illustrated in Figure 11.24, although the precise shape will depend very much

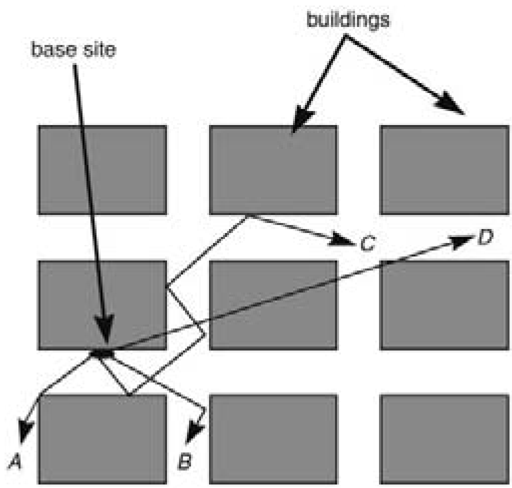


Figure 11.23 Variation of propagation mechanisms with distance for NLOS microcells

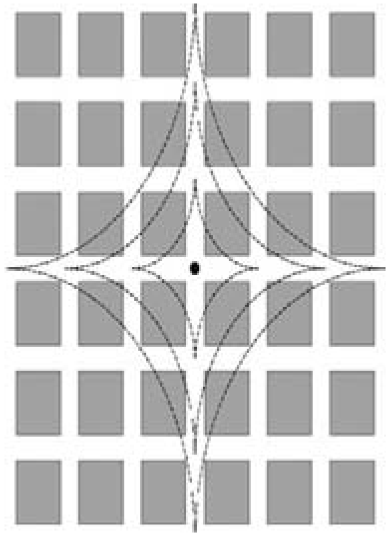


Figure 11.24 Microcellular propagation along and across streets: base site (•), (---) path loss contours

upon the building geometry. Confirmed by measurement, the curve segments forming the diamonds in Figure 11.24 have been shown to indicate that the dominant mechanism of propagation into side streets is diffraction rather than reflection [35].

The variation of the microcell shape with base antenna height in a Manhattan grid structure has been investigated in detail using the multiple diffraction integral [36], and it is shown that there is a smooth transition from a diamond shape to nearly circular as the antenna height increases. It has also been shown [37] that the characteristic diamond cell shape is obtained by considering only the vertical corner diffraction plus reflections from building walls. This work also showed that the distance at which the transition between the various mechanisms occurs depends strongly on the distance between the base station and the nearest street corners.

For low antenna heights, the strong scattering present in microcells prevents the efficient use of sectorisation since the free space antenna radiation pattern is destroyed. Efficient frequency reuse can still be provided, however, by taking advantage of the building geometry. In regular street grid structures, co-channel microcells should be separated diagonally across the street directions, and with sufficient spacing to ensure that cells do not overlap within their breakpoint distance, in order to maintain high signal-to-interference levels.

In more typical environments, where the buildings are not regular in size, more advanced planning techniques must be applied, particularly when frequencies are shared between microcell and macrocell layers; see for example Dehghan [38] and Wang [39].

11.4.5.2 Site-specific ray models

Prediction of the detailed characteristics of microcells requires a site-specific prediction based on detailed knowledge of the built geometry. Electromagnetic analysis of these situations is commonly based on the geometrical theory of diffraction (GTD) and its extensions (Section 8.4). These models are capable of very high accuracy, but their practical application has been limited by the cost of obtaining sufficiently detailed data and the required computation time. More recently, progress in satellite remote sensing has reduced the cost of the necessary data and advanced ray-tracing techniques and cheap computational resources have advanced the art to the point where ray-tracing models are entirely feasible. Nevertheless, many operators consider the costs to be prohibitive and prefer to deploy their microcells based on the knowledge of experienced planning engineers together with site-specific measurements.

11.4.6 Discussion

For the practical application of microcell propagation models, there is an important trade-off between the accuracy of the prediction and the speed with which the prediction can be made. Microcells often have to be deployed very quickly, with little engineering effort, and by people who are not necessarily radio experts. Rules of thumb and very rapid statistical planning tools are very important. Also, even with a very high resolution topographic database, propagation may often be dominated by items of street furniture (signs, lamp posts etc.) and by details of the antenna siting and its interaction with the objects on which it is mounted, which no database could hope to have available. These features may also change rapidly with time, as is certainly the case when dealing with the effects of traffic. For example, when a double-decker bus passes close by a microcell antenna, the coverage area of the microcell may change dramatically for a short time. Either these items must be entered by hand or, more likely, systems of the future will have to be capable of adapting their characteristics to suit the environment which they find, by taking measurements from the active network and responding accordingly.

These factors will dramatically change the way in which propagation models are applied, from being processes that are run at the start of a system deployment, and then used to create a fixed set of predictions and recommendations for deployment, to real-time processes which operate within the base station, with assistance from the mobiles, which are optimised on an ongoing basis and are used by the system to assess the likely impact of changes to system parameters such as transmit powers, antenna patterns and channel assignments.

11.4.7 Microcell shadowing

The log-normal distribution is applied to shadow fading in microcells, just as for macrocells (Section 11.3). Some measurements have suggested that the location variability increases with range, typically in the range 6–10 dB [40]. In order to account for microcell shadowing cross correlation, the shadowing can be separated into two parts, one of which is caused by obstructions very local to the mobile and is therefore common to all paths, and another which is specific to the transport of energy from the mobile to a particular base station [41].

11.4.8 Conclusions

Propagation in microcells can be modelled using either empirical or physical models, just like the macrocells in section 11.2. In either case, however, the clutter surrounding the base station has a significant impact on the cell shape and this must be accounted for to avoid serious prediction errors. In particular, a simple path loss exponent model is inadequate and dual-slope behaviour must be accounted for. This clutter also creates difficulties in deploying antennas, since the clutter disrupts the free-space antenna radiation pattern. Nevertheless, the enormous potential offered by microcells in creating high-capacity cellular systems makes them increasingly attractive methods of providing outdoor coverage in areas with high user densities [38].

11.5 References

1 SAUNDERS, S.R.: 'Antennas and propagation for wireless communication systems' (John Wiley, Chichester, 1999)
2 SHANNON, C.: 'A mathematical theory of communication', *Bell Syst. Tech. J.*, 1948, **27**, 379–423 and 623–56. Reprinted in SLOANE, N.J.A. and WYNER, A.D.: 'Claude Elwood Shannon: collected papers' (IEEE Press, New York, 1993)
3 CLARKE, R.H.: 'A statistical theory of mobile radio reception', *Bell Syst. Tech. J., 1968,* **47**, (6), pp.957–1000
4 PARSONS, J. D. 'The mobile radio propagation channel' (John Wiley, Chichester, 1992)
5 LEE, W.C.Y.: 'Estimate of local average power of a mobile radio signal', *IEEE Trans. Veh. Tech.*, 1985, **34** (1), pp.22–27
6 EGLI, J.J, 'Radio propagation above 40 MC over irregular terrain', *Proc. IRE*, 1957, pp.1383–91
7 DELISLE, G.Y, LEFEVRE, J., LECOURS, M., and CHOUINARD, J.: 'Propagation loss prediction: a comparative study with application to the mobile radio channel', *IEEE Trans. Veh. Tech.*, 1985, **26**, (4), pp.295–308
8 OKUMURA, Y., OHMORI, E., KAWANO, T. and FUKUDA, K.: 'Field strength and its variability in VHF and UHF land mobile radio service', *Rev. Electr. Commun. Lab.*, 1968, **16**, pp.825–73
9 HATA, M.: 'Empirical formula for propagation loss in land mobile radio services', *IEEE Trans. Veh. Tech.*, 1980, **29**, pp.317–25
10 International Telecommunication Union: 'Recommendation ITU-R P.529-2: Prediction methods for the terrestrial land mobile service in the VHF and UHF bands', Geneva, 1997
11 Final report of the COST-207 management committee: 'Digital land mobile

radio communications' Commission of the European Communities, L-2920, Luxembourg, 1989
12 KOZONO, S. and WATANABE, K.: 'Influence of environmental buildings on UHF land mobile radio propagation', *IEEE Trans. Commun.*, 1977, **25**, (10), pp.1133–43
13 COST 231 final report: 'Digital mobile radio: COST 231 view on the evolution towards 3rd generation systems', Commission of the European Communities and COST Telecommunications, Brussels, 1999
14 IKEGAMI, F., TAKEUCHI, T., and YOSHIDA, S.: 'Theoretical prediction of mean field strength for urban mobile radio', *IEEE Trans. Antennas Propag.*, **39**, (3), pp.299–302
15 SAUNDERS, S.R., and BONAR, F.R.: 'Prediction of mobile radio wave propagation over buildings of irregular heights and spacings', *IEEE Trans. Antennas Propag.*, 1994, **42**, (2), pp.137–44
16 SAUNDERS, S.R., and BONAR, F.R.: 'Explicit multiple building diffraction attenuation function for mobile radio wave propagation', *Electron. Lett.*, 1991, **27** (14), pp.1276–77
17 WALFISCH, J., and BERTONI, H.L.: 'A theoretical model of UHF propagation in urban environments, *IEEE Trans. Antennas. Propag.*, 1988, **36**, (12), pp. 1788–96
18 CHRYSANTHOU, C., and BERTONI, H.L.: 'Variability of sector averaged signals for UHF propagation in cities', *IEEE Trans. Veh. Tech.*, 1990, **39** (4), pp.352–58
19 SAUNDERS, S.R., and BONAR, F.R.: 'Mobile radio propagation in built-up areas: a numerical model of slow fading', Proc. 41st IEEE Vehicular Technology Society Conference, 1991, pp.295–300
20 JAKES, W.C.: 'Microwave mobile communications', (IEEE, New York, 1994)
21 IBRAHIM, M.F., and PARSONS, J.D.: 'Signal strength prediction in built-up areas', *IEE Proc. F*, 1983, **130**, (5), pp.377–84
22 REUDINK, D.O.: 'Comparison of radio transmission at X-band frequencies in suburban and urban areas', *IEEE Trans. Antennas Propag.*, 1972, **20**, p.470
23 OTT, G.D.: 'Data processing summary and path loss statistics for Philadelphia HCNTS measurements program'. Quoted in JAKES [20]
24 BLACK, D.M., and REUDINK, D.O.: 'Some characteristics of mobile radio propagation at 836 MHz in the Philadelphia area', *IEEE Trans. Veh. Tech.*, 1972, **21** (2), pp.45–51
25 GREENSTEIN, L.J., ARMITAY, N., CHU, T.S. *et al.*: 'Microcells in personal communication systems', *IEEE Commun. Mag.*, 1992, **30**, (12), pp.76–88
26 SARNECKI, J., VINODRAI, C., JAVED, A., O'KELLY, P. and DICK, K.: 'Microcell design principles', *IEEE Commun. Mag*, 1993, **31**, (4), pp.76–82
27 MADFORS, M., WALLSTEDT, K., *et al.*: 'High capacity with limited spectrum in cellular systems', *IEEE Commun. Mag.*, 1997, **35**, (8), pp.38–46
28 GREEN, E.: 'Radio link design for microcellular systems', *BT Tech. J*, 1990, **8**, (1), pp.85–96
29 HARLEY, P.: 'Short distance attenuation measurements at 900 MHz and

1.8 GHz using low antenna heights for microcells', *IEEE J. Sel. Areas Commun.*, 1989, **7**, (1), pp.5–11

30 CHIA, S.T.S.: 'Radiowave propagation and handover criteria for microcells', *BT Tech. J.*, 1990, **8**, (4), pp.50–61

31 XIA, H.H., BERTONI, H.L., MACIEL, L.R., LINDSAY-STEWART, A. and ROWE, R.: 'Microcellular propagation characteristics for personal communications in urban and suburban environments', *IEEE Trans. Veh. Tech.*, 1994, **43**, (3), pp.743–752

32 BULTITUDE, R.J.C., and HUGHES. D.A.: 'Propagation loss at 1.8 GHz on microcellular mobile radio channels'. Proc. IEEE international symposium on *Personal, indoor and mobile radio communications*, (*PIMRC '96*), October 1996, pp.786–90

33 XIA, H.H., BERTONI, H.L., MACIEL, L.R., LINDSAY-STEWART, A., and ROWE, R.: 'Radio propagation characteristics for line-of-sight microcellular and personal communications', *IEEE Trans. Antennas Propag.*, 1993, **41**, (10), pp.1439–1447

34 DERSCH, U., and ZOLLINGER, E.: 'Propagation mechanisms in microcell and indoor environments' *IEEE Trans. Veh. Tech*, 1994, **43**, (4), pp.1058–66

35 GOLDSMITH, A.J., and GOLDSMITH, L.J.: 'A measurement-based model for predicting coverage areas of urban microcells', *IEEE J. Sel. Areas Commun.*, 1993, **11**, (7), pp.1013–23

36 MACIEL, L.R., and BERTONI, H.L.: 'Cell shape for microcellular systems in residential and commercial environments', *IEEE Trans. Veh. Tech.*, 1994, **43**, (2), pp.270–278

37 ERCEG, V., RUSTAKO, A.J., and ROMAN, P.S.: 'Diffraction around corners and its effects on the microcell coverage area in urban and suburban environments at 900 MHz, 2 GHz, and 6 GHz', *IEEE Trans. Veh. Tech.*, 1994, **43**, (3), pp.762–66

38 DEHGHAN, S., and STEELE, R.: 'Small cell city', *IEEE Commun. Mag.*, 1997, **35**, (8), pp.52–59

39 WANG, L.C., STUBEA, G.L., and LEA, C.T.: 'Architecture design, frequency planning, and performance analysis for a microcell/macrocell overlaying system', *IEEE Trans. Veh. Tech.*, 1997, **46**, (4), pp.836–48

40 FEUERSTEIN, M.J., BLACKARD, K.L., RAPPAPORT, T.S., and SEIDEL, S.Y.: 'Path loss, delay spread and outage models as functions of antenna height for microcellular system design', *IEEE Trans. Veh. Tech.*, 1994, **43**, (3), pp.487–98

41 ARNOLD, H.W., COX, D.C., and MURRAY, R.R.: 'Macroscopic diversity performance measured in the 800 MHz portable radio communications environment', *IEEE Trans. Antennas Propag.*, 1988, **36**, (2), pp.277–80

Chapter 12

Propagation in rain and clouds

J.W.F. Goddard

12.1 Introduction

Water appears in the atmosphere in a variety of forms, usually referred to by the term hydrometeor, which includes particles as diverse as cloud, rain drops, snowflakes, ice crystals, hail and graupel. Of these, rain, hail, graupel and snow are generally recognised as precipitation. The effects that hydrometeors have on communications systems are dependent both on the system frequency and the type of particle present. At any given instant, of course, more than one type of particle will affect a given link. For example, an earth–space link may often encounter rain over the lower part of the path, and snow at greater heights.

We will consider first the various forms of hydrometeor, together with their relevance to radiowave propagation. A brief theoretical framework will then be presented as a background to the model development process and, finally, specific effects of hydrometeors on systems will be considered.

12.2 Rain

12.2.1 Rain drop size distributions

In modelling the effects of rain on radiowaves propagating through it, or scattered from it, the usual problem is to numerically relate two physically distinct quantities. For example, we might wish to relate a microwave-specific attenuation to a rain rate or to a radar reflectivity, a rain rate to a reflectivity, or a differential attenuation between two polarisations to an absolute attenuation in one etc. All these quantities are obtained as integrals over the drop size distribution function $N(D)$, defined such that $N(D)dD$ is the number of drops per cubic metre with drop diameters D between D and $D + dD$. (As all but the smallest drops are nonspherical, D is usually defined as being that of an equivolumic sphere.)

Nearly always the contributions to the two quantities to be related have a different diameter dependence, and some assumptions about the shape of the function $N(D)$ must be made.

Satisfactory results are generally obtained by using an ideal mathematical form for $N(D)$, described by a small number of free parameters. Large departures from the ideal forms undoubtedly occur in real rain, but the situation is eased by the fact that relationships between quantities are in practice often only required to hold on a mean statistical basis. Also, fine structure in $N(D)$ tends to disappear in the integration.

Much modelling to date has used an exponential form for $N(D)$:

$$N(D) = N_0 \exp(-\Lambda D) \tag{12.1}$$

Early work [1] indicated that Λ tends to increase with rain rate R, and that these distributions can be reduced to a one-parameter family by treating the relation as deterministic, with:

$$\Lambda = 4.1\ R^{-0.21} \tag{12.2}$$

where Λ is in mm^{-1} and R is in mm h^{-1}.

Once Λ is determined from R, so is N_0 by the requirement that the integral of $N(D)D^3\ V(D)$, where V is the drop terminal velocity, actually gives R. Typical values of N_0 were found to be around 8000 $\mathrm{mm}^{-1}\ \mathrm{m}^{-3}$, with this form of distribution becoming known as the Marshall–Palmer distribution. The parameter, Λ, can be shown to be related to the median volume drop diameter of the distribution, D_0, through the relation:

$$\Lambda = 3.67 / D_0 \tag{12.3}$$

The Laws–Parsons distribution [2], which predates the Marshall–Palmer approach and is also much used, is an empirically measured form for $N(D)$ which is tabulated numerically rather than expressed mathematically. It is similar to the Marshall–Palmer form except that it has slightly fewer small drops.

Subsequent work which followed the development of disdrometers, instruments capable of automatically recording the size distribution of drops, showed that N_0 could, in fact, vary quite considerably, both from one event to another and even within a rain event. In addition, the shape of the distribution was also found to vary significantly, which led to a more general form of distribution, the gamma-type distribution, being proposed [3]:

$$N(D) = N_0 D^m e^{-(3.67+m)^{D}/_{D_0}} \tag{12.4}$$

Positive values of m reduce the numbers of drops at both the large and small ends of the size spectrum, compared with the exponential distribution obtained when $m = 0$. In contrast, negative values of m increase the numbers of drops at each end of the spectrum. Figure 12.1 shows this effect for three different distributions, all with the same rainfall rate (5 mm h^{-1}) and the same D_0 (1 mm).

It is interesting to note from Figure 12.1 that the three distributions have very different numbers of small and large drops, but the same integrated rainfall rate.

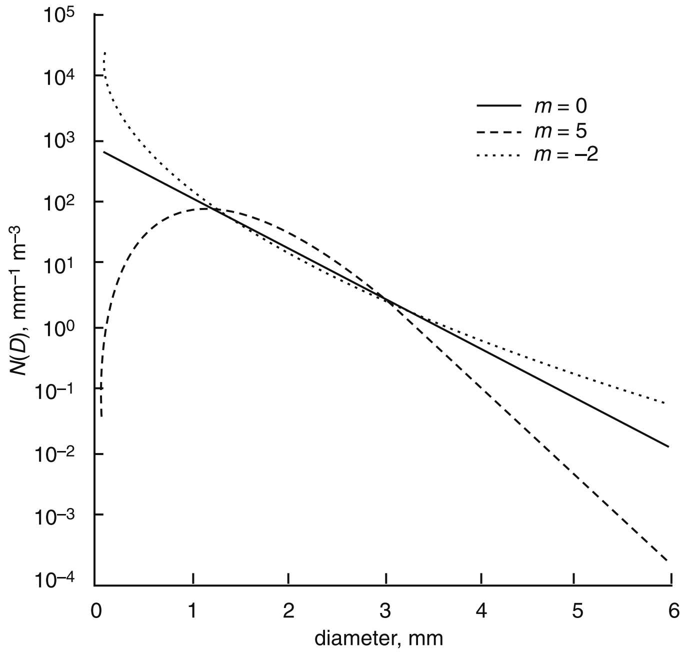


Figure 12.1 *Examples of three different drop-size distributions, with a rainfall rate of 5 mm h^{-1}, and a D_0 value of 1 mm*

The reason for this can be seen in Figure 12.2, which indicates the contribution from each drop size to the total rainfall rate. For example, drops between 1 and 2 mm diameter contribute 60 per cent of the total rainfall rate of 5 mm h^{-1}.

It must be stressed that there is at present no single model for drop-size distributions which is generally accepted as representing physical reality, even as a statistical mean over many rain events. Observations of quantities which are very sensitive to large drops, such as radar differential reflectivity, and certain cross-polarisation measurements, have tended to indicate the presence of fewer very large drops than in an exponential distribution. They have been well fitted using a gamma distribution with m between 3 and 5. This model incidentally lowers the assumed density of small drops (below 1 mm diameter) even though the experimental data may have little to say about them. Other measurements, particularly of attenuation at 30 to 300 GHz, are more sensitive to small drops and make no strong implication about the presumed cut-off point for large drops, although work in the UK has suggested that a log-normal form of distribution would be appropriate for millimetric attenuation prediction [4]. Deductions about small drops have sometimes been contradictory [3,4].

It has generally been found that the Marshall–Palmer distribution is satisfactory for statistical predictions of attenuation in the 10 to 30 GHz range, and probably remains a good estimate for statistical work at slightly higher

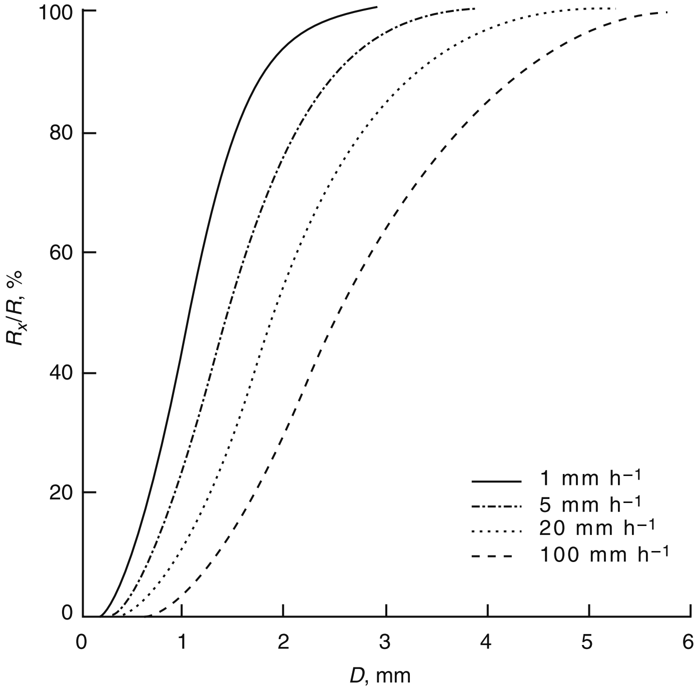


Figure 12.2 Contributions of different drop sizes to integrated rain fall rate, for various exponential drop-size distributions with $N_0 = 8000\ mm^{-1}m^{-3}$

frequencies. Fortunately, for particular modelling purposes it is not essential that the assumed distribution represents physical truth at all drop sizes.

12.2.2 Rain drop shapes

Rain drops in free fall are not spheroids (except for the smallest drops). Instead, their true shape is an oblate spheroid with a flattened base. As the drop size increases above 4 mm, the base actually becomes concave. In scattering calculations, it is usual to model the shapes as simple oblate spheroids. A wide range of size–shape relations have been proposed, largely based on wind-tunnel measurements, and Figure 12.3 shows two of these.

12.3 Other forms of hydrometeor

12.3.1 Liquid water clouds and fog

Nonprecipitating clouds containing only liquid water are not very significant for frequencies below about 40 GHz. The liquid water content is too low to cause much absorption of energy; the droplets are too small to cause much scattering of energy, and being virtually spherical do not cause measurable cross

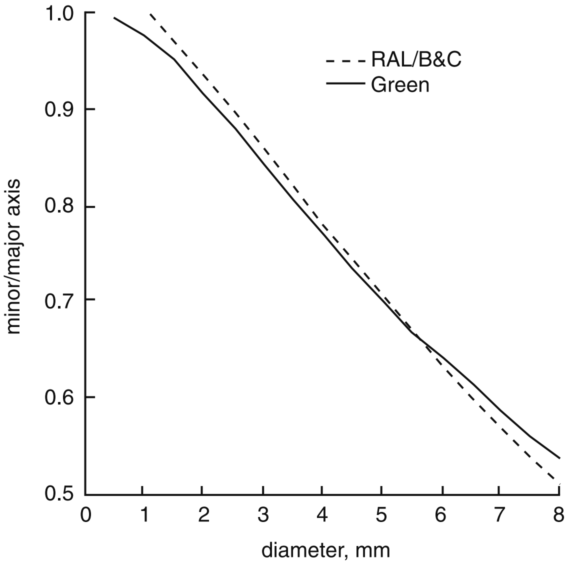


Figure 12.3 Rain drop shape as a function of diameter, for two representative models

polarisation. Liquid water contents vary from about 0.1 g m^{-3} in stratiform clouds, through to 0.5 g m^{-3} in cumulus clouds, up to around 2 g m^{-3} in cumulonimbus clouds. The liquid water content typically peaks about 2 km above the cloud base then decreases towards the top of the cloud. For much greater detail, see Mason [7]. Section 12.6.1 describes the frequency-dependent specific attenuation to be expected in liquid water clouds.

Fog can be considered to have similar physical properties to cloud, except that it occurs close to the ground.

12.3.2 Ice hydrometeors

There is an enormous variety of forms of atmospheric ice particles, and the physics of their formation is extremely complex. However, some simple classifications may be made for radio purposes. Reference [4] gives a much more detailed discussion.

12.3.3 Ice crystals

From the radio propagation point of view, single crystals at high altitudes appear to be the most important form of atmospheric ice, as they give rise to cross-polarisation effects (Section 12.7). A very crude picture which nevertheless appears adequate to interpret radio effects (particularly cross polarisation) is to divide them broadly into plate and prism (or needle) forms. The first group includes simple hexagonal plates with typical diameters up to 0.5 mm and thicknesses of perhaps 1/10 of the diameter, and the classical dendritic form with complex hexagonal symmetry. The latter may reach 5 mm in diameter, with a thickness of probably less than 0.05 mm. The second group includes a variety of long thin shapes including both needles and hexagonal prisms; typical lengths are around 0.5 mm, and the ratio of length to breadth ranges from one to five. For modelling radio scatter when the crystals are small compared with the wavelength, a start can be made by treating them as either very oblate or very prolate spheroids.

The type of crystal depends in a complex way on the temperature at which it formed, and hence on the height. In the range –8° to –27 °C, plate types are typical, with dendrites dominant in the subrange –14° to –17 °C. Below –27 °C prisms are formed. Both classes can form in the 0° to –8 °C range, but they are probably less important for propagation, except where differential phase effects are of concern.

12.3.4 Snow

Snow consists of aggregated ice crystals, with large flakes forming only at temperatures just below freezing when the crystal surfaces become particularly sticky. In stratiform rain most of the ice is present as large flakes down to a few hundred metres above the melting level, and at greater heights single crystals are mainly present. Dry snow does not appear to be very important for propagation of radiowaves, at least below 30 GHz. This is partly because of its low density structure (generally around 0.1 g m^{-3}), giving it a permittivity close to unity, and (for polarisation effects) partly because the flakes tend to tumble without a preferred orientation.

12.3.5 Hail and graupel

Particles of hail and graupel are mainly formed by the accretion of supercooled cloud droplets in convective storms. Hail particles have densities close to that of water, and are often of roughly spherical shape, although a wide variety of shapes have been recorded. Graupel has a density intermediate between those of hail and snow and is usually of conical shape. Although of a higher apparent permittivity than snow, dry hail and graupel are only weakly attenuating below 30 GHz and do not appear to be very important for wave propagation effects. However, when they begin to melt, they scatter like very large raindrops. Noticeable polarisation effects might be expected but these seem to be of little significance on a statistical basis, at least in a climate such as that in the UK.

12.3.6 Melting layer

In stratiform rain, partially melted snowflakes exist within a height interval of about 500 m around the 0 °C isotherm, and this melting layer can produce intense radio scattering effects, due to a combination of factors. The melting particles combine a large size with a large apparent permittivity, intrinsically scattering and attenuating strongly compared with the dry ice above and smaller raindrops below. The effect is magnified by their high spatial density, relative to the rain, because of their relatively small fall velocity. Low density snow flakes fall at about 1 m s^{-1}, and raindrop fall speeds are significantly higher (see Figure 12.4). Fall speeds increase gradually as the melting process becomes complete, with a consequent decrease in particle number densities.

The consequences of these effects are shown in Figure 12.5, where the 3 GHz backscattered radar reflectivity is shown as a function of height during a typical stratiform rain event.

Also shown in the Figure is the difference in scattered power between horizontal and vertical polarisation, where it can be seen that the large melting snowflakes scatter the horizontal polarisation much more strongly than the vertical, because of their much larger horizontal dimensions. The melting phase has been modelled in some depth, see for example Reference [8], where it is predicted that the strongly enhanced scatter apparent at frequencies below about 15 GHz rapidly disappears above that frequency due to non-Rayleigh scatter from the large particles present.

Although effects of the melting layer on most earth–space paths are small because of the limited length of path within the layer, in climates such as the UK's it is likely that terrestrial link paths frequently traverse large lengths of melting layer and could be subject to more attenuation than expected from the rain rate seen at the ground (see Section 12.6.2).

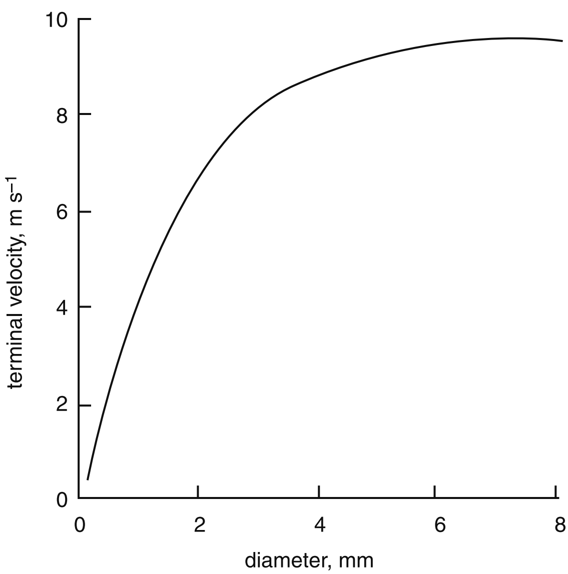


Figure 12.4 Terminal velocity of rain drops as a function of diameter

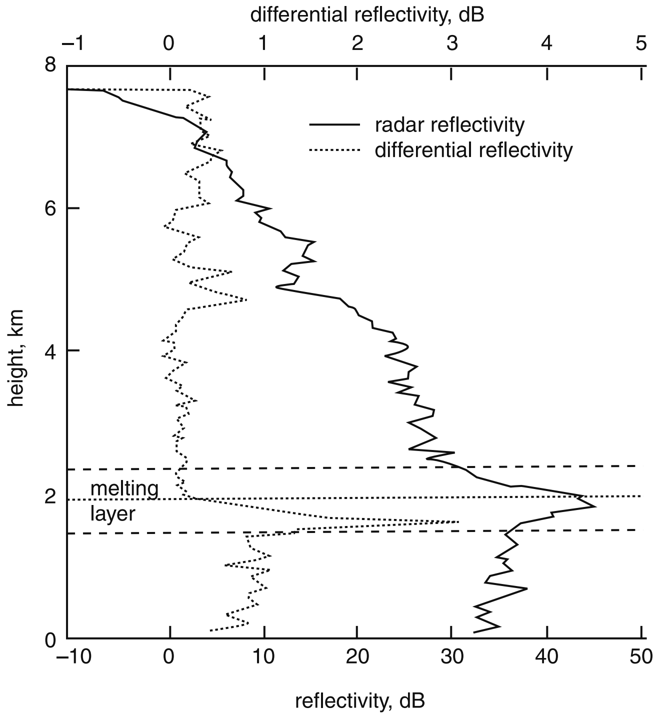


Figure 12.5 Radar reflectivity and differential reflectivity (as functions of height) obtained with the Chilbolton radar

Above the melting layer, the lower permittivity of snow reduces the magnitude of differential scatter to close to zero, except where, as near 5 km height, there are significant quantities of highly asymmetrical ice plates. Below the melting layer, oblate raindrops produce measurable differential scatter, in this case of the order of 1 dB. In this example, the changes in reflectivity and differential reflectivity with respect to height within the rain are due to wind shear effects, rather than, for example, evaporation. These shear effects can be seen in the full radar scan of the same event, shown in Figure 12.6. This Figure also gives an indication of the variability of rain on a horizontal scale, even in stratiform conditions, with moderately heavy precipitation of around 10 mm h^{-1} at 20 km range, but much weaker rain of around 1 mm h^{-1} at longer ranges. The difference in mean drop sizes between these two regions can also be inferred from the differential reflectivity values.

12.4 Refractive indices of ice and water

For understanding the interactions of radiowaves with hydrometeors, a vital factor is obviously the complex refractive index n (or equivalently the permittivity

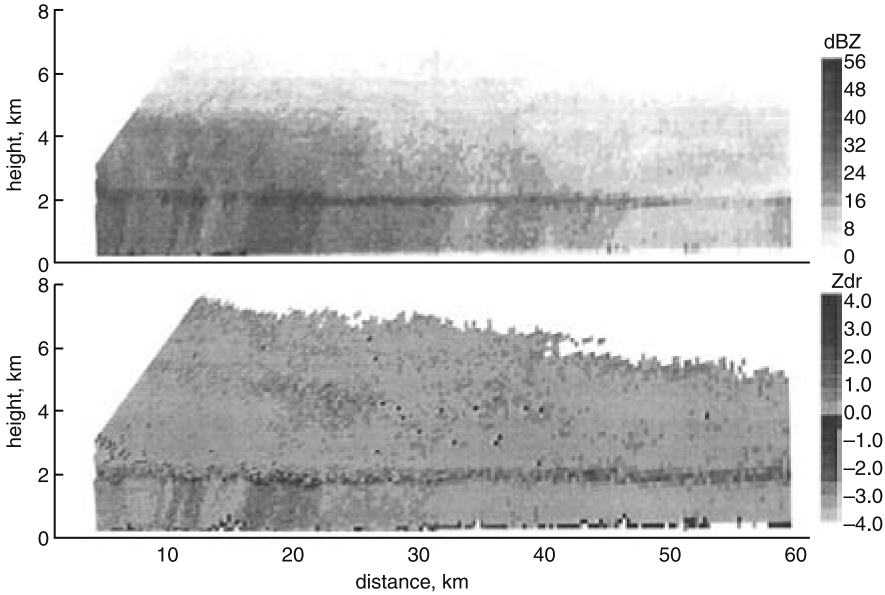


Figure 12.6 Vertical radar section through rain obtained on 14 September 1994 (top – reflectivity; bottom – differential reflectivity)

ε_r, equal to n^2) of the water or ice forming the particle. Figure 12.7 shows the dependence of n on frequency for water. The high values for water arise from the polar nature of the water molecule, and are consistent with the well known value of about 80 for the dielectric constant of water at low frequencies. The peak in Im(n) around 20 GHz can be regarded as the 22 GHz line of water vapour, greatly broadened in the liquid state.

Although not shown on the Figure, Re(n) for ice is very close to 1.78 over the entire frequency range shown, while at 0 °C, Im(n) falls from 0.025 at 300 MHz to 0.0002 at 300 GHz. (Corresponding values for –20 °C: 0.006 to 0.00004.)

A plane wave travelling in a medium of refractive index n has its field intensity reduced by $\exp(-2\pi\ \mathrm{Im}(n))$ in travelling through one free-space wavelength. Thus in water at 300 GHz and 0 °C, waves are attenuated in voltage by about $\exp(-2\pi\ 0.5) = 0.04$ in travelling through 1 mm.

This shows that waves penetrating raindrops at 300 GHz and above are well attenuated within the diameter of the drop.

Further insight can be obtained by considering the quantity $z = (\varepsilon_r - 1)/(\varepsilon_r + 2)$. A plot for water is included in Figure 12.7. It is important to note that, although Im(n) peaks at a certain frequency (dependent on temperature), Im(z) rises almost linearly with frequency throughout most of the frequency range. This is because Re(n) is falling with frequency. Re(z) is close to 1 (>0.94) up to 30 GHz, then falls to 0.79 (0.91) at 100 GHz and 0.63 (0.74) at 300 GHz, for 0 °C (30 °C). For ice Re(z) is always close to 0.42 and Im(z) is about 0.0009 at 0 °C and

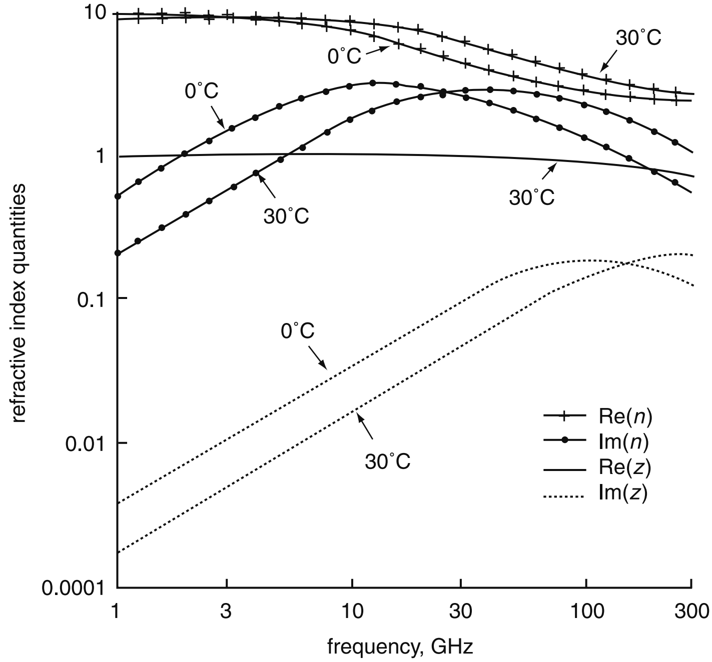


Figure 12.7 Real and imaginary components of refractive index of water(n)

10 GHz, and falls with frequency. Note that it is very much smaller than the value for water.

Except for Re(n) and Re(z) for ice, and Re(z) for water below 100 GHz, all the quantities have a noticeable temperature dependence. Behaviour outside the frequency range shown is complex and further details can be found in Reference [9]. Lower frequencies are almost unaffected by hydrometeors, but higher frequencies must be regarded as being outside the radio frequency range.

12.5 Hydrometeor scattering theory

12.5.1 General

The far field scattered from a particle is commonly described by a dimensionless function S of the scattering angles [10], and is defined by:

$$E_{\text{scat}} = E_{\text{inc.}}\, S\frac{(\theta,\varphi)}{jkr} e^{-jkr+j\omega t} \quad (k = 2\pi/\lambda) \tag{12.5}$$

where r is the radial distance from the particle. For forward scatter, S is written $S(0)$. The total extinction cross section of the particle is given by:

$$C_{\text{ext}} = \lambda^2/\pi \ \text{Re}[S(0)] \tag{12.6}$$

A plane wave propagating through a medium containing N randomly distributed particles per unit volume experiences an attenuation of $NC/2$ nepers per unit distance (1 neper is a voltage ratio of e, a power ratio of e^2, or 8.7 dB):

$$a = 8.7 \frac{N\lambda^2}{2\pi} \mathrm{Re}\,[S(0)] \qquad \text{dB/unit distance} \tag{12.7}$$

The corresponding specific phase shift is:

$$\beta = \frac{N\lambda^2}{2\pi} \mathrm{Im}\,[S(0)] \qquad \text{radians/unit distance} \tag{12.8}$$

For rain with a distribution of drop sizes, the term NS can be rewritten as an integral over D of $N(D)S(0,D)$.

12.5.2 Rayleigh scattering region

A scattering particle of radius a and refractive index n is in the Rayleigh scattering region when it is both electrically small ($2\pi a/\lambda << 1$) and phase shifts across it are small ($2\pi na/\lambda << 1$). In this condition Rayleigh's approximation can be used, which assumes that:

(a) the scattered field is that of a dipole;
(b) the dipole moment induced in the particle is related to the incident electric field in the same way as for electrostatic fields.

At radio frequencies, the approximation is especially useful for cloud droplets and atmospheric ice crystals, and it gives partial insight into rain scatter.

If P is the induced dipole moment per unit incident field, $S(0)$ is given by:

$$S(0) = j\, P\omega^3\, \mu_0/4\pi c \tag{12.9}$$

then P can be written as:

$$P = \varepsilon_0\,(\varepsilon_r - 1)\,\xi v = \varepsilon_0 U v \tag{12.10}$$

where v is the particle's volume and ξ is the ratio of the internal to the external field. This ratio is given by $\xi = 3/(\varepsilon_r + 2)$ for a sphere. For a very oblate spheroid (flat plate), ξ is 1 when the field is applied along the longest axis, and is $1/\varepsilon_r$ when applied along the shortest. For a very prolate spheroid (thin needle) it is 1 when the field is parallel to the long axis, and is $2/(\varepsilon + 1)$ when at right angles to it. Values of U for water and ice spheres, and ice needles and plate are shown in Table 12.1. Imaginary parts are not shown as they are very small for ice.

Table 12.1 Values of U for different hydrometeors

	Water sphere	Ice sphere	Ice needle	Ice plate
U_{parallel}	$3z$	1.26	2.32	2.32
$U_{\text{perpendicular}}$	$3z$	1.26	1.04	0.68

Specific attenuation and phase shift in a cloud of the particles are given by:

$$\alpha = (\pi V/\lambda) \,.\, \mathrm{Im}(U) \quad \text{nepers/unit distance} \tag{12.11}$$

$$\beta = (\pi V/\lambda) \,.\, \mathrm{Re}(U) \quad \text{radians/unit distance} \tag{12.12}$$

Here, $V = Nv$ is the fractional volume of space occupied by the particles. Important features of these two expressions are their direct proportionality to frequency (if n is frequency independent) and to total particle volume, regardless of how distributed between sizes. The first equation includes only extinction due to absorption and the Rayleigh approximation can be improved by adding to C_{ext} a scatter cross section equal to the total power reradiated by the induced dipole, which is $P^2/3\varepsilon_0{}^2\lambda^4$. This term varies as D^6f^4 and becomes noticeable for raindrops at electrical sizes which are approaching breakdown of the Rayleigh approximation.

12.5.3 Optical and resonance scattering regions

The optical scattering region is that in which the incident wavelength is much less than the diameter of the scattering particle. The larger raindrops are approaching this condition at 100 GHz and above. The scattering may be described by a ray model. Provided the rays which enter the particle are well attenuated within the diameter – which is true in the high radio range but not of course at truly optical frequencies – the extinction cross section approaches a value of twice the scatter's geometric area. The factor of 2 is known as the extinction paradox, see Reference [10].

Between the Rayleigh and optical regions lies the resonance region in which no simple approximation for the scattered fields is available. Mie's theory gives exact results for a sphere but requires considerable computation. Several numerical methods have been developed for nonspherical particles, for example Reference [11]. Figure 12.8 shows extinction cross sections for raindrops as functions of drop size and frequency, and several features of the three scattering regions can be observed.

Taking Figures 12.8*a* and *b* together, we can see cross sections increasing roughly as D^3f^2 in the Rayleigh region. Eqn. 12.7 predicts this kind of dependence, with absorption increasing in proportion to particle volume and to f.Im(z), and it was already noted that Im(z) increases roughly as f over most of the frequency range. In Figure 12.8*a*, a transition to D^2 behaviour can be seen for large drops, the transition occurring at a smaller size as the frequency is raised. Both graphs show fine structure due to resonance. In Figure 12.8*a*, it is seen that the resonances become increasingly damped at the higher frequencies; at the lower frequencies, as D is increased towards the first strong resonance, the dependence becomes a good deal steeper than D^3. This is particularly clear at 5 GHz. Similarly, Figure 12.8*b* shows a much steeper dependence than f^2 at intermediate sizes and lower frequencies, and the relative independence of frequency for larger particles at the higher frequencies is also clear.

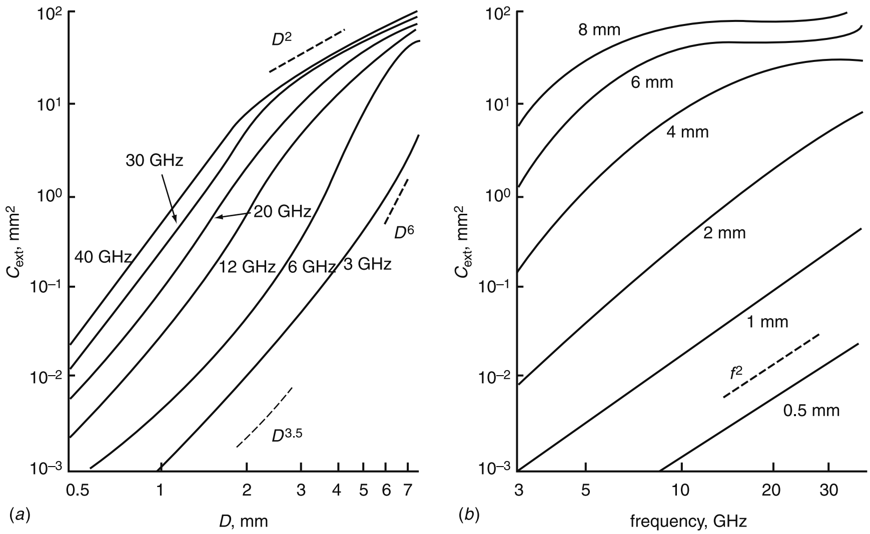


Figure 12.8 *Extinction cross sections*
(a) versus drop diameter
(b) versus frequency

12.6 Attenuation effects

12.6.1 Rain and cloud

The total attenuation from rain is due to the sum of the contributions of each individual drop. It is therefore necessary to assume a form of drop size distribution, $N(D)$, as described in Section 12.2.1. Then the total attenuation, A, can be computed from:

$$A = 4.34\, L \int_0^\infty \cdot C_{\text{ext}} N(D) dD \quad \text{decibels} \tag{12.13}$$

Figure 12.9 shows the results for specific attenuation, using Eqn. 12.12 over a 1 km path together with the extinction cross sections in Figure 12.8 and assuming an exponential distribution with $N_0 = 8000$ mm^{-1}m^{-3}. The behaviour can be well understood from the previous discussion of cross section dependencies on size and frequency, remembering that fine structure due to resonance is smoothed out by the variety of drop sizes.

Cloud liquid water attenuation is simply related to the concentration of liquid water, and can be calculated effectively from Eqn. 12.11. The specific attenuation in a water cloud of density 1 g m^{-3} is shown in Figure 12.10 as a function of frequency.

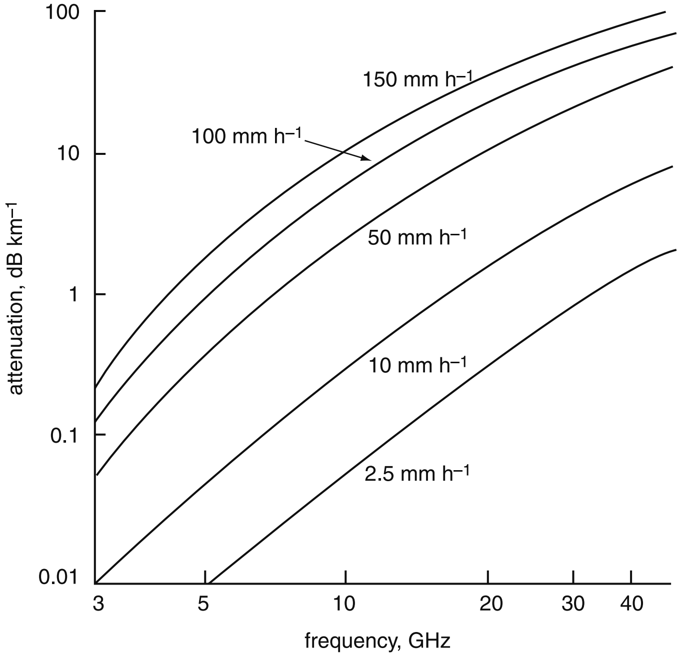


Figure 12.9 Specific attenuation due to rain

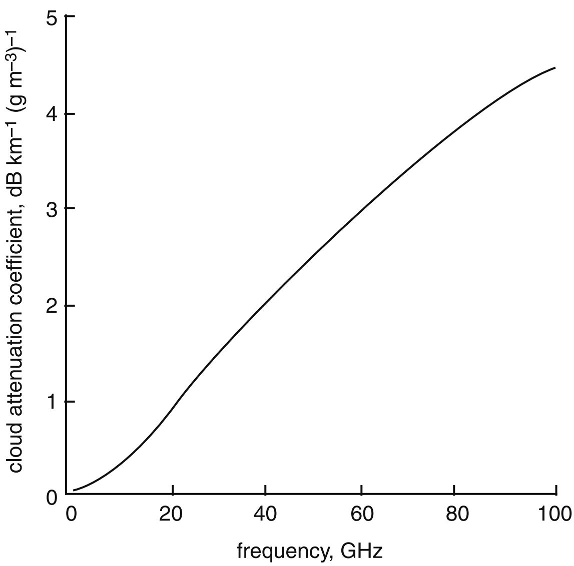


Figure 12.10 Specific attenuation due to cloud, 0.5 g m^{-3} at 0°C

12.6.2 Melting layer

It was noted in Section 12.3.6 that attenuation in the region of melting snowflakes can be significantly larger than in the rain below. Several models have been developed in attempts to quantify the scattering and attenuation effects, but they are hampered by the complexity of the melting process, and a lack of knowledge (or extreme variability) of the initial conditions. Example model calculations [11] are plotted in Figure 12.11; the frequency is 20 GHz and a Marshall–Palmer drop size distribution has been assumed. The extra attenuation in the melting layer is predicted to also depend on the initial density of the melting snowflakes. In Figure 12.11 densities of 0.2 and 0.3 g m^{-3} have been assumed, and it can be seen that the lower density almost doubles the attenuation.

12.7 Polarisation dependence

Because hydrometeors are not spherical, if a wave is propagated through them along a line-of-sight path it will usually change its polarisation as it travels. The generation of such a cross-polarised component may be a problem for communication systems using polarisation orthogonality to maintain isolation between channels.

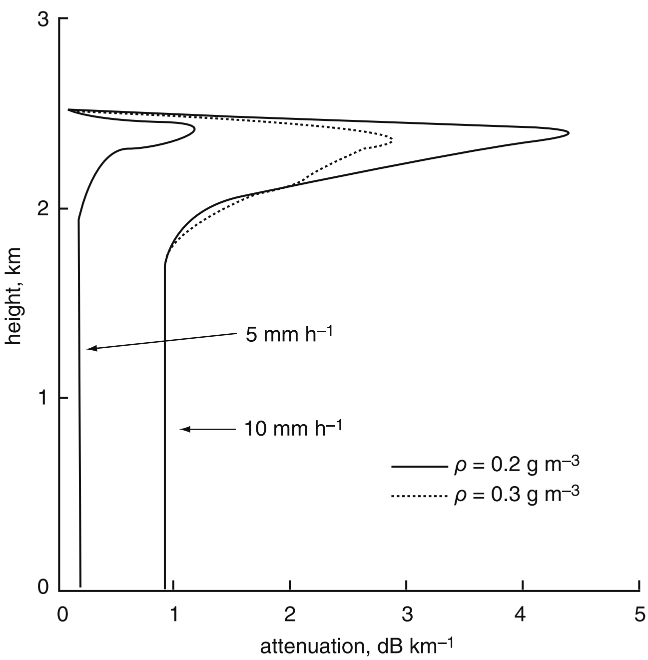


Figure 12.11 Attenuation in rain and melting layer at 20 GHz

12.7.1 Concept of principal planes

For any line-of-sight path through a volume of hydrometeors, there always exist two polarisations which will propagate over that link without changing. These polarisations may be frequency dependent and time varying as the rain, ice etc. change. They are referred to as principal planes of the link but need not theoretically be of linear polarisation. Waves transmitted in the principal planes will arrive unchanged but, because of the nonspherical particles, they will experience different attenuation and phase shifts. Consequently, any transmitted polarisation that is not one of the link's principal planes will be cross polarised on reception.

A great simplification for rain effects is to assume that the raindrops are all rotationally symmetric, and their symmetry axes are all vertical. It is then physically obvious that the principal planes are linear vertical and horizontal polarisations. This description is a very good approximation for most purposes.

12.7.2 Differential attenuation and phase shift in rain

As we discussed briefly in Section 12.2.2, raindrops have a mean shape which is close to an oblate spheroid, and so radiowaves propagating through them will suffer differential attenuation and phase shift. Calculations of the magnitude of these effects are shown in Figure 12.12 as a function of frequency. At low frequencies, this confirms the expectation from Rayleigh theory that attenuation and phase shift will be greater in horizontal polarisation. The reversal in sign of the differential phase at high frequency is a purely resonance phenomenon in which large drops produce negative differential phase outweighing a positive contribution from the smaller ones. It is important to note that, although differential attenuation, and phase below 18 GHz, increase with frequency for a given rain event, they decrease for a given fade depth. This is partly because less deformed smaller drops make a greater relative contribution to the total attenuation as frequency is raised.

12.7.3 Attenuation – XPD relations

If a given microwave link experiences a given fade depth due to rain, this could be because of a variety of drop size distributions. Even approximating the drop size distributions as a one-parameter family, with Λ dependent on rain rate (see Eqns. 12.1 and 12.2), that rain rate would depend on the path length over which the rain was operating to produce the given fade.

Despite this variability, calculations show that the cross polarisation is closely related to total attenuation. The possible variations in drop size distribution for a given fade tend to vary the differential attenuation and differential phase in opposite senses, tending to keep the cross polar amplitude (but not its phase) fairly constant. Thus for system design calculations it is useful and convenient to approximate the relation between cross polarisation and attenuation as a deterministic one.

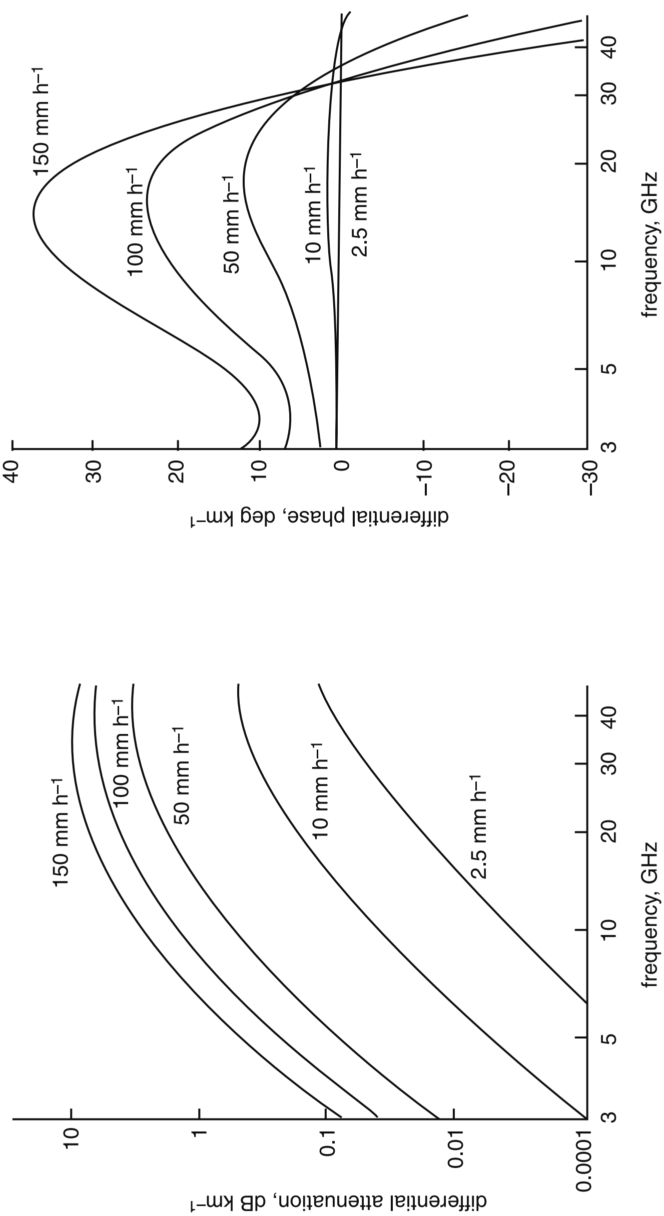


Figure 12.12 ***Differential propagation (horizontal–vertical) in rain***

The cross polar magnitude is usually expressed as cross polar discrimination (XPD), defined as the decibel ratio, at the receiving point of a link, of the field intensity in the wanted or normal polarisation to that induced by the propagation medium in the orthogonal polarisation. XPD is worst (numerically lowest) when the transmitted polarisation is either circular or is a linear polarisation bisecting the principal planes of the medium.

As a first-order approximation, it can be assumed that the cross polar field intensity (voltage) is proportional to the total amount of rain traversed and thus to the total fade. Reverting to the decibel XPD scale this leads to the equation:

$$XPD = U - 20 \log A + I(\theta) \tag{12.14}$$

where A is the fade depth of the copolarised signal in dB, U is a constant for a particular link frequency. $I(\theta)$ is the improvement factor applicable when the transmitted polarisation is not one of the worst case ones just mentioned. It is approximately given by $20 \log |\sin 2\theta|$, if the transmitted polarisation is linear and makes an angle of θ with the principal planes of the rain medium.

Some experimenters have derived a relationship in which the term $20 \log A$ is replaced by $V \log A$, V being a second frequency dependent constant. This can make some allowance for the tendency of deeper fades to be associated with higher rain rates and thus larger raindrops. Between 8 and 35 GHz, ITU-R recommends: $U = -30 \log f$, where f is the link frequency in GHz, with $V = 20$ between 8 and 15 GHz, and $V = 23$ between 15 and 35 GHz [15]. Relations at higher frequencies appear not to be well established. Note that the f dependence of U is consistent with the previous comment that differential propagation reduces with f for a given fade depth. Eqn. 12.14 assumes a horizontally propagating wave. For an elevated path, the term $-40 \log(\cos \text{elevation})$ should be added to XPD because raindrops appear more spherical when viewed at a greater elevation. At 10 GHz, a fade of 15 dB which is quite severe would give a worst case XPD of 6.5 dB at zero elevation.

12.7.4 Rain drop canting angles

It has been shown experimentally that the rain medium exhibits small tilts of its principal planes, so that horizontally and vertically polarised waves do suffer measurable cross polarisation. This effect is called the canting angle of rain drops. Cross polar measurements [16] suggest that a drop falls so that its symmetry axis is parallel to the velocity of air flow relative to the drop. If a drop falls through a vertical wind shear, its horizontal velocity is not equal to that of the local air and a horizontal component of drag force is produced. The air velocity vector relative to the drop is thus tilted from vertical, hence the symmetry axis is also tilted.

Calculations show that terrestrial links within about 40 m of the ground could see canting angles up to about 5°, and experimental values have been of this order. As wind shears reduce with height the model implies equally strongly that tilts seen on earth–satellite paths at elevations of more than a few degrees would be unlikely to see angles greater than 1°, which is negligible for practical purposes. This has not been conclusively demonstrated in experimental data, due to measurement errors.

12.7.5 Ice crystal principal planes

Of the two general types of ice crystal shape thought to affect elevated earth–satellite paths, it is known that plate types generally fall with their flat faces in a horizontal plane. For a medium of these crystals the principal planes are clearly linear vertical/linear horizontal.

Needle crystal types are thought to fall with their symmetry axes (which are also their longest axes) in a horizontal plane. It is also believed that the azimuths of these long axes may be systematically aligned in one direction by wind shear or sometimes by atmospheric electrostatic fields. The evidence for this comes from observations of rapid changes in XPD coinciding with changes in field strength measured by nearby electric field probes [17].

Experimental data suggest that all principal plane angles occur occasionally, but show [18] that they spend about 80 per cent of the time within ±15° of vertical/horizontal when the event is significant. This is consistent with the picture that needles are usually accompanied by plates at a lower altitude.

12.7.6 Differential phase shift due to ice

Statistics of ice differential phase shift on elevated paths are inferable from measurements made at a few sites using satellite beacons. In Europe a typical value exceeded for 0.01 per cent of the time at about 30° elevation would be about 10° at 12 GHz [18] corresponding to a worst case XPD of 22 dB. Such statistics are difficult to relate to any features of the ground rainfall. The phase shift can, however, be scaled linearly to much higher frequencies still in the Rayleigh region. It is also very difficult from any ground-based radio or radar measurement to make any detailed inferences about ice particle shapes and size distributions. Ice may exhibit resonant scatter in the millimetric region, which would be hard to model on the basis of existing data.

12.8 Spatial–temporal structure of rain

12.8.1 Temporal (rainfall rate)

Rainfall rate is obviously the key factor in determining the amount of attenuation likely to be suffered by a link. The most widely used statistics are those of annual rainfall, expressed as rain rate exceeded for a given percentage of the year. Because of the extreme variability of rainfall statistics at a given location (see Figure 12.13), averages of statistics over many years are recommended. As the Figure also shows, even individual gauges can give different results, perhaps due to calibration differences.

If sufficiently accurate data are not available for a particular location, then Recommendation ITU-R P. 837 provides annual statistics for the whole world, as a function of latitude and longitude.

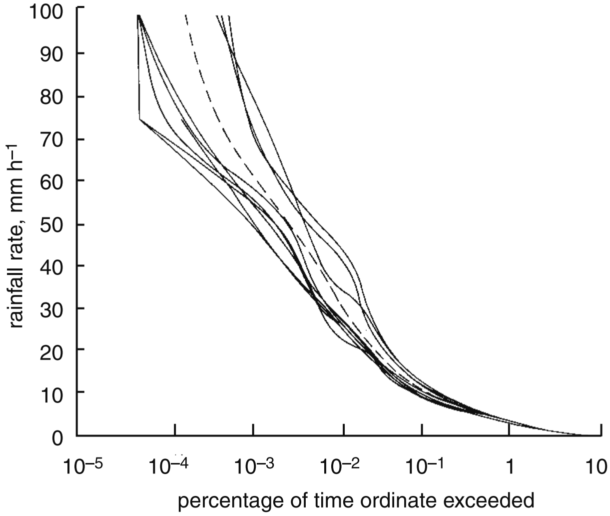


Figure 12.13 Annual statistics from three gauges and three years, in southern UK

12.8.2. Spatial (vertical and horizontal structure)

The height of the 0 °C isotherm plays a key role in the magnitude of attenuation on earth–space paths, because of the different permittivities of ice and water discussed in Section 12.6. For a particular rain event, although the rainfall rate determines the specific attenuation, the height of the 0 °C isotherm determines the length of path over which this is integrated. In temperate climates, like the UK, the 0 °C isotherm height varies significantly with season, with typical summer heights of 3 km, and winter heights of 1 km or even less in, for example, the northern latitudes of the UK. However, in tropical latitudes, the height is not only higher but also less variable. For modelling purposes, the annual mean is usually required, and Recommendation ITU-R P.839 provides appropriate statistics.

One final point that should be made about rain structure is that it is inhomogeneous, both horizontally and vertically. Although vertical variability is only an issue in tropical regions, where the 0 °C isotherm is high, horizontal variability has an impact in all climates, and particularly in heavy precipitation which can be very localised. Models generally cope with this by adding a scaling factor, dependent on path length, called a path reduction factor.

12.9 Bistatic scatter in rain

Bistatic scatter refers to a situation where a cloud of particles scatters a signal *via* an indirect path between two antennas which have intersecting beams but are not strongly coupled by a direct path. This is usually considered as a cause of interference, although it is also sometimes used intentionally in the form of bistatic radars. An example of an interference geometry is shown in Figure 12.14, where a terrestrial link is shown sharing the same frequency band as an earth station.

When no rain is present, the only path for interference from the terrestrial link into the earth station would be through the latter's far sidelobes, and even this route would normally be blocked by local terrain features. However, if rain occurred in the common volume of the two systems, then energy could be scattered from the terrestrial transmitter into the main beam of the earth station, with an increase in the possibility of interference. Figure 12.15 shows data from an experimental link operating between Chilbolton and Baldock in the UK, at a frequency of 11.2 GHz, together with some results from a prediction method developed within COST 210 [19] and subsequently adopted by the ITU-R [20]. In the absence of rain, no signal is received at the Baldock satellite monitoring station from the transmitter at Chilbolton. However, when rain is present in the common volume of the transmitter and receiver, a clear signal is seen. The Figure shows the statistics of the transmission loss over a two-year period. The transmission loss is defined as the ratio of the transmitted power to the received power. The power received from a volume containing rain can be calculated using the bistatic radar equation of which an approximate form is:

$$P_r = \frac{P_t G_t G_r \lambda^4 N V |S(\theta,\varphi)|^2}{64\pi^4 R_1^2 R_2^2} \qquad (12.15)$$

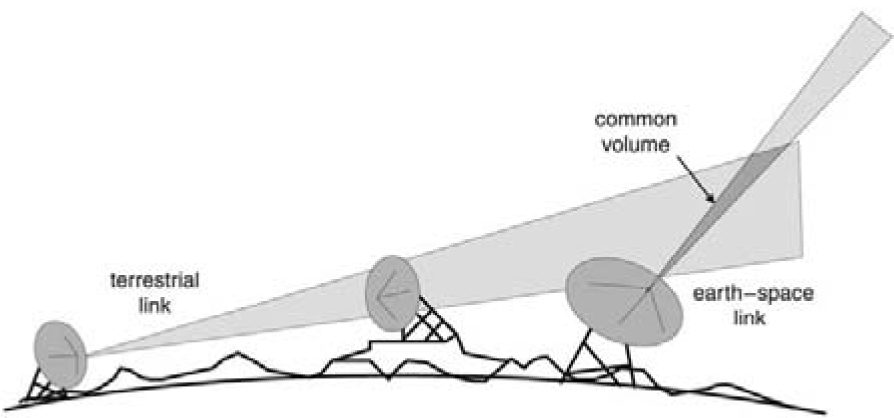


Figure 12.14 An illustrative rain-scatter interference geometry

Here R_1, R_2 are distances from the antennas to the beam intersection volume, t denotes transmission and r denotes reception, G denotes antenna gain and P power. V is the volume of the beam intersection and is assumed to contain N particles per unit volume with scattering function S at the appropriate angle. In using this equation, careful account must be taken of the two antenna polarisations; as it stands, the equation does not take account of attenuation. At the higher frequencies, attenuation on the paths to and from the scattering volume, and within it, is very significant and may lead to intense rain producing less scatter than moderate rain. Attenuation also limits the distance over which interference is likely to be significant at frequencies above 30 GHz.

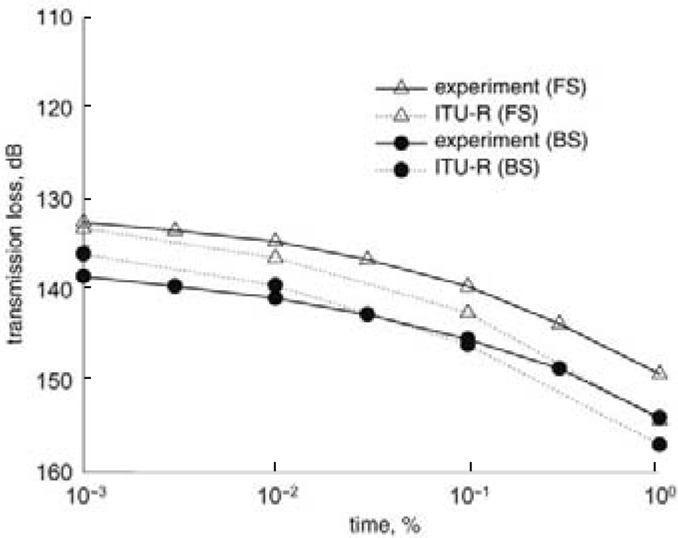


Figure 12.15 Statistics of transmission loss at 11.2 GHz, tx-rx distance 131 km (FS – forward scatter; BS – backward scatter)

Except in the Rayleigh and optical regions, the scattering functions $S(\theta\varphi)$ of raindrops exhibit very complex angular variations and require detailed calculation. For example, Figure 12.16 shows the bistatic reflectivity factor as a function of scattering angle (θ) for a rainfall rate of 20 mm h^{-1} at a frequency of 11 GHz (for further details of this and other aspects of interference due to hydrometeor scatter, see Reference [19]). It is obvious from this Figure that the scatter geometry can affect the scattered signal by tens of dBs when the plane of polarisation is parallel to the scatter plane.

It is important to appreciate that a signal arriving *via* bistatic scatter may be very different in its time waveform and spectral shape from the transmitted signal, as it is essentially incoherent, and randomly fluctuating. The transfer function of the path is strongly dispersive and frequency selective in both ampli-

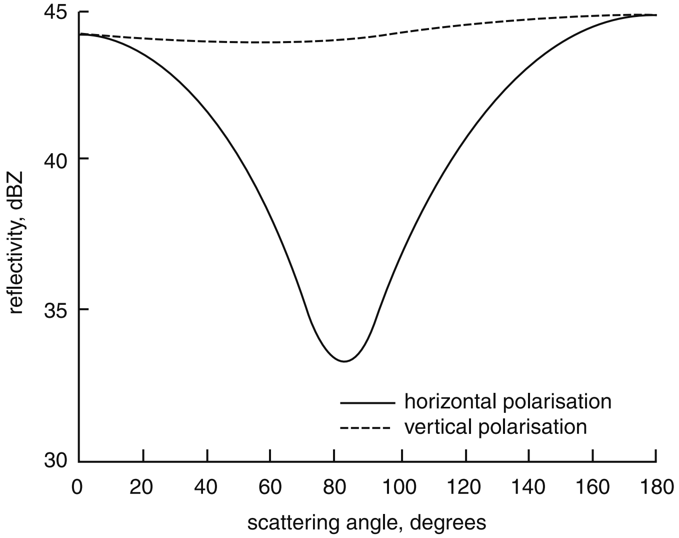


Figure 12.16 Angular dependence of rain scatter at 11 GHz (rainfall rate of 20 mm/h)

tude and phase, and varies randomly. The arriving signal may be described roughly as the sum of geometry, the spread of time delays between the individual paths may be of the order of microseconds, and the relative phasing between them changes as the particles move. A consequence of this is that it may be difficult to distinguish interference due to rain scatter from a decrease in signal-to-noise due to attenuation of the wanted signal, particularly as the two phenomena will usually occur together.

If a CW signal is transmitted, the bistatic signal is usually Rayleigh distributed and fluctuates on a time scale of milliseconds. It is believed that if a signal with a modulation bandwidth $>>1/T$ is transmitted, where T is the time delay spread of the paths, the received signal will resemble Gaussian noise. The short-term spectrum of this noise will display intense frequency selectivity on a scale of $1/T$, with the structure itself changing randomly on the millisecond scale.

12.10 References

1 MARSHALL, J.S., and PALMER, W. McK.: 'The distribution of raindrops with size', *J. Meteor.*, 1945, **5**, pp.165–166

2 LAWS, J.O., and PARSONS, D.A.: 'The relation of rain-drop size to intensity', *Trans. Amer. Geophys. Union*, 1943, **24**, pp.432–460

3 ULBRICH, C.W.: 'Natural variations in the analytical form of the raindrop size distribution', *J. Clim. Appl. Meteorol.*, 1983, **22**, pp.1764–1775

4 HARDEN, B.N., NORBURY, J.R., and WHITE, W.J.K.: 'Use of a log-mormal distribution of raindrop sizes in millimetric radio attenuation studies'. IEE Conf. Publ. 169, Part 2, pp.87–91

5 MEDEIROS FILHO, F.C., COLE, R.S., and SARMA, A.D.: 'Millimetre-wave rain induced attenuation: theory and experiment'. IEE Proc. H, 1986, **133**, pp.308–314
6 MANABE, T., IHARA, T., AWAKA, J., and FURUHAMA, Y.: 'Rain attenuation statistics of millimeter waves and raindrop size distribution'. Proceedings URSI Commission F Symposium on *Wave propagation: remote sensing and communications*, University of New Hampshire, Durham, New Hampshire, USA, 1986, pp.10.5.1–10.5.4
7 MASON, B.J.: 'The Physics of Clouds' (Oxford University Press, London, England, 1971, 2nd edn.)
8 KLAASSEN, W.: 'Attenuation and reflection of radio waves by a melting layer of precipitation', *IEE Proc. H* 1990, **137**, pp.39–44
9 RAY, P.S.: 'Broadband complex refractive indices of ice and water', *Appl. Opt.*, 1972, **11**, pp.1836–1844
10 VAN DE HULST, H.C.: 'Light scattering by small particles' (Dover Publications, Inc., New York, USA, 1981)
11 HOLT, A.R., UZUNOGLU, N.K. and EVANS, B.G.: 'An integral equation solution to the scattering of electromeagnetic radiation by dielectric spheroids and ellipsoids', *IEEE Trans. Antennas Propag.*, 1976, **26**, (**5**), pp.706–712
12 D'AMICO, M., HOLT A.R. and CAPSONI, C.: 'An anisotropic model of the electromagnetic properties of the melting layer', *Radio Sci.*, **33**, pp.535–552
13 PRUPPACHER, H.R., and PITTER, R.L.: 'A semi-empirical determination of the shape of clouds and raindrops', *J. Atmos. Sci.*, **28**, pp.86–94.
14 GODDARD, J.W.F. and BAPTISTA, J.P.V.: 'Multi-parameter radar applied to propagation studies', Olympus *Utilisation* conference, Seville, Spain, 1993, ESA Pub. WPP-60, pp.609–613
15 Recommendation ITU-R P.530–9: 'Propagation data and prediction methods for the design of terrestrial line-of-sight systems', 2000
16 BRUSSAARD, G.: 'A meteorological model for rain-induced cross-polarisation', *IEEE Trans. Antennas Propag.*, 1976, **24**, (1), pp.5–11
17 HAWORTH, D.P., MCEWAN, N.J., and WATSON, P.A.: 'Relationship between atmospheric electricity and microwave radio propagation', *Nature*, 1977, **266**, pp.703–704
18 'Influence of the Atmosphere of Radiopropagation on Satellite Earth Paths at Frequencies Above 10 GHz, (COST Project 205)', Commission of the European Communities, 1985, report EUR 9923.
19 'Influence of the atmosphere on interference between radio communications systems at frequencies above 1 GHz. (COST Project 210), Commission of the European Communities, 1991, report EUR 13407 EN
20 Recommendation ITU-R P.452-9: 'Prediction procedure for the evaluation of microwave interference between stations on the surface of the earth at frequencies above about 0.7 GHz', 2000.

Chapter 13

The propagation aspects of broadcasting at VHF and UHF

J. Middleton

13.1 Introduction

Broadcasting is defined in the ITU Radio Regulations as 'a radiocommunication service in which the transmissions are intended for direct reception by the general public'. European broadcasting at VHF and UHF takes place in the frequency ranges shown in Figure 13.1. In the UK, 88–108 MHz is used for FM radio, 218–230 MHz is used for digital radio and 470–860 MHz is used for analogue and digital television.

The propagation of radiowaves relies on fundamental principles and is

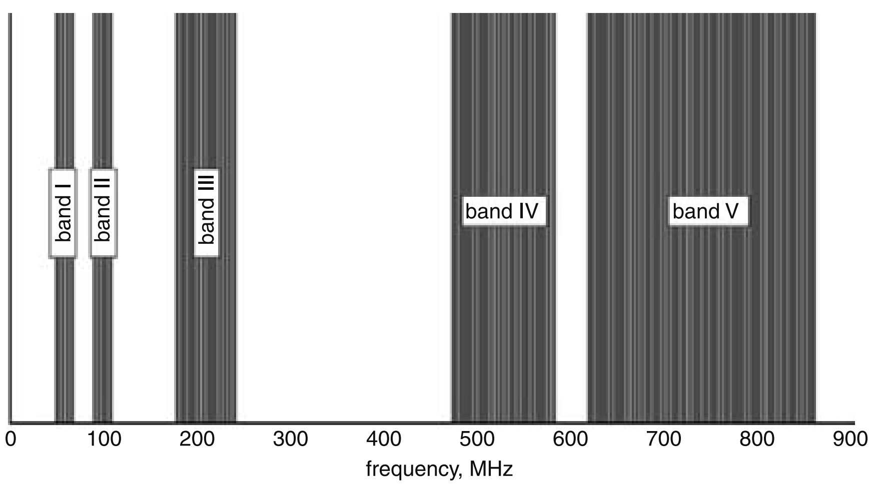


Figure 13.1 Spectrum allocation for VHF/UHF broadcasting

obviously independent of the application. However, the particular requirements of the broadcaster, where transmitters are required to cover large areas, with little control over the receiving installation, mean that certain aspects of propagation have greater significance than for, say, a point-to-point link.

13.2 Limit of service determination

Broadcasters define coverage in terms of field strength, unlike point-to-point services where the concept of transmission loss is used. For example, the limit of service for television in Band IV is 64 dBμV m^{-1}. Using a typical roof-mounted receiving antenna, this level of field strength will give sufficient signal at the television receiver to provide a noise-free picture.

This limit of service field strength is based on the requirements of the service. For example, it can be shown that the minimum acceptable signal/noise ratio at the video output of a domestic television receiver is about 33 dB:

$$\frac{S}{N} = 20 \log \left[\frac{Pk - Pk \text{ picture voltage}}{\text{RMS noise voltage}} \right] = 33 \text{ dB}$$

For a vestigial sideband system, the corresponding RF signal/noise = 33 + 8 = 41 dB.

The thermal noise voltage at the input of a domestic TV receiver is given by

$$V_N = 10 \log[k t_0 \, b r] + N + 120 \text{ dB}\mu\text{V m}^{-1}$$

where

k	$= 1.38 \times 10^{-23}$	the Boltzman constant
t_0	$= 290\text{K}$	reference temperature
b	$= 5.5 \times 10^6$ Hz	receiver bandwidth
r	$= 75\ \Omega$	receiver input impedance
N	$= 7$ dB	receiver noise figure

Therefore $V_N = 9$ dBμV and the minimum voltage at the receiver input:

$$V_R = 9 + 41 = 50 \text{ dB}\mu\text{V}$$

This input voltage is related to the field strength E at the receiving antenna by the following equation:

$$E = V_R + L - G + 20 \log f - 33.6 \text{ dB}$$

where, in this example:

L	$= 4$ dB	feeder loss
G	$= 10$ dB	antenna gain
f	$= 495$ MHz	centre of Band IV

Table 13.1 Typical coverage limits

Frequency band	*Band II*	*Band III*	*Band IV*	*Band V*
Type of service	FM radio reception at 10 m for 50% of locations	DAB reception at 1.5 m for 50% of locations	analogue TV reception at 10 m for 50% of locations	analogue TV reception at 10 m for 50% of locations
Field strength at edge of coverage area (dBμV m^{-1})	54	35	64	70

Hence, the minimum field strength, $E = 50 + 14 = 64$ dBμV m^{-1}. Table 13.1 shows typical coverage limits for other services.

The reason for quoting coverage area limits in this way is because the broadcaster has no control over the receiving installation. However, by making certain reasonable assumptions, he can provide a field of sufficient intensity to guarantee that the viewer/listener will receive a high quality service.

13.3 Protection ratios

Well beyond the normal coverage limit of a transmitter, when the signal has fallen to a sufficiently low level, the frequency can be reused, either for the same service or for a different one. Clearly, the broadcaster must be certain that the distant transmission, despite its low level, is not going to cause interference to the new one. To do this, the concept of protection ratio is employed. Its definition is as follows:

protection ratio for a given level of impairment = difference in dB between the wanted and unwanted signal powers at the input to the receiver

where impairment is defined according to an ITU-R scale as follows:

Grade 5 imperceptible
Grade 4 perceptible but not annoying
Grade 3 slightly annoying
Grade 2 annoying
Grade 1 very annoying

Grade 3 (slightly annoying) is normally used when determining protection ratios for the television service. For a 625-line system I-PAL transmission interfered with by another similar transmission on the same frequency, the required protection ratio is 40 dB for continuous interference. Tropospheric refraction effects can cause VHF and UHF signals to travel well beyond the horizon, although usually only for short periods of time (typically for between one and ten per cent of the time). For this reason, the protection ratio is relaxed to 30 dB for tropospheric interference.

As an example, consider the 1 MW Crystal Palace television transmitter located on high ground near Croydon. It has a 64 dBμV m^{-1} service area with a radius of approximately 50 km and uses channels 23, 26, 30 and 33. How far away can these channels be reused without being interfered with by Crystal Palace?

Assuming a 30 dB protection ratio, the level of Crystal Palace at the edge of the new service area must not exceed $64 - 30 = 34$ dBμV m^{-1}. The ITU field strength prediction method, Recommendation ITU-R P.370-7[1], indicates that this field strength level would occur about 240 km from the Crystal Palace transmitter. In reality, terrain effects can modify this distance. For instance, the Wrekin transmitter, near Shrewsbury, uses the same channels as Crystal Palace but is only 219 km away.

Protection ratio values depend on the types of service being considered and on their frequency separation. In general, maximum values occur when the interfering signal has the same frequency as the wanted service and fall for increasing frequency difference. Figure 13.2 shows the protection ratio curve for a television signal interfered with by a narrowband or CW signal.

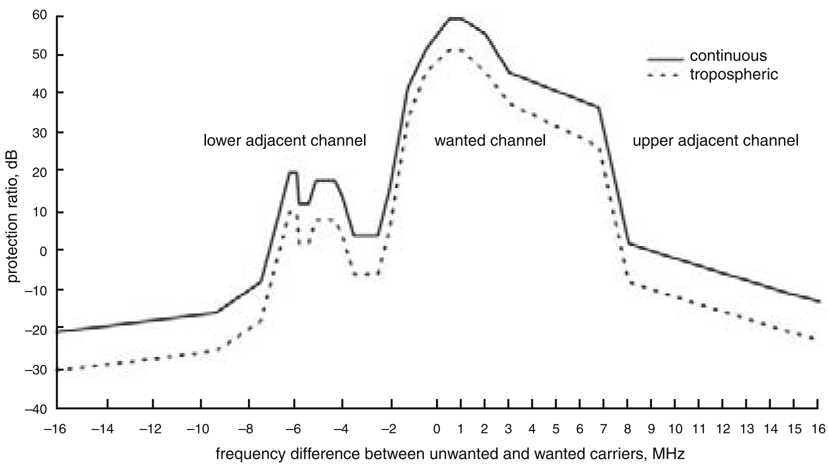


Figure 13.2 Typical protection ratio curves. Vision protection ratios for 625-line, system I-PAL television interfered with by a CW signal

[1] This recommendation has been replaced by Recommendation ITU-R P.1546

13.4 Characteristics of propagation

13.4.1 Ground reflections and surface roughness

It is shown in Chapter 6 (Section 6.5) that direct and ground reflected signals can, under certain circumstances, pass periodically in and out of phase as path length increases and produce a resultant which varies by +6 dB to −30 dB or less relative to the direct signal, as shown in Figure 13.3.

It could be argued that, in this example, the height of the receiving antenna, shown as 100 m, does not represent a realistic situation. However, this is the height above the reflecting surface and could easily occur in reality, for example a sloping hillside above a flat plain. The main point of this example is to demonstrate, in theory, that large areas could be unserved owing to two-ray effects. The main requirement is for the ground surface to be considered smooth enough to produce a specular reflection, analogous to the surface of a mirror. If the ground surface is rough, the signal would be scattered after reflection. The received signal would then consist of a large number of reflected components arriving in random phase relative to the direct signal component and the occurrence of deep fades would be reduced.

A convenient criterion for establishing the greatest degree of roughness, still supporting specular reflection, was first developed by Rayleigh, see also Chapter 6 (Section 6.7). Figure 13.4 shows a reflecting surface with roughness represented by rectangular steps of depth d. A ray striking the surface at point A has a path length from transmitter to receiver which is shorter than that which strikes the lower surface at point C. The path difference is given by $AC - AB$ from which the phase difference is given by:

$$\Delta\phi = \frac{4\pi d \sin\theta}{\lambda}$$

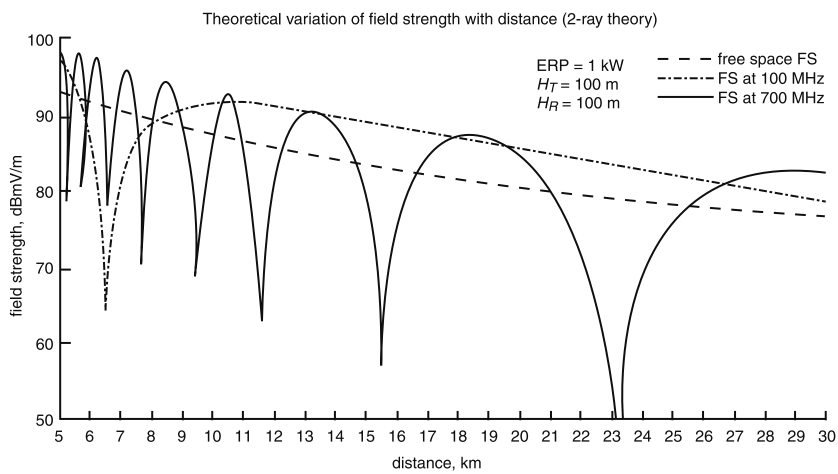


Figure 13.3 Example of two-ray theory. Theoretical variation of field strength with distance

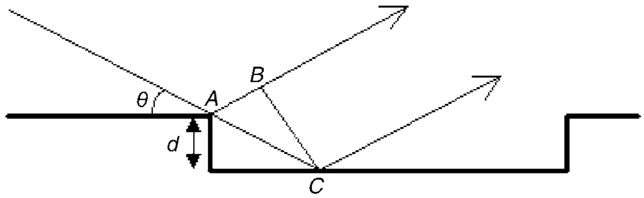


Figure 13.4 Rayleigh roughness criterion

A surface is usually considered to be rough if this phase difference exceeds $\pi/4$ radians.

Thus, for typical grazing angles of 1°, the criterion is satisfied for $d > 10$ m at 100 MHz or $d > 2$ m at 600 MHz. These figures show that at Band II frequencies and small grazing angles, specular reflection is likely to occur, even where there are quite substantial surface irregularities, such as might occur in a residential area. Conversely, at Bands IV and V, ground surfaces may generally be considered as rough. Thus, large field strength variations, resulting from direct and ground reflected signals being out of phase, are more prevalent at lower VHF than at UHF.

13.4.2 Other multipath effects

The foregoing considers the propagation from transmitter to receiver as a two-dimensional vertical slice. Only two rays are considered, the direct and the ground reflected. This is a simplification and, in the three-dimensional world, there are likely to be many more reflections from hills and buildings on either side of the direct path. Tall buildings in urban areas, because of their smooth flat surfaces, can cause serious problems to broadcast reception. Even when line of sight to the transmitter, the direct signal arriving at the antenna of a radio or television receiver is likely to be accompanied by several reflected rays arriving slightly later in time because of the longer path lengths. The effect can be seen on television receivers in the form of delayed images slightly to the right of the main picture. In the case of FM radio, multipath effects cause the demodulated audio to sound distorted.

Another multipath effect is found where tidal water exists on the path between a transmitter and the target service area, for example across an estuary. Provided the surface of the water is not too disturbed, specular reflection can easily occur at UHF. At certain times of day, when the sea level is at the right height, the reflected ray can interfere destructively with the direct ray and result in deep fades lasting for several hours.

For domestic antenna installations, using a directional yagi fixed at roof height, these multipath effects can be minimised by careful positioning of the antenna. However, multipath reception in a mobile environment is more difficult to overcome because of the low height, omnidirectional antennas which are inevitably used on vehicles. Only now, with the development of advanced digital modulation methods such as COFDM, can this problem be overcome.

13.4.3 Diffraction

Again, diffraction theory is discussed in detail elsewhere. Since, to some extent, diffraction loss is inversely proportional to wavelength then transmissions in Bands IV and V will be attenuated more over the same obstacle than at Band II. The result of this is that the UHF television service needs many more relay stations than the Band II FM radio service since coverage deficiencies occur wherever there are significant terrain obstructions. For example, within the Crystal Palace television service area, which has a radius of approximately 50 km, over 50 relay stations have been built to serve those populated areas which have no service from the main transmitter. For the whole of the UK and Northern Ireland, there are 51 main stations and over 1000 relay stations. This effect is less of a problem at Band II since the longer wavelengths of the FM sound transmissions allow the signals to diffract round hills with less attenuation than at UHF. Consequently, fewer stations are required for this service.

13.4.4 Tropospheric refraction

At distances greater than about 100 km, VHF/UHF signal levels can become highly variable with time due to variations in the refractivity of the troposphere. Under certain conditions, inversion layers and ducts are formed which allow the signals to travel well beyond the horizon. Since this is a meteorological mechanism, signal strength can vary on a seasonal and even diurnal basis. Figure 13.5

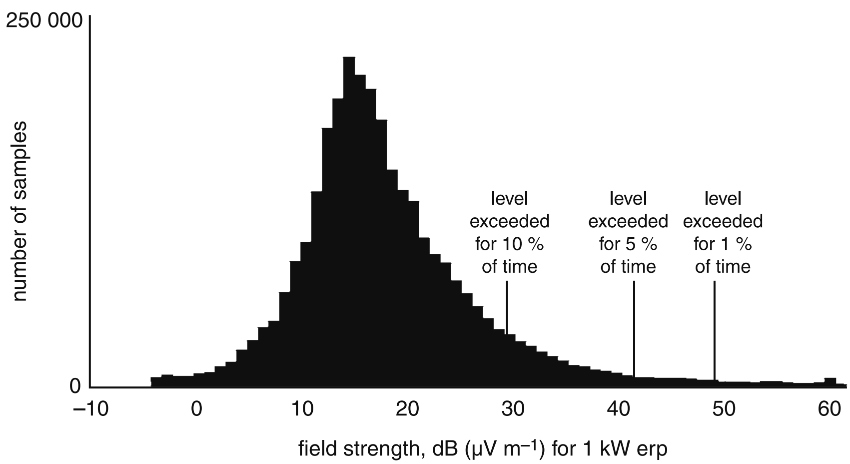


Figure 13.5 Typical field strength distribution showing enhanced propagation effects. Field strength distribution for Mendip to Daventry transmission at 218 MHz, March 1995 to February 1996; path length = 152 km

shows the variation of a test transmission on 218 MHz, from Mendip to Daventry, over a 12-month period between 1995 and 1996.

It can be seen that for low percentages of time the signal levels are several tens of decibels greater than the median level.

Enhanced propagation tends to occur during anticyclonic weather and, for land paths, a distinct diurnal effect can often be observed, with maximum signal levels occurring during the night. This can be explained as follows. After a fine clear day, the earth's surface may cool quickly by radiation after sunset, giving rise to an increase of air temperature with height in the lower troposphere which leads to the formation of an inversion layer and ducting. The following morning, after sunrise, the inversion layer is destroyed by solar heating and the signal level falls. This effect is clearly demonstrated in Figure 13.6.

Propagation ducts are also formed over sea paths due to variations of humidity and temperature with height. Unlike land paths, diurnal variation occurs less frequently and enhanced propagation can last for several days, as seen in Figure 13.7.

These tropospheric effects can have serious consequences when frequencies have to be reused in other parts of the country and careful planning is required to minimise interference problems. In some cases, directional transmitting antennas have to be used, particularly where interference to continental Europe is likely. In the case of television reception, where fixed, roof-mounted receiving antennas are used, it is assumed that they will be directional to discriminate against unwanted interference.

13.5 The measurement of field strength

As seen in Section 13.2, coverage is defined in terms of field strength, unlike point-to-point services where the concept of transmission loss is used. At some

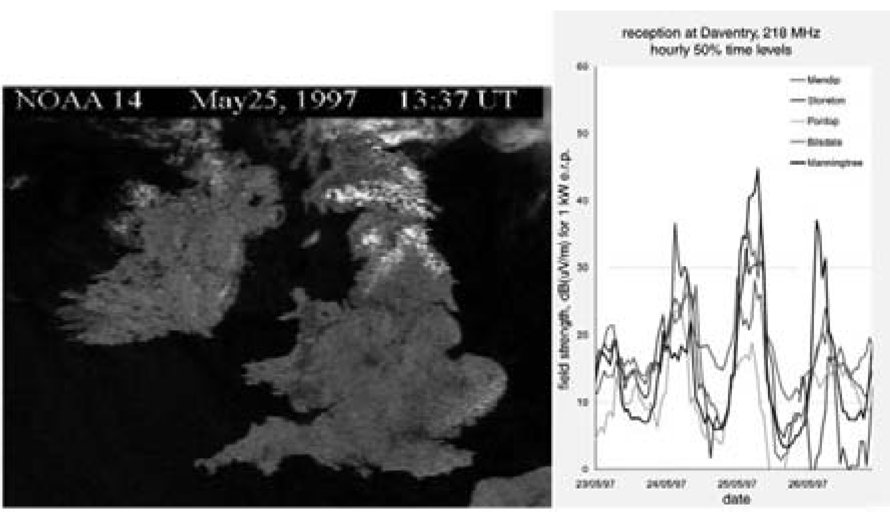


Figure 13.6 Example of diurnal variation over a land path

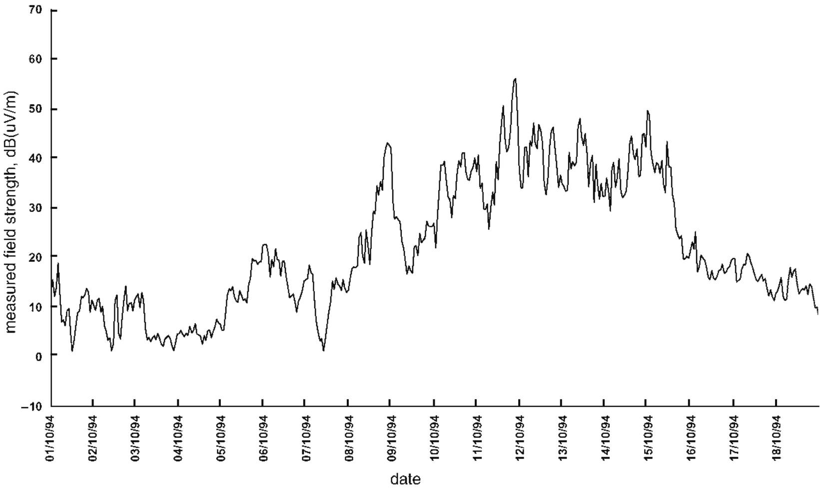


Figure 13.7 Example of sea path propagation. Kippure to Walney Island, 223 MHz – hourly 50% time levels

stage in the development of a new service, the broadcaster will need to confirm this coverage. To do this, a calibrated measuring antenna and receiver are used.

A field strength measuring receiver is essentially a tuneable RF voltmeter. It enables the level of a signal, when applied to its input terminals, to be measured when tuned to the frequency of that signal. This level may be displayed on a meter or LCD or transferred on a data link to a computer. Most commercially available receivers allow a choice of bandwidth and detector type commensurate with the transmission system being measured.

The measuring antenna is usually a ruggedised yagi or log-periodic design the gain of which is accurately known. When the antenna is placed in an RF field of E dBμV m^{-1}, a voltage is induced at its terminals which can be measured on the receiver. The voltage, V dBμV, at the receiver is related to the field strength, E dBμV m^{-1}, at the antenna by

$$E = V + \underbrace{L - G + 20\log f - 32}_{\text{antenna factor}}$$

where

L is the loss in dB of the feeder which connects the antenna to the receiver
G is the gain in dB of the antenna in dB relative to a $\lambda/2$ dipole,
f is the frequency in MHz

and a characteristic impedance of 50 Ω is assumed.

The sum of the terms to the right of V is known as the antenna factor, i.e. the number of decibels which must be added to the receiver voltage to give the equivalent field strength. The derivation of this equation is given in Section 13.8.

Two methods of measurement are normally used. For the television service, where it is assumed that the domestic receiving antenna is roof mounted, the measuring antenna is mounted on a pneumatic mast so that it can be elevated to a height of 10 m. The field strength is then measured with the vehicle stationary although, at UHF, particularly in urban or suburban areas, the measurements are made over a short distance to average out any multipath effects. This method, although giving an accurate assessment of the coverage, can be a very time-consuming process.

For evaluating in-car radio reception, the measuring antenna is usually a $\lambda/4$ whip mounted on the vehicle roof. The field strength is measured while the vehicle is moving and the results are logged on a computer. Some form of positional information can also be fed to the computer either from a GPS receiver or some other radio-based location system.

13.6 The prediction of field strength

When planning a new service the broadcaster needs to know where to build the transmitters and what powers to radiate to get the most effective coverage. We have seen in Section 13.3 that he also needs to know where he can reuse the same frequencies without causing interference. To do this some form of field strength prediction technique is required. There are two approaches to propagation prediction, empirical and deterministic.

A well known empirical method, recommended by the ITU for many years, is ITU Recommendation P.370. It has been used for VHF and UHF planning, particularly where coordination is required between neighbouring countries. The method was based on a set of curves of field strength plotted against distance, obtained from an extensive measurement campaign. The curves allowed field strengths to be predicted for 50, 10, 5 and 1 per cent time for land and sea paths. Although path profile information was not essential, accuracy was improved through a series of corrections for which knowledge of the terrain at both ends of the propagation path was required. Although not generally renowned for its accuracy of prediction, this method has been widely used for national and international planning since it had the benefit of impartiality. Similar methods may now be used with the current successor Recommendation, ITU-R P.1546.

The deterministic approach attempts to mathematically model the propagation path and associated propagation conditions. The so-called BBC method is an example of this, and most broadcasting organisations have similar methods. A terrain height database is required from which is constructed a path profile. This determines the existence of sea or land and characterises the terrain into a series of discrete geometric shapes, i.e. smooth earth, knife edge, wedge or cylinder. A path loss calculation is then made according to how the profile approaches or intrudes into the line between the transmitter and receiver. Figure 13.8 shows a typical example.

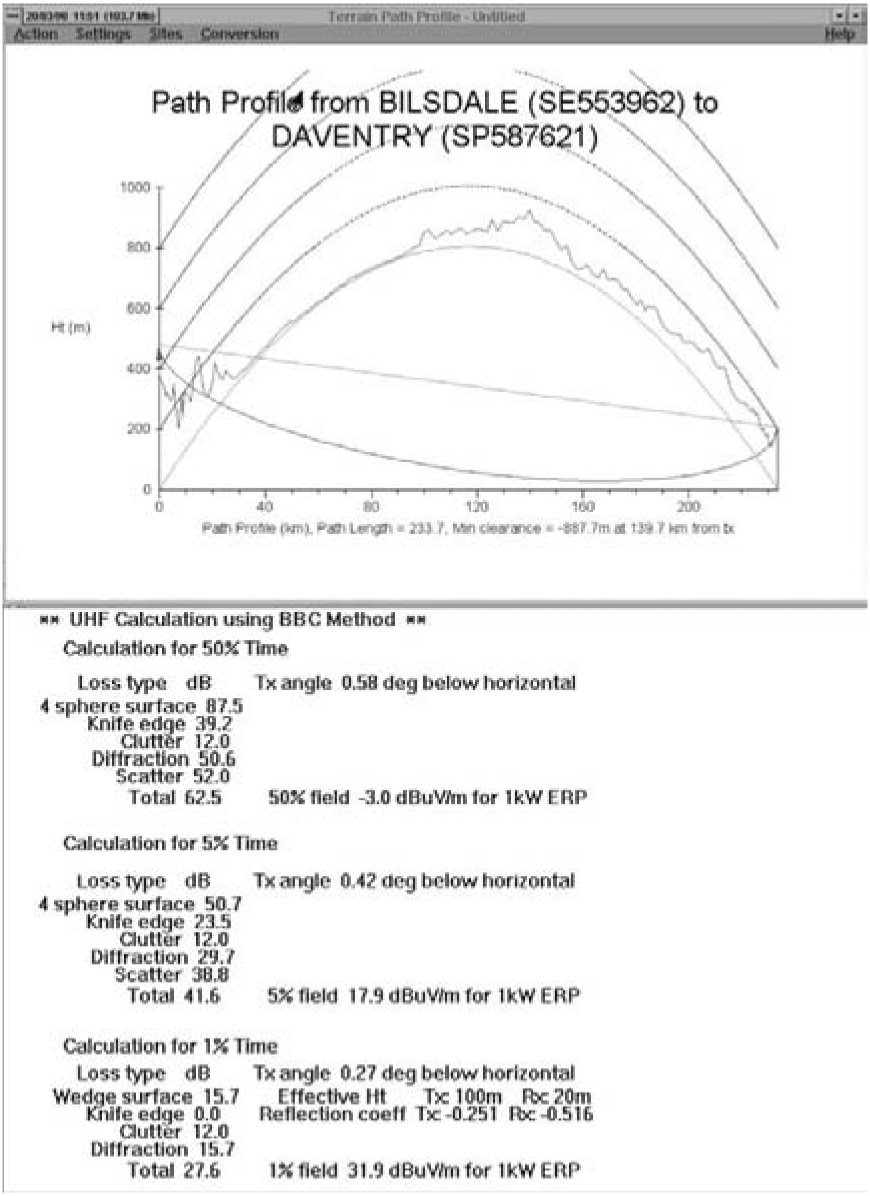


Figure 13.8 Example of field strength prediction using terrain data

For land paths an interpolation is made between a multiple knife-edge loss and a loss calculated over a stylised surface (wedge or cylinder). For sea paths, a spherical surface loss calculation is used. The path loss is completed by adding a clutter loss to take account of buildings and trees. Account is also taken of tropospheric effects such as ducting and scatter propagation. To increase the accuracy of the results the theoretical calculations are modified by an optimisation factor based on many field strength measurements made in populated areas. The standard deviation of the prediction error is normally about 7 dB.

The deterministic method tends to be more accurate than the empirical one although not as much as might be expected. This is mainly because it only models a single path between transmitter and receiver. In reality, the level of the received signal is the resultant of many additional signals due to reflections from hills or buildings. As a result, these programs are used only to provide a rough guide of transmitter coverage. Their main application is to calculate low-percentage time interference levels in conjunction with network planning software.

Nevertheless, modern high-speed computers, with large memories, are enabling new approaches to field strength prediction to be considered, such as the parabolic equation method. This method models the propagation environment between the transmitter and receiver on a three-dimensional basis and has the potential for making more accurate predictions. However, it may be several years before such methods are used on a regular basis for planning purposes.

13.7 Digital broadcasting

So far, the discussion has been about conventional analogue broadcasting systems. New digital techniques are now being developed which have the potential for greatly improving the quality and efficiency of broadcasting.

13.7.1 DAB

The implementation of digital audio broadcasting (DAB) is well advanced in the UK and uses a COFDM (coded orthogonal frequency division multiplex) system which has been standardised throughout Europe by the EUREKA 147 consortium. UK transmissions are in Band III (218 to 230 MHz) and comprise a 1.5 MHz wide block which can carry five high quality audio services.

A fundamental feature of the system is the ability to operate satisfactorily in areas having high levels of multipath propagation. This is achieved by the incorporation of a guard interval in the time domain. Provided the longest multipath delay time does not significantly exceed this guard interval then all received signal components add constructively. Therefore, the system is capable of providing excellent coverage to moving vehicles.

From the viewpoint of signal processing within the receiver, a multipath signal is indistinguishable from another suitably delayed transmission carrying exactly the same modulation. It follows that a network of DAB transmitters can cover a large area using a single frequency block without mutual interference. This concept of the single frequency network (SFN) makes for high spectral efficiency which is ideally suited to national broadcasting.

Another advantage of the SFN is the concept of network gain. This arises from the fact that, in many areas, the coverage from adjacent transmitters will overlap hence signal powers are added in the receiver. There is also a statistical component due to the location variation distributions of the different fields. The overall standard deviation of the composite received signal is less than that of a single signal. Hence, the power margin to achieve 99 per cent of locations relative to, say, 50 per cent coverage is less than for a single transmitter case.

Signals which arrive outside the guard interval will degrade reception of the main signal and this can happen with distant DAB transmissions, albeit from the same network, when they are received due to tropospheric refraction. The effect will be to cause service areas to shrink slightly. If this becomes a problem in a particular area then, being a single frequency network, the solution is relatively easy – provide a low power fill-in station to serve that area.

13.7.2 DTT

Digital terrestrial television (DTT) is also being implemented in the UK – the first stage of the transmitter network, consisting of 80 transmitters, is now complete. However, unlike DAB, the requirements of the system mean that a single frequency network is not feasible. Instead, six 8 MHz wide blocks are radiated from each transmitting site on frequencies interleaved with the existing UHF television channels. Each block or multiplex can carry up to five programs with a quality comparable to system I-PAL. Clearly there is potential for interference to and from the existing analogue television transmissions and great care has gone into defining and measuring the transmitting antenna radiation patterns. In some cases, directional antennas having patterns with 20 to 30 dB notches have been specified.

13.8 Appendix

13.8.1 Field strength measurement and the antenna factor

If we consider a halfwave dipole in an electric field of e $V\,\mathrm{m}^{-1}$ the voltage induced across the terminals of this dipole is $e.\lambda/\pi$ V (see Figure 13.9*a*). The derivation of this voltage is given in Section 13.8.2. The equivalent circuit of such an antenna, terminated in a load resistance R_L, is shown in Figure 13.9*b* where R_A is the radiation resistance of the antenna.

The voltage across the load, v_L is given by:

$$v_L = \frac{e\lambda}{\pi} \frac{\mathrm{iR_L}}{R_A + R_L} \quad \mathrm{V}$$

It is usual to express field strength in dBμV m^{-1} hence:

$$V_L = E + 20 \log\frac{\lambda}{\pi} - 20 \log(1 + \frac{R_A}{R_L}) \quad \mathrm{dB\mu V/m}$$

where $V_L = 20 \log v_L$ and $E = 20 \log e$.

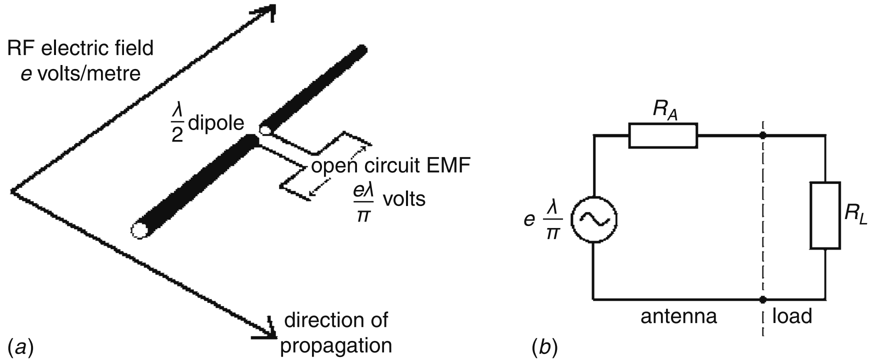


Figure 13.9 Sketch of a dipole antenna in an electromagnetic field, and its equivalent circuit

If the load is matched to the antenna, i.e. $R_A = R_L$ then:

$$V_L = E + 20\log\frac{\lambda}{\pi} - 6 \quad \text{dB}\mu\text{V}$$

The radiation resistance of a dipole approximates to 73 Ω in free space, even at 10 m AGL for VHF and UHF. However, receivers normally standardise their input impedance to 50 Ω so that impedance conversion is necessary to obtain maximum power transfer from antenna to measuring receiver. This is done with a matching transformer, usually in the form of a balun since most receivers have an unbalanced input. The equivalent circuit of a 50 Ω dipole is shown in Figure 13.10*a* with a further simplification in Figure 13.10*b*.

When such an antenna is terminated with 50 Ω then the voltage across the load is given by:

$$V = E + 20\log\frac{\lambda}{\pi} + 10\log\frac{50}{73} - 6 \quad \text{dB}\mu\text{V}$$

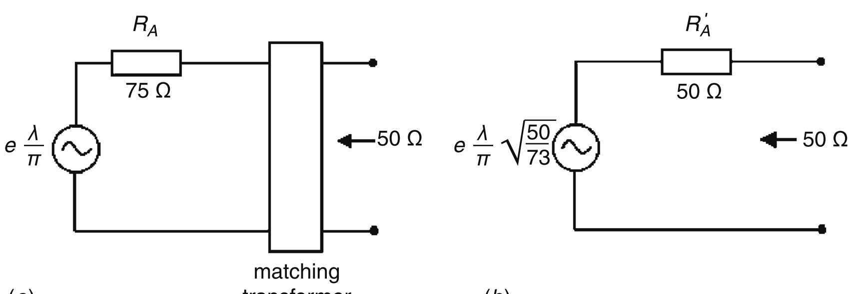


Figure 13.10 Equivalent circuits for matching to 50 Ω

which reduces to:

$$V = E + 20\log\frac{\lambda}{\pi} - 7.6 \quad \mathrm{dB\mu V}$$

If the antenna in use has a gain G dB relative to a $\lambda/2$ dipole and is connected to the receiver through a feeder with a loss L dB then the equation becomes:

$$V = E + G - L + 20\log\frac{\lambda}{\pi} - 7.6 \quad \mathrm{dB\mu V}$$

This can be rewritten as:

$$E = V + L - G + 20\log f - 32 \quad \mathrm{dB\mu V\ m^{-1}}$$

where f is the frequency in MHz.

The sum of the terms to the right of V is the antenna factor, i.e. the number of decibels which must be added to the receiver voltage to give the equivalent field strength.

Occasionally, equipment is encountered which has a characteristic impedance of 75 Ω. In this case the equation becomes: $E = V + L - G + 20\log f - 33.7$.

13.8.2 The voltage induced in a halfwave dipole

The effective area of an antenna is given by

$$\frac{\lambda^2}{4\pi}g \quad \mathrm{m^2}$$

where g is the antenna gain relative to an isotropic source.

If the antenna is placed in a field whose power density is s W m^{-2} the power extracted by the antenna is given by:

$$p = s\frac{\lambda^2}{4\pi}g \quad \mathrm{W}$$

A plane wave in free space with a field strength of e V m^{-1} has a power density given by:

$$s = \frac{e^2}{120\pi} \quad \mathrm{W\ m^{-2}}$$

Hence the power received by the antenna is:

$$p = e^2\frac{\lambda^2}{\pi^2}\frac{g}{480} \quad \mathrm{W}$$

The impedance of a $\lambda/2$ dipole in free space is 73 Ω. When connected to a matched load this power will be developed in the load and the equivalent open-circuit EMF, v_{oc}, is related to the received power by:

$$p = \frac{{v_{oc}}^2}{4.73} \quad \mathrm{W}$$

Hence, for a halfwave dipole ($g = 1.64$), the open-circuit EMF, v_{oc}, can be defined in terms of field strength e by combining these two power relationships to give:

$$v_{oc} = e\frac{\lambda}{\pi} \quad \mathrm{V}$$

The ratio λ/π is called the effective length of the halfwave dipole antenna.

Chapter 14

Propagation aspects of mobile spread spectrum networks

M.A. Beach and S.A. Allpress

14.1 Introduction to spread spectrum

Spread spectrum has been adopted as the core radio access method in many of the third-generation (3G) wireless standards [1], where high capacity networks supporting multimedia-like services are desired [2]. This selection was based on numerous research initiatives conducted both privately within numerous organisations [3], as well as publicly under initiatives such as the UK DTI LINK programme and the European RACE[1] and ACTS[2] programmes, addressing the selection of the primary air interface technique for the universal mobile telecommunication system (UMTS). Relevant examples include the RACEII [4] ATDMA and CODIT projects, as well as the UK DTI/SERC LINK CDMA [5] programme. More recently, a body called the Third Generation Partnership Project (3PP) [6] was established with the aim of harmonising the numerous proposals for 3G systems on a global basis.

Networks employing spread spectrum [7] access methods are extremely good examples of wideband radio access. Here a narrowband message signal is spread to a much wider bandwidth than is strictly necessary for RF transmission, then at the receiver the wideband representation is despread to yield an estimate of the original narrowband message signal. Ratios of 10 kHz message bandwidth to several MHz RF channel bandwidth are commonplace in such systems.

There are principally two techniques employed to spread the spectrum. In frequency hopping (FH) spread spectrum systems the narrowband message signal is modulated with a carrier frequency which is rapidly shifted in discrete increments in a pattern dictated by a spreading function. This spreading

[1] Research into Advanced Communications for Europe
[2] Advanced Communications Technologies Societies

function is also available at the receiver, and enables it to retune to the correct channel at the end of each hop. Although the simplest spread spectrum technique to visualise, only a few have attempted to commercially exploit this method for personal communication networks. The Geotek [8] system was one such embodiment and an analysis of FH propagation in urban cellular environments is given by Purle [9]. More recently, FH has been adopted in the Bluetooth [10] personal area network (PAN) standard. However, much of the current interest in spread spectrum systems is focused towards the implementation and deployment of direct sequence systems within 3G networks.

Direct sequence (DS) spread spectrum systems take an already modulated narrowband message signal and apply secondary modulation in the form of pseudo random noise (PN), thereby spreading the spectrum. This secondary modulation process is usually implemented in the form of phase shift keying (PSK), and the PN sequence is known as the spreading waveform or sequence, or code. At the receiver, the incoming spread spectrum waveform is multiplied with an identical synchronised spreading sequence thus resulting in the recovery of the information or message signal. By associating a unique spreading code with each user in the network, multiple users can be simultaneously overlaid in both the time and frequency domains, and the concept of code division multiple access (CDMA) realised. Given the considerable interest and investment made by the early pioneers of this technology (companies such as Qualcomm [11], InterDigital [12] and others [13]), the generic term CDMA has become synonymous with cellular systems specially employing the DS spread spectrum technique.

The user capacity of all cellular radio access techniques is self interference limited [13]. However, CDMA differs from frequency and time multiple access methods in as much as the primary source of this interference arises from users simultaneously occupying the same frequency and time allocations within both the same cell and cochannel cells. In CDMA terminology this is usually referred to as intra- and intercell interference respectively. The resultant capacity of such systems can be significantly increased by employing signal processing techniques at the receiver which are able to exploit the wideband propagation characteristics of the DS-CDMA channel. It is this fundamental aspect of high capacity CDMA cellular networks that is described in this Chapter.

14.2 Properties of DS-CDMA waveforms

In order to fully understand the propagation and sensitivity analysis of wireless networks employing CDMA access techniques, it is first necessary to consider the correlation properties of the spreading codes frequently employed in such systems. In the context of this discussion it is only necessary to consider the autocorrelation properties [15] of binary pseudo random or *m*-sequences, as this gives a measure with which the code can be distinguished from a time-shifted version of itself. *M*-sequences display an almost ideal two-valued autocorrelation function, $R_{ac}(\tau)$, as given in Eqn. 14.1, where N is the period of the sequence measured in chips[3]:

$$R_{ac}(\tau) = \sum_{i=1}^{N} a_i(a_{i=\tau}) = \begin{cases} 1, \text{ for } \tau = kN \\ -1/N, \text{ for } \tau \neq kN \end{cases} \tag{14.1}$$

Further, the normalised form of this function is illustrated in Figure 14.1. Here, it can be clearly seen that the autocorrelation function contains large triangular correlation peaks corresponding to the synchronisation of the waveforms, thus aiding both code acquisition and synchronisation functions which are necessary in the receiver in order to despread the DS-CDMA waveform.

In addition, it will be shown in the following sections that this technique can be used to resolve multipath energy present at the receiver in a manner similar to that employed in some wideband channel sounders [16].

The spectral occupancy, $S(f)$, of the DS-CDMA waveform can be derived from the autocorrelation function given above using the Wiener-Khintchine relationship. This is given by the following expression, where T_c is the chip duration of the spreading code, and it can be seen to consist of individual spectral lines within a $\sin^2 x/x$ envelope variation as illustrated in Figure 14.2:

$$S(f) = -\frac{1}{N}\delta(f) + \frac{N+1}{N^2}\operatorname{sinc}^2(\pi f T_c)\sum_{m=-\infty}^{\infty}\delta\left(f + \frac{m}{NT_c}\right) \tag{14.2}$$

The waveform illustrated in Figure 14.2 would indicate that DS-CDMA systems require an infinite RF bandwidth allocation, however in order to permit inter-working with other systems, the transmitted spectrum is filtered. Filtering does result in a reduction of the correlation function as illustrated in Figure 14.3 and

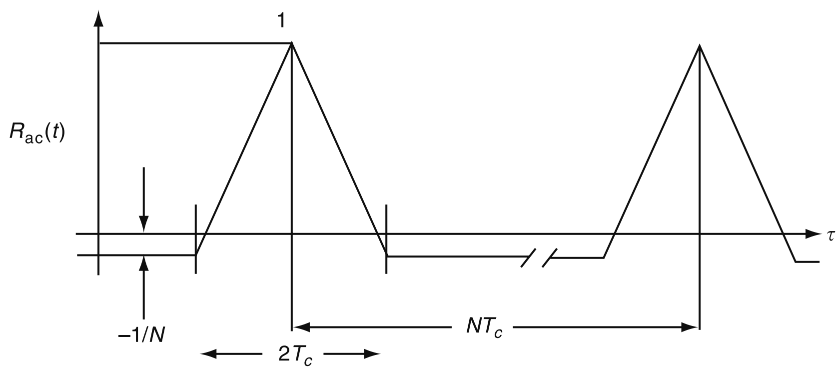


Figure 14.1 Normalised autocorrelation function of an m-sequence

[3] Each binary component of the pseudo random sequence is termed a chip in order to distinguish it from the message data bits, or symbols

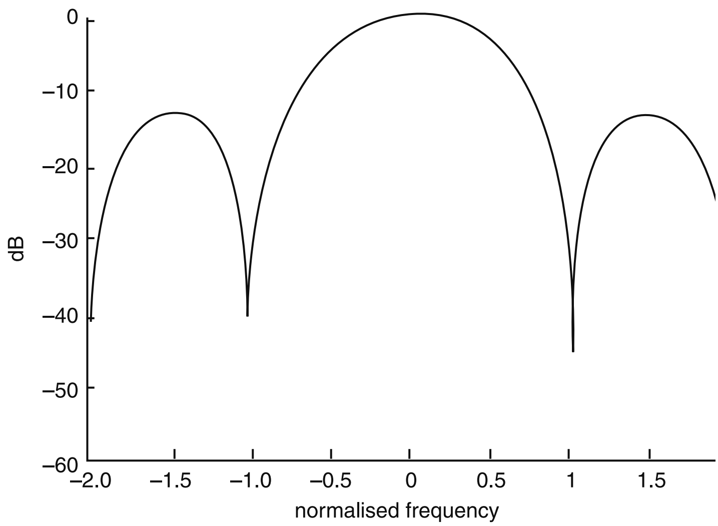


Figure 14.2 Envelope of the power spectrum density of an m-sequence

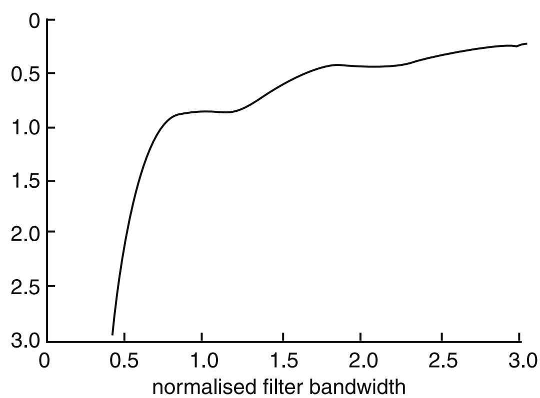


Figure 14.3 Correlation loss due to filtering

distortion of the triangular in-phase correlation peak as given in Figures 14.4*a* and *b*. These illustrations assume a conventional *m*-sequence waveform.

Practical realisations of CDMA systems employ filtering constraints similar to that shown in Figure 14.4*b*. For example, the Qualcomm IS95 CDMA cellular system [17] employs a 1.2288 Mchip s^{-1} spreading code rate operating within a baseband channel mask of 590 kHz passband and an ultimate stopband attenuation of 40 dB at 740 kHz. In the UMTS terrestrial radio access network (UTRA or UTRAN) the chipping rate is 3.84 Mchip s^{-1}, with the transmission mask specified in terms of adjacent channel leakage ratio (ACLR) [18].

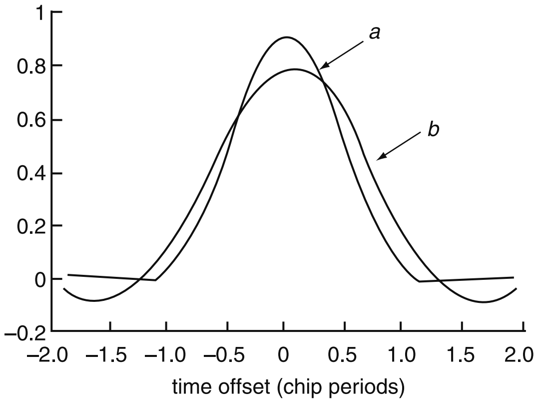


Figure 14.4 Distortion of correlation peak with a brick-wall filter passband equal to
(a) width of main-lobe
(b) central half of main-lobe

14.3 Impact of the mobile channel

The long-term fading statistics, path and shadow loss, of both narrowband and wideband signal propagation in the mobile radio channel have been shown to exhibit very similar characteristics according to theory and also by measurement [19]. However, the short-term statistics or the multipath fading characteristics of these channels differ significantly, since CDMA systems tend to operate using channel bandwidths far in excess of the coherence bandwidth of the mobile channel.

The impact of the mobile radio channel upon the CDMA waveform is most readily understood by considering the form of the correlation function after transmission through an environment similar to that shown in Figure 14.5*a*. The correlator output (see Figure 14.5*b*) shows three peaks corresponding to the time synchronisation between the local *m*-sequence in the mobile receiver and

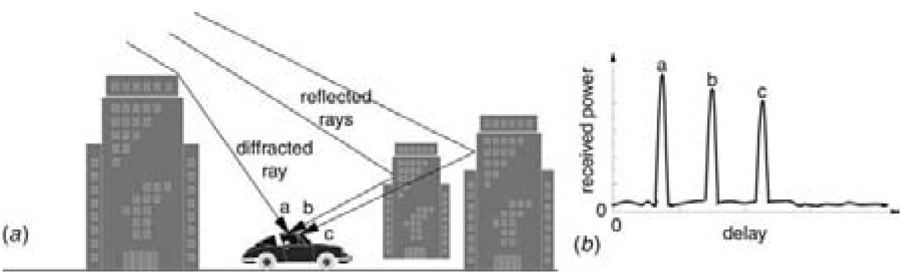


Figure 14.5 Resolution of multipath components using DS-CDMA
(a) urban cellular environment
(b) correlator output

the multipath components a, b and c. The relative delay between these peaks corresponds to the different path delays in the channel, and thus if the total multipath delay (T_m) is known, then the maximum number of multipath components (L_m) which can be resolved by the receiver is given by [20]:

$$L_m \leq \frac{T_m}{T_c} + 1 \tag{14.3}$$

Thus, it can be seen that the degree of multipath resolution is directly related to the chipping rate of the CDMA waveform. With reference to Figure 14.1, resolution of individual multipath rays would be possible if an infinite chipping rate was employed. As already stated, it is necessary to bandlimit such systems, and thus the multipath activity observed is clustered together in discrete time intervals, or bins, separated by T_c. Hence, as the mobile moves within the environment the signal energy within each bin of the correlator output will fade.

The severity of fading will depend upon the amount of multipath activity present in each bin, and this is related to both the physical nature of the scattering volume as well as the time resolution of the system. Since each bin of the correlator contains energy contributions from different multipath components, these will tend to fade independently and also have different instantaneous Doppler components. Thus, as shown in Section 14.5, the bin level fading statistics in an urban cellular environment vary from Rayleigh-like characteristics through to Rician distributions with relatively high K factors as the chipping rate is increased.

The inherent ability of the CDMA technique to produce multiple replicas of the same information bearing signal leads to the important concept of path diversity reception. Here, multiple despreading circuits can be used to decode the wanted signal contributions contained within the active bins of the correlator, and then combining these using the familiar techniques of switched, equal gain or maximal ratio combining in order to obtain a better estimate of the message signal. This is also known as rake reception based on the analogy with a garden rake being used to rake the impulse response of the channel, with each prong, branch or finger corresponding to a time bin in the correlator. This architecture is illustrated in Figure 14.6, where $r(t)$ is the received signal, $d(t)$ is the message estimate and $A_0\exp[j\phi L]$ corresponds to amplitude and phase offsets of the spreading waveform $c(t)$.

In addition to the time domain or temporal distortion caused by the multipath channel, it also follows from Figure 14.5*b* that the channel must also inflict severe frequency domain distortion. The received spectrum for a 10 Mchip s^{-1} system operating in a simulated urban environment is illustrated in Figure 14.7 using channel parameters as given in Table 14.1. Despite the numerous frequency domain nulls in the sinc envelope of the CDMA signal, data demodulation is still possible because of the inherent frequency diversity associated with spread spectrum systems.

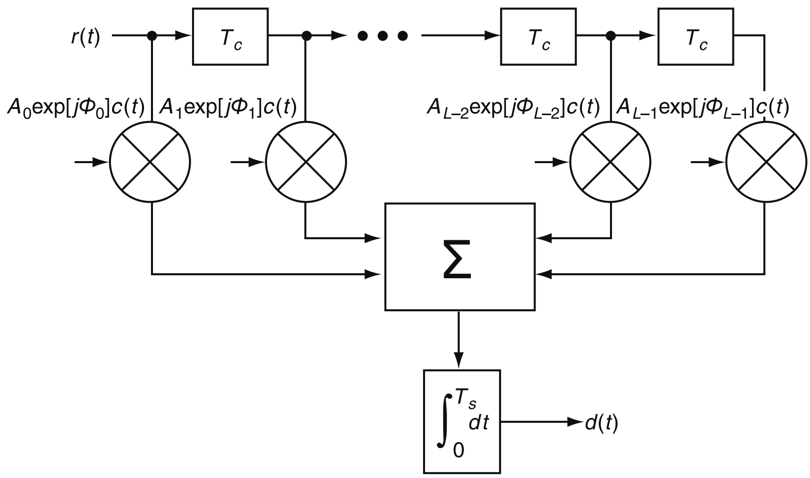


Figure 14.6 Conceptual diagram of a rake receiver

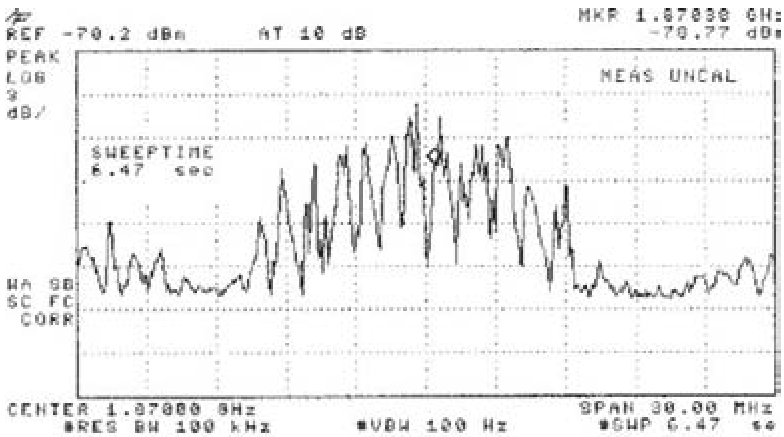


Figure 14.7 Typical received spectrum in a multipath environment[4]

14.4 DS-CDMA system performance

The capacity analysis of CDMA cellular networks is a complex process and, since capacity can be traded for quality, these systems exhibit a property known as soft capacity [11]. Test bed or field trial validation [4,5,21,22] can be used to provide a rigorous analysis of a subset of teleservices, but lacks flexibility if a system-wide perspective is required. Bit level simulations require many hours of CPU time due to the high signalling rates involved, but can yield useful results if

[4] Analyser parameters have been selected in order to emphasise the frequency-selective fading characteristics of the mobile channel

Table 14.1 Fading statistics of the synthesised urban channel

Ray	Excess delay (ns)	Mean amplitude (dB)	Fading	Fade rate (Hz)
1	0	0	Rician, $K = 7.0$ dB	100
2	60	–1.0	Rayleigh	31
3	125	–2.0	Rayleigh	96
4	250	–4.2	Rayleigh	35
5	500	–8.3	Rayleigh	94
6	1000	–16.7	Rayleigh	80

techniques such as importance sampling are employed, as well as directly exploiting propagation data available from field experiments. Mathematical modelling [23] offers by far the most computationally efficient approach for determining the relative performance and sensitivity analysis of DS-CDMA.

In order to develop a simulation or mathematical model of a cellular CDMA system, it is initially advantageous to make a number of simplifying assumptions. Here it has been assumed that ideal power control exists within the network in order to circumvent the near–far problem. This is particularly acute in DS systems, since all users share the same frequency band for either the up or downlinks of the system, and thus the power control scheme [24] must ensure that the cell site receivers are not captured[5] by a single user who may be in close proximity to the base station. Furthermore, in DS-CDMA it has been found that the power control loop must compensate for both the slow and fast fading present in the channel.

The capacity or bandwidth efficiency of a DS-CDMA cellular system can be evaluated using the approach of Pursley [27], where the total interference level (N_I) present at the input of a user's demodulator can be expressed in terms of mutual interference and thermal noise effects (N_0) for a total system capacity of K CDMA users. Mutual interference can be attributed to the $(K - 1)$ CDMA users also present within the same frequency band transmitting at a symbol energy of E_s, and further the power spectral density of this interference can be shown to be gaussian. The actual power level of this interference is scaled by the processing gain (G_p) of the spread spectrum system after despreading and prior to the input of the demodulator, thus giving:

$$N_I = \frac{(K-1)E_s}{G_p} + N_0 \tag{14.4}$$

For multiple access systems it can be shown[6] that $G_p = 3T_s/T_c$, where T_s is the coded symbol duration of the information signal.

The bandwidth efficiency (η) of the system can be expressed in terms of

[5] Recent work on multi-user detection [25,26] potentially offers an alternative solution
[6] Assuming asynchronous random spreading codes

symbol rate of the vocoder (R_s), number of symbol states employed in the RF modulation scheme (M) and the bandwidth of the spread spectrum transmission (W_{ss}):

$$\eta = \frac{R_s K(\log_2 M)}{W_{ss}} \tag{14.5}$$

For a large number of users, $K - 1 \cong K$, $W_{ss} = 1/T_c$ assuming that the transmitted spectrum has been filtered to half of the main lobe (see Figure 14.4*b*), and $R_s = 1/T_s$. Also, provided that a sufficient interleaving depth has been employed to ensure a discrete memoryless channel, the benefits of convolutional error protection coding can be evaluated as given below. Here the term $(E_s/N_I)_{REQ}$ is the ratio of coded symbol energy to interference energy required to support a given error rate to the vocoder when using a rate r_c convolutional code, with E_s/N_I representing the thermal noise level in the receiver. In the following results the thermal noise power has been assumed to remain at a constant level of 10 dB, and QPSK modulation ($M = 4$) with a ½ rate, constraint length 7 convolutional code has been employed for the air interface throughout:

$$\eta \approx 3r_c(\log_2 M)\left[\left(\frac{N_I}{E_s}\right)_{REQ} - \frac{N_0}{E_s}\right] \tag{14.6}$$

Eqn. 14.6 gives the bandwidth efficiency of a single isolated CDMA cell. In a cellular CDMA system users in the adjacent cells also contribute to the mutual interference level since they operate within the same CDMA frequency band, and thus reduce the capacity when compared with the single cell system. This is termed the frequency reuse efficiency (L_{fr}) of the system, and if the interference signals are modelled as zero mean gaussian noise sources uniformly distributed within an hexagonal cellular floor plan, then it can be shown that $L_{fr} = 0.67$ [11] assuming a fourth-order path loss exponent.

In order to reduce the impact of mutual interference on the overall capacity of DS-CDMA systems, several techniques can be employed. First, it is directly possible to exploit the fact that users are only engaged in active conversation for approximately 35 per cent of the time during a call, and thus users can be gated out of the system during the numerous periods of silence. This is known as the voice activity factor (G_{va}), and is equal to the reciprocal of the voice activity of the users. In addition, directional or sectorised base station antennas can be employed to reduce the amount of interference received by a cell site receiver. If an idealised three-sector base-station antenna system is assumed with a top-hat response, then the antenna gain (G_{sa}) is 3. Combining these effects gives:

$$\eta \approx 3r_c\,(\log_2 M)G_{va}G_{sa}L_{fr}\left[\left(\frac{N_I}{E_s}\right)_{REQ} - \frac{N_0}{E_s}\right] \tag{14.7}$$

The expression for the bandwidth efficiency of the system is dominated by the term $(N_I/E_s)_{REQ}$, and diversity reception techniques can be employed to reduce the symbol energy necessary to maintain a given bit error rate (BER), and thus enhance the capacity of the network. In particular, CDMA receivers make use of path diversity reception techniques as illustrated in Figure 14.5, and using the concept of the coded channel described by Simon [20], the bandwidth efficiency of a CDMA system can be calculated including the effects of diversity reception methods operating in both log-normal Rayleigh and log-normal Rician mobile radio channels [28].

Figure 14.8 gives the BER performance *versus* E_s/N_I for varying orders of internal diversity while operating in a log-normal Rayleigh channel with a standard deviation of 8 dB. In these results the diversity branches have been combined using maximal ratio combining (MRC), and it can be seen for a fixed BER that increasing the number of diversity branches results in a considerable reduction of the required E_s/N_I.

Figure 14.9 illustrates the BER performance *versus* E_s/N_I when equal gain combining (EGC) is employed for a system operating under the same channel conditions as given in Figure 14.8. Again, it can be seen that there are considerable benefits to be obtained from employing diversity reception techniques, and when comparing MRC against EGC there is an approximate power saving of 3 dB in favour of MRC for fifth-order path diversity at a BER of 10^{-3}.

Figure 14.10 gives the bandwidth efficiency *versus* diversity order for both MRC and EGC architectures operating in the log-normal Rayleigh channel for a BER of 10^{-3}, assuming no base station sectorisation (G_{sa} = 1). Again, the benefits of employing MRC combining against EGC can be clearly seen, and also that there are considerable benefits to be attained in terms of capacity enhancement when employing up to four diversity branches, whereas increasing

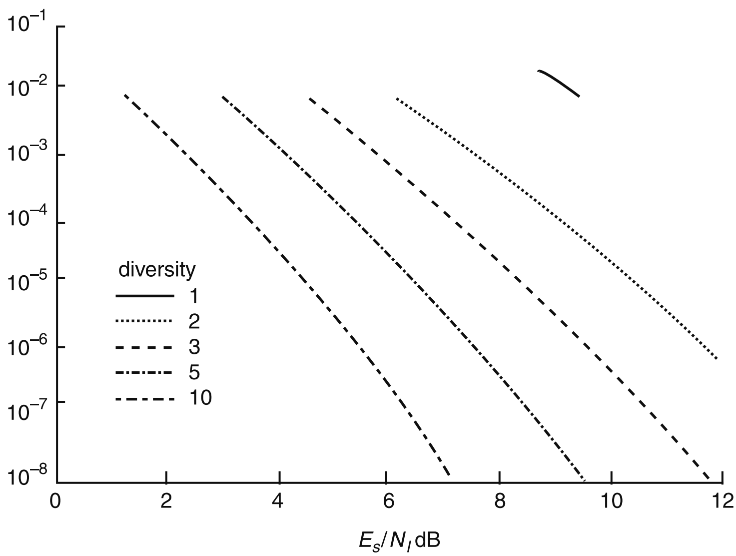


Figure 14.8 BER versus E_s/N_I for MRC operation in a log-normal Rayleigh channel

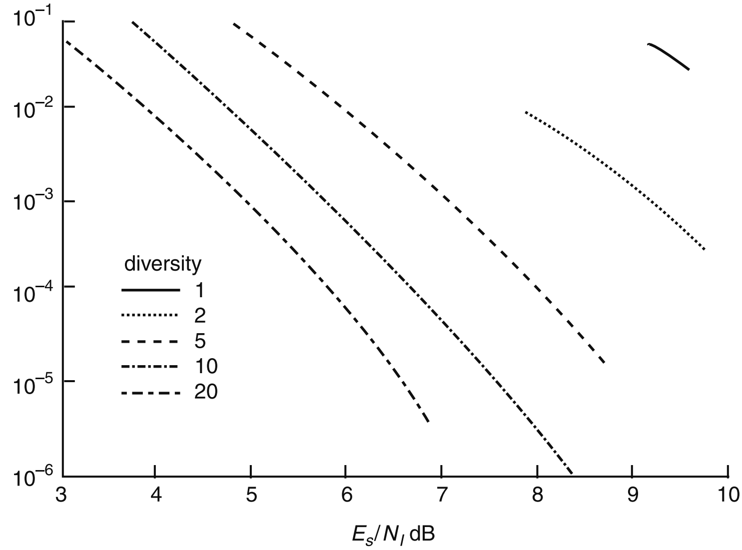


Figure 14.9 BER versus E_s/N_I for EGC operation in a log-normal Rayleigh channel

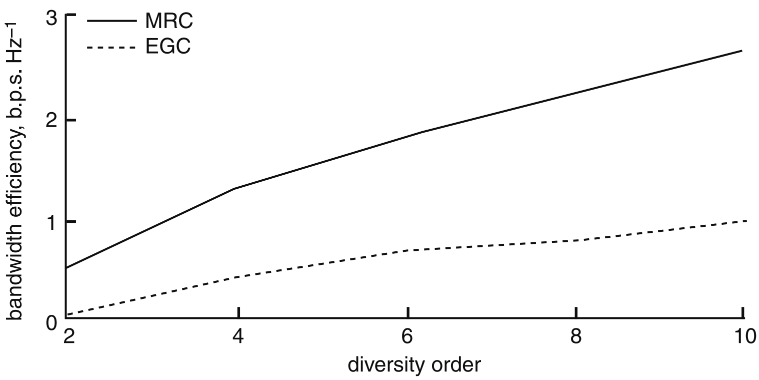


Figure 14.10 Bandwidth efficiency versus diversity order for MRC and EGC operating in a log-normal Rayleigh channel

the diversity order beyond this level tends to result in diminishing returns with respect to complexity and cost for any further increase in capacity.

Results giving the BER performance of a CDMA rake receiver with MRC processing for varying numbers of diversity branches operating in a log-normal Rician channel with a K factor of 3 dB and a standard deviation of 2.77 dB are given in Figure 14.11. When compared with the results given for the log-normal Rayleigh channel, there is a considerable reduction in the required E_s/N_I to maintain a given BER because of the deterministic nature of the Rician channel. Further, it can also be seen that the BER performance is less sensitive to the diversity order of the MRC architecture.

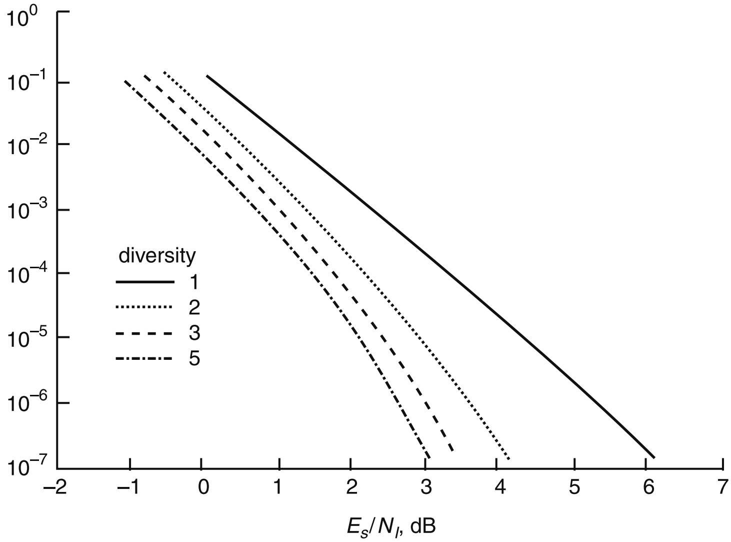


Figure 14.11 BER versus E_s/N_I for MRC operation in a log-normal Rician channel

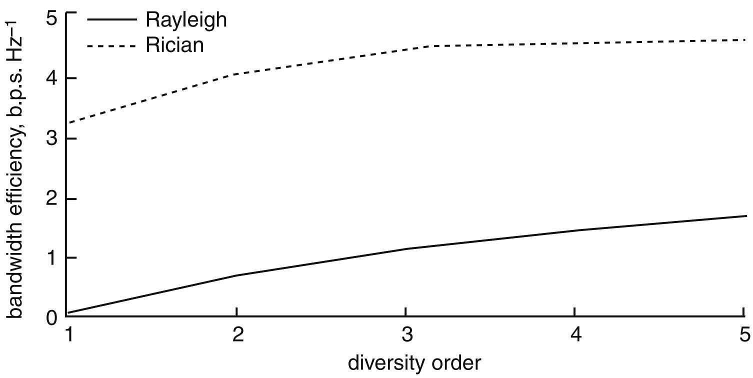


Figure 14.12 Bandwidth efficiency for a MRC rake architecture operating in both log-normal Rayleigh and Rician channels

Figure 14.12 illustrates the relative bandwidth efficiencies of MRC rake reception techniques while operating in both log-normal Rayleigh and Rician channels for a BER of 10^{-3}, again assuming no base station sectorisation. The operational characteristics obtained for these channels are extremely different, with the system performance being very sensitive to the diversity order for the Rayleigh case, and with considerably reduced sensitivity shown in the case of the Rician channel.

The sensitivity of DS-CDMA to the assumptions made regarding the propagation characteristics of the mobile fading channel has prompted several investigations [29,30] into the characteristics of this medium when viewed through a wideband rake CDMA receiver. Based upon the discussion so far, the key issues to be resolved by engineers considering such systems are as follows:

- can sufficient independent diversity branches be resolved from the channel?
- what are the bin (diversity branch) statistics of the mobile channel?

14.5 DS-CDMA channel measurements and models

The use of wideband correlation channel sounding techniques allows the extraction of the impulse response of the multipath channel. By employing continuous sampling techniques, the statistics directly related to the operation of CDMA rake receivers can be obtained. Before examining the fading statistics of a wideband urban channel, a discussion on multipath resolution based on the impulse response of a typical channel provides a useful insight into the problems of the CDMA system designer.

A typical impulse response of an urban cellular service area in the City of Bristol is illustrated in Figure 14.13. This response was taken using a sliding-correlator [31] sounder operating at 1.8 GHz with a path resolution of 50 ns. It can be seen that most of the energy arrives at the mobile receiver within a 1 μs excess delay window, although there is some low level activity out to 2 μs. Also shown in this Figure are the discrete time bins of a 1 μs resolution rake receiver. If such an architecture was employed in this situation, then only one path or branch of an internal diversity processing architecture could be supported if a 10 dB diversity power window is assumed. In order to support the degree of diversity processing as suggested in the previous section, it is necessary to increase

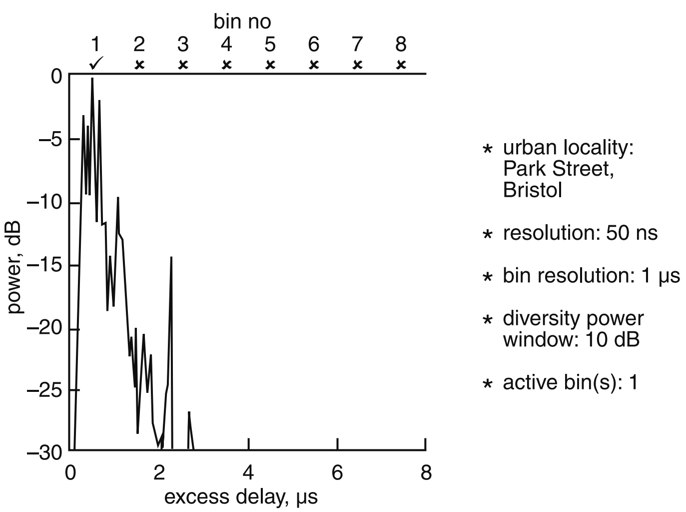


Figure 14.13 Impulse response of a typical urban mobile channel with 1 μs rake bin resolution

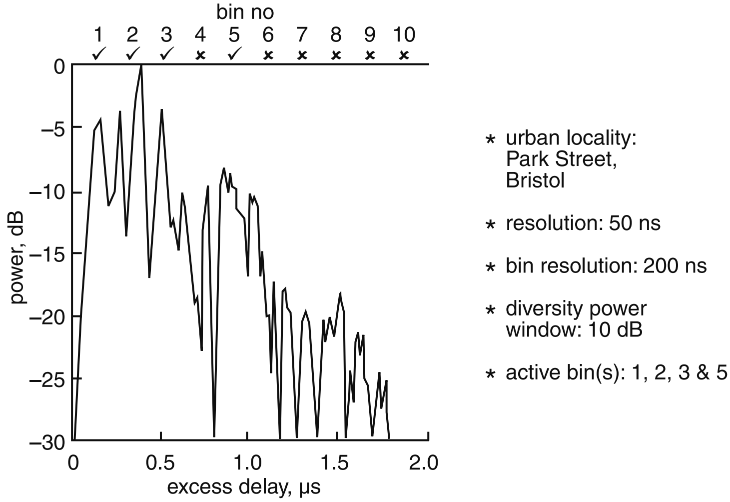


Figure 14.14 Impulse response of a typical urban mobile channel with 200 ns rake bin resolution

the bin resolution to 200 ns as illustrated in Figure 14.14. Here it can be seen that up to four diversity branches can be supported.

In order to investigate the multipath resolution and the bin statistics of a DS-CDMA system, modifications were made to a sliding-correlator wideband channel sounder operating within the 1.8 GHz PCN band, thus emulating a multifingered rake receiver. In the following sets of results, Figures 14.15 to 14.19, a 255 chip *m*-sequence was clocked at 1.25, 2.5, 5, 10 and 20 Mchip s^{-1} while the receiving system moved within an urban locality.

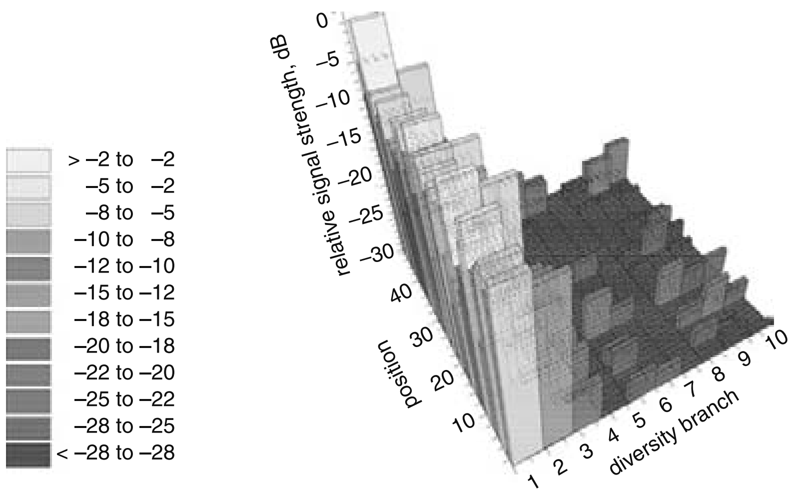


Figure 14.15 Urban DS-CDMA rake emulation at 1.25 Mchip s^{-1}

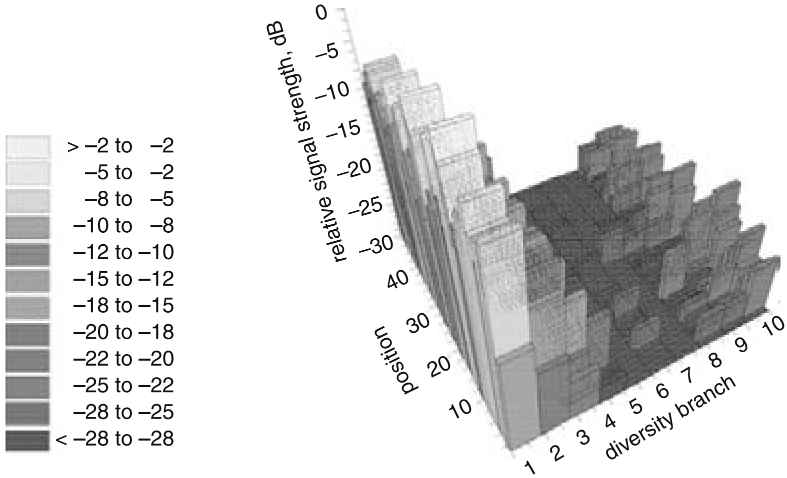


Figure 14.16 *Urban DS-CDMA rake emulation at 2.5 Mchip* s^{-1}

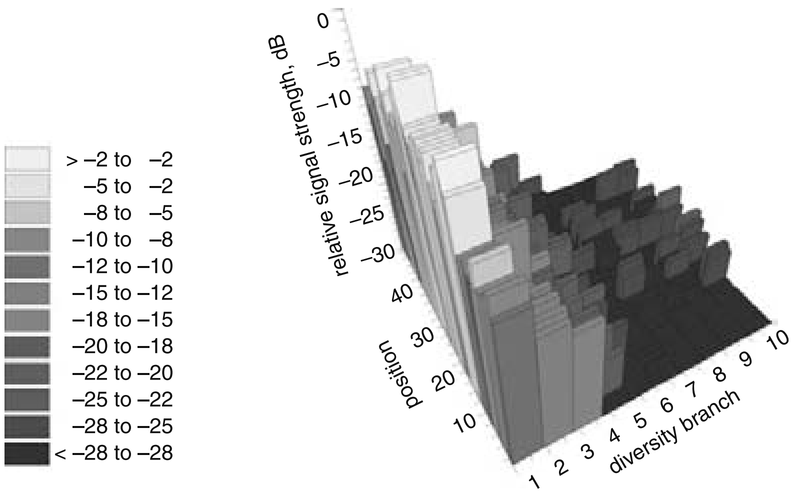


Figure 14.17 *Urban DS-CDMA rake emulation at 5 Mchip* s^{-1}

From the above results it can be clearly seen that as the chipping rate is increased the multipath activity spreads to more diversity branches or rake bins, thus making more independently fading samples available for diversity combining in the receiver. Also, increasing the chipping rate or spread spectrum bandwidth has a most notable affect on the signal strength variability of each branch, but in particular that of the first bin as illustrated in Table 14.2. Increasing the chipping rate reduces the signal variability, and this has a significant impact upon the E_s/N_I required to support a given error rate. In addition, higher chipping rates reduce the sensitivity of the system to power control errors, since the dynamics of the fast fading component have been significantly reduced.

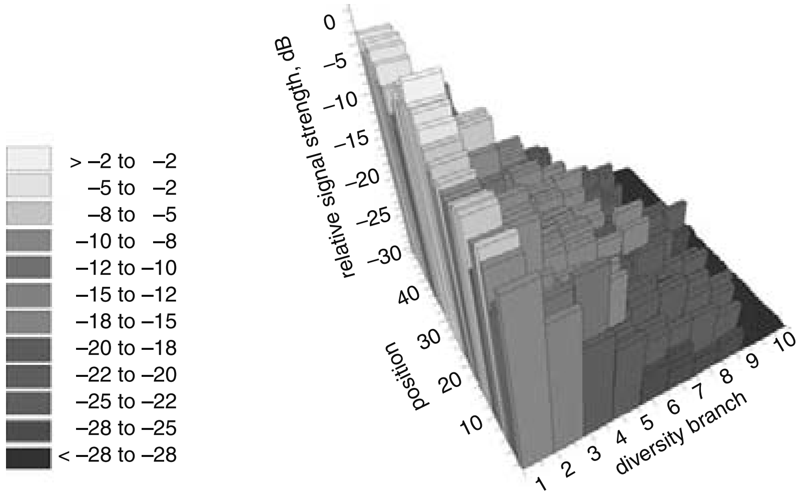


Figure 14.18 Urban DS-CDMA rake emulation at 10 Mchip s[−1]

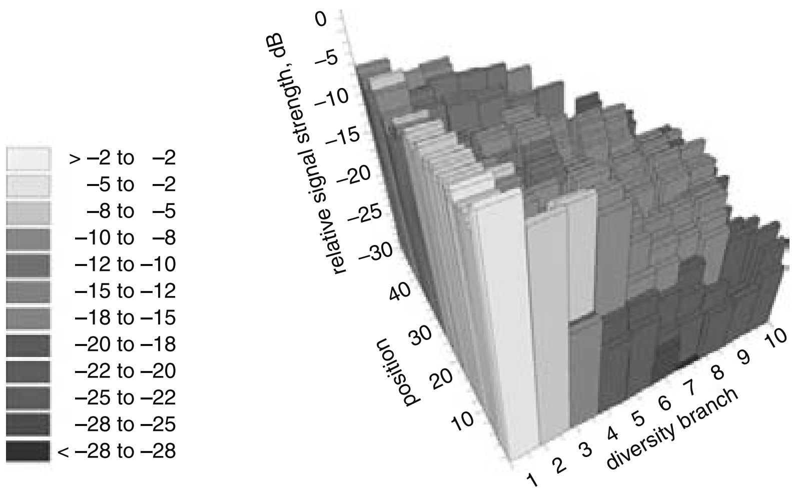


Figure 14.19 Urban DS-CDMA rake emulation at 20 Mchip s[−1]

Table 14.2 Rician K-factor of dominant bin

Chipping rate (Mchip s^{-1})	1.25	2.5	5	10	20
K factor (dB)	−6.04	0.78	3.22	3.99	6.68

The practical results presented here can be shown to have a good agreement with the theoretical work of Holtzman [33], as given in Figures 14.20 to 14.22. In Figure 14.20 the mean signal power of the measured data is shown as a function of the diversity bin number for the varying chipping rates considered during the field trial experiments. It can be seen that the mean signal power contained within each branch reduces with increasing diversity order, but this occurs with a diminishing rate as the chipping rate is increased.

Increasing the chipping rate of the system results in less energy being available on a per diversity bin basis for the same operational environment. This is shown in Figure 14.21 in terms of the available power as additional branches are

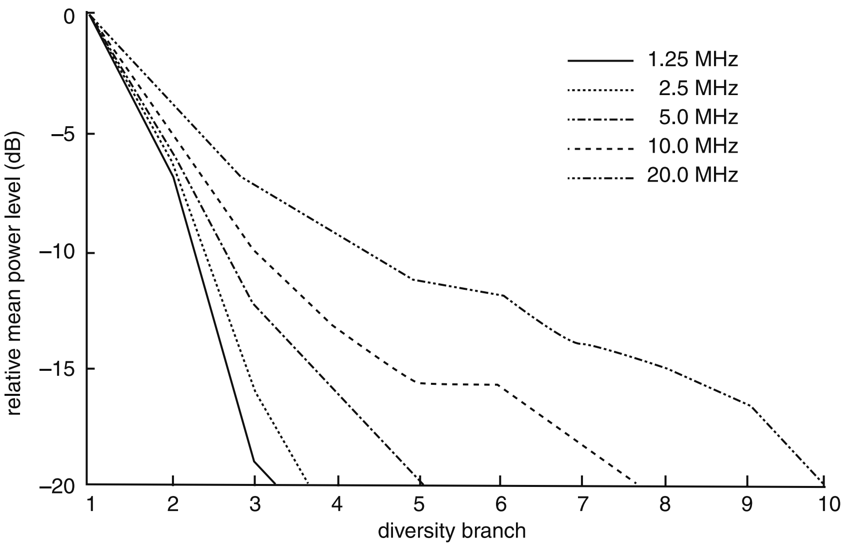


Figure 14.20 Mean signal power versus diversity branch

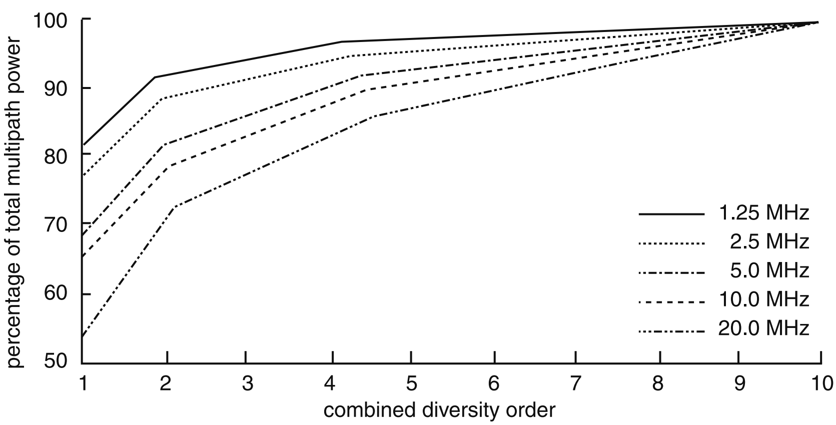


Figure 14.21 Available power versus combined diversity order

combined for the various chipping rates considered. From here it can be seen that low chipping rate systems (~1 Mchip s^{-1}) extract most of the available multipath power with a single branch architecture, whereas higher chip rate systems can make use of higher-order diversity.

The signal strength variability can also be considered in terms of the coefficient of variation [33] and the normalised standard deviation. The former is defined as the ratio of the standard deviation relative to the mean, whereas the latter is given by first normalising the received powers measured in dBs and then calculating the standard deviation relative to the mean. These results are illustrated in Figure 14.22, where it can be seen that as the chip duration is increased[7] the coefficient of variation of the first diversity branch tends to 1, and also the value of the normalised standard deviation approaches 5.57 dB. These values correspond to the Rayleigh case.

14.6 Support of handover in cellular networks

Cellular CDMA networks should not be considered as single isolated cell structures. This has already been discussed in terms of mutual interference arising from users in adjacent cells operating on the same CDMA frequency band. In addition, mobile users will frequently move from cell to cell, and thus it is necessary to handover service provision from one cell site to the next. Currently, all FDMA and TDMA cellular networks employ a technique called hard

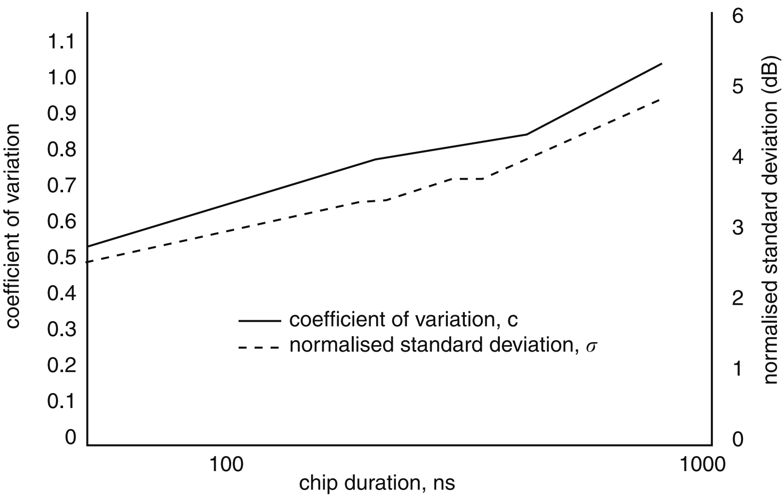


Figure 14.22 Coefficient of variation and normalised standard deviation versus chip duration

[7] Increasing the chip duration corresponds to a decrease in the chipping rate or bandwidth of the CDMA transmission

handover, where the air interface is physically switched between cell sites. Hysteresis is normally employed to prevent the switching ping-ponging between two cell sites at the radio coverage boundaries, however, service quality is often poor and calls are frequently dropped during handover.

If the path diversity process inherent in the CDMA technique is again considered as illustrated in Figure 14.5, the multipath components a and c could be considered to originate from one cell site, and b from another. Now by designing the network such that multiple base stations can transmit and receive information associated with the same mobile call within the handover regions of the system, the concept of soft handover is realised [18]. There are numerous benefits associated with the soft handover process when compared with that of hard handover; first it can be engineered as a make before break switch thus reducing the number of dropped calls. In addition, the diversity gain of the rake signal processing can be directly exploited in terms of a reduction of the E_s/N_I required at the boundaries of the service area, thus resulting in a net capacity increase when compared with hard handover CDMA as illustrated in Figure 14.23.

14.7 Conclusions

From the results presented in Section 14.4 it can be seen that the capacity of a CDMA network can be traded against the complexity of the diversity signal processing architecture employed within the receiver, and that the gain available from path diversity processing is extremely sensitive to the short-term propagation characteristics of the channel. Furthermore, the chipping rate of these

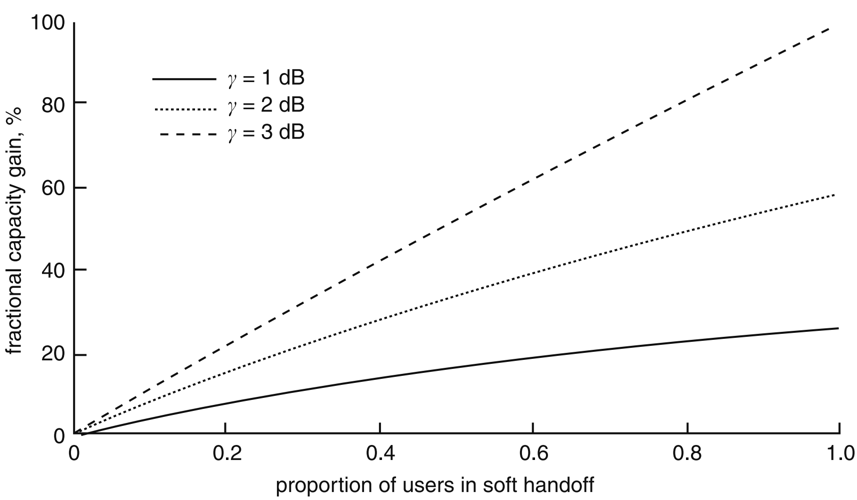


Figure 14.23 Capacity gain versus diversity gain (γ) due to soft handover [34]

systems must be sufficient to ensure that multipath energy can be extracted from the channel (see Figures 14.12 and 14.14) for use in the diversity process, with the RF bandwidth being directly related to this value.

It has also been shown that the channel statistics of the mobile radio channel vary when observed through a DS-CDMA receiver as the chipping rate is varied, and this has a significant impact on the overall system performance. For low chipping rates (≤1.25 Mchip s^{-1}) the measured channel displayed Rayleigh-like statistics with a Rician K factor of –6dB. As the chipping rate was increased to 20 Mchip s^{-1}, strong Rician statistics were observed. The terms narrowband and wideband CDMA are now associated with these systems, with chipping rates of approximately 1 Mchip s^{-1}, and rates in excess of 5 Mchip s^{-1}, defining the respective categories. Further, the benefits illustrated here substantiate the choice of chipping rates now employed in UTRA and similar systems. However, it should be noted that the discussion give in Sections 14.2 to 14.5 assumes that *m*-sequences are employed as the spreading codes, whereas in UTRA a hybrid code structure based on these sequences is employed [18].

A common figure of merit used to compare channel impulse responses and air interface performance is the RMS delay spread. This measure has not been discussed here in the context of CDMA systems, since it provides no useful information regarding system performance. This is because no phase information of multipath signals arriving within a particular time window is given, and also there is too much ambiguity in the form of the impulse response data. This is illustrated in Figure 14.24, where the RMS delay spread of both impulse response measurements is 200 ns. The response shown in Figure 14.24*a* contains two rays of equal amplitude arriving within a 400 ns time window, whereas Figure 14.24*b* again contains two rays, but with the second path arriving 700 ns after the first at a power level of –10 dB.

Although both impulse responses have the same RMS delay spread, when processed by a rake receiver they produce very different results. For example, if a chipping rate of 2 Mchip s^{-1} was employed, then Figure 14.24*a* would produce severe bin fading in the form of the two-ray model scenario. No bin fading would be observed by the receiver for the environment of Figure 14.24*b*.

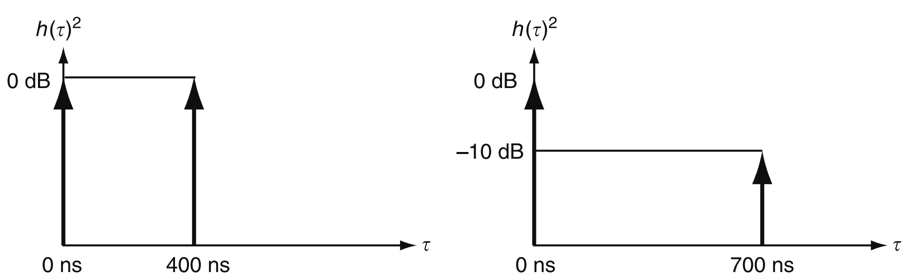


Figure 14.24 Illustration of ambiguity of RMS delay spread

14.8 Acknowledgments

The authors wish to express their sincere thanks to colleagues at the Centre for Communication Research (University of Bristol), in particular to Prof. Joe McGeehan, Dr. Simon Swales, Dr. Andy Nix and Dr. Joseph Cheung for their invaluable help and support of this work. We also greatly value the early contributions from Dr. Amnon Hammer on CDMA capacity analysis.

14.9 References

1 RICHARDSON, K.W.: 'UMTS overview', *Electron. Commun. Eng. J.*, 2000, **12**, **3**, pp.93–100
2 BULL, D.R., CANAGARAJAH, C.N. and NIX, N.R.: 'Insights into mobile multimedia communications' (Academic Press, 2000)
3 PAJUKOSKI, K. and SAVUSALO, J.: 'Wideband CDMA test system'. Proceedings of IEEE international conference on *Personal indoor and mobile radio communications*, PIMRC'97, Helsinki, Finland, 1–4 September 1997, pp.667–672
4 'CEC, RACE mobile telecommunications workshop''. DG XIII/B, RA129–93, Metz, France, 16–18 June 1993
5 BEACH, M.A.: 'A CDMA air interface for mobile access', 'Digital mobile radio towards future generation systems' (COST231 final report), ch. 7.2
6 http://www.3gpp.org
7 DIXON, R.C.: 'Spread spectrum systems' (John Wiley & Sons, Inc., 1984, 2nd edn.)
8 LIVNEH, N. *et al.*: 'Frequency hopping CDMA for future cellular radio'. IEEE 42nd VTC, Denver, May 1992, pp.400–404
9 PURLE, D.J. *et al.*: 'Investigation of propagation through a frequency hopped channel', *Electron. Lett.*, 1992, **29**, **9**, pp.785–6
10 http://www.bluetooth.org/
11 GILHOUSEN, K.S. *et al.*: 'On the capacity of a cellular CDMA system', *IEEE Trans. Veh. Technol.*, 1991, **40**, pp.303–312
12 SCHILLING, D.S.: 'Broadband CDMA overlay'. IEEE 43rd VTC, New Jersey, May 1993, pp.452–455
13 OMURA, J.K.: 'Spread spectrum radios for personal communication services'. IEEE International symposium on *Personal, indoor and mobile radio communications*, November 1992
14 PARSONS, J.D: 'The mobile radio propagation channel' (Wiley, 2000, 2nd edn.)
15 GOLOMB, S.W.: 'Shift register sequences' (Aegean Park Press, California, 1982, revised edn.)
16 Digital land mobile radio communications, final report. CEC COST207, CD-NA-12160-EN-C, Brussels, 1989
17 YANG, S.C.: CDMA RF system engineering (Artech House, 1998)
18 HOLMA, H, and TOSKALA, A.: 'WCDMA for UMTS', (Wiley, 2000)
19 LEE, W.C.Y.: 'Overview of cellular CDMA', *IEEE Trans. Vch. Technol.*, 1991, **40**, pp. 291–310

20 SIMON, M.K. *et al.*: ‘Spread spectrum communications’ (Computer Science Press, Rockville, 1985, vols. I–III)
21 LEE, J.S., & MILLER, L.E.: ‘CDMA systems engineering handbook’ (Artech House, 1998)
22 NIKULA, E., TOSKALA, A., DAHLMAN, E., GIRARD, L., and KLEIN, L.: ‘FRAMES multiple access for UMTS and IMT2000’, *IEEE Pers. Commun.*, 1998, pp.16–24
23 BEACH, M.A. *et al.*: ‘An evaluation of direct sequence CDMA for future mobile personal communication networks’. IEEE VTC, St. Louis, May 1991, pp.63–70
24 GILHOUSEN, K.S.: ‘Mobile power control for CDMA’, *IEEE Trans. Commun.*, 1992, pp.34–38
25 WIJAYASURIYA, S.S.H. *et al.*: ‘Rake Decorrelating Receiver for DS-CDMA mobile radio networks’, *Electron. Lett.*, 1992, **29**, pp.395–396
26 HONIG, M., and TSATSANIS, M.K.: ‘Adaptive techniques for multiuser CDMA receivers’, *IEEE Signal Process. Mag.*, 2000, **17**, (3), pp.49–61
27 PURSLEY, M.B.: ‘Performance evaluation for phased-coded spread spectrum multiple access communications – part I: system analysis’, *IEEE Trans. Commun.*, 1977, **COM-25**, pp.795–799
28 ALLPRESS, S.A.: ‘Optimising signalling rate and internal diversity order for the mobile cellular DS-CDMA systems’. PhD thesis, University of Bristol, Bristol, December 1993
29 ALLPRESS, S.A. and BEACH, M.A.: ‘Measurement & characterisation of the wideband DS-CDMA radio channel’. IEE ICAP93, Edinburgh, March 1993, pp.429–432
30 ALLPRESS, S.A. and BEACH, M.A.: ‘On the optimum DS-CDMA channel bandwidth for personal communication systems’. IEEE 43rd VTC, New Jersey, May 1993, pp.436–439
31 COX, D.C., and LECK, R.P.: ‘Correlation bandwidth and delay spread multipath propagation statistics for 910 MHz urban mobile radio channels’, *IEEE Trans. on Commun.*, 1985, **23**, pp.1275–1280
32 HOLTZMAN, J.M. and JALLOUL, L.M.A.: ‘Rayleigh fading effect reduction with wideband DS-CDMA signals’. Technical Report, Rutgers University, August 1991
33 MILLER, I. and FREUD, J.: ‘Probability and statistics for engineers’, (Prentice-Hall, 1985, 3rd edn. 1985)
34 CHEUNG, J.C.S. *et al.*: ‘On the impact of soft handoff on cellular direct sequence code division multiple access system’. CEC RACE PLATON/UOB/037/WPD2/07/ 93/Issue1, July 1993

Chapter 15

Basic physics of the ionosphere

H. Rishbeth

15.1 Introduction

15.1.1 What is the ionosphere?

The ionosphere is the part of the atmosphere in which free electrons are sufficiently numerous to influence the propagation of radiowaves. Very roughly, it extends in height from 70 to 600 km, although its upper limit is not well defined. In ionospheric parlance the term region (D, E, F) denotes parts of the atmosphere, the D/E boundary being conventionally at 90 km height and the E/F boundary at 150 km. The term layer refers to ionisation within a region, such as the F1 and F2 layers in the F region and the E, E2 and sporadic E layers in the E region (Figure 15.1).

The term plasma is used for an ionised gas; the ionosphere is a weak plasma because, even in the F region by day, no more than one part in 10^6 of the gas is ionised. The ionosphere is electrically neutral to a high degree of approximation, as positive and negative charged particles are always created and destroyed together.

15.1.2 Where does the ionosphere fit into the atmosphere?

Figure 1.2 of Chapter 1 shows the main regions of the neutral atmosphere: the troposphere (0–12 km), where the temperature decreases upwards; stratosphere (12–30 km), a turbulent region of fairly uniform temperature; mesosphere (30–80 km), heated by absorption of solar UV radiation by ozone; and the thermosphere above 80 km, strongly heated by solar extreme ultraviolet (EUV) and X-rays. In the thermosphere, the temperature increases rapidly upwards towards the limiting exospheric temperature, which on average is about 1000 K. In the exosphere, above about 600 km, individual atoms move freely in satellite-like orbits controlled by the earth's gravity. The solar EUV and X-rays dissociate

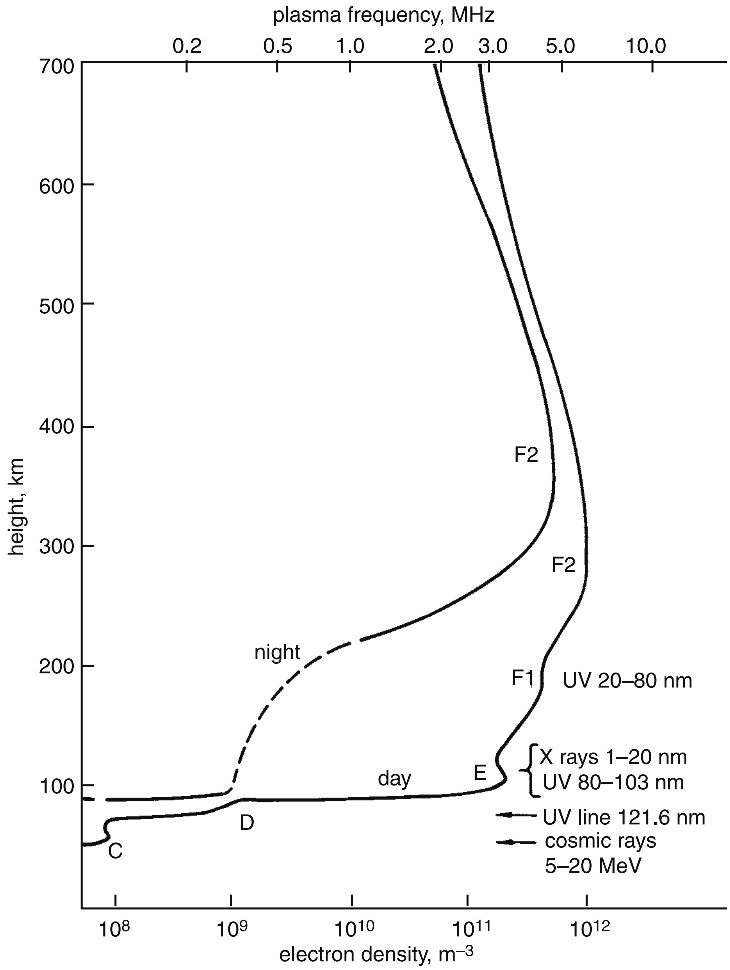


Figure 15.1 Typical mid-latitude N(h) electron density profiles for moderate solar activity, showing the radiations that produce the ionospheric layers (Contemp. Phys., 1973 ***16****, 230 (Taylor & Francis, London))*

and ionise the atmospheric gases in the upper mesosphere and thermosphere to create the ionosphere.

15.1.3 When was the ionosphere discovered?

Gauss in 1839 speculated on the existence of electrical phenomena in the upper atmosphere, and Balfour Stewart in 1883 postulated a conducting layer as the site of the currents that cause the small daily variations of the earth's magnetic field. This conducting layer, now called the ionosphere, can be said to have been discovered by Marconi when he sent radio signals across the Atlantic in 1901.

Lodge in 1902 correctly surmised that it contains free electrons and ions produced by solar ionising radiation. Research really began in 1924 when Appleton and Barnett and Breit and Tuve measured the height of the reflecting layers.

15.1.4 How do we study the ionosphere?

Ground-based techniques sound the ionosphere with radiowaves, which are returned from the ionosphere in three ways: by total reflection (ionosondes); by partial reflection or coherent scatter from small-scale structures in the ionosphere; or by incoherent scatter from the ion and electrons. These techniques are classed as radar; there is a historical link between the invention of radar and ionospheric research. Rockets and satellites enable *in situ* measurements and space/ground radio propagation experiments. See Table 15.1.

15.1.5 Why do we study the ionosphere?

The ionosphere is important because of its role in radiocommunications, and also scientifically because the electrons are easier to detect experimentally than the neutral air and thus act as tracers for studying the upper atmosphere. The electrons and ions form a weakly ionised gas or plasma, and the plasma physics associated with detailed ionospheric structure is relevant to radio propagation. Other current research topics are prediction and forecasting of ionospheric conditions on various time scales (space weather) and long-term studies of global change.

15.1.6 What are the most important characteristics of the ionosphere?

For radio propagation, the most important ionospheric characteristics are the critical frequencies (penetration frequencies at vertical incidence) for the ordinary magnetoionic mode (the extraordinary critical frequencies, which exist because of the birefringence caused by the Earth's magnetic field, are less important). For any layer the ordinary critical frequency is related to the peak electron density (electron concentration) N_m by:

$$f_0 = 80.6\ N_m \quad \text{or} \quad f_0 \approx 9\sqrt{N_m} \tag{15.1}$$

in which the MKS numerical factors are derived from physical constants. Other important characteristics are the heights h_m of the peaks of the layers, and the total or height-integrated electron content of the whole ionosphere.

15.2 The environment of the ionosphere

15.2.1 Ionising solar radiations

The sun emits radiations in the X-ray and extreme ultraviolet (EUV) parts of the electromagnetic spectrum, which comprise spectral lines and a background

Table 15.1 Experimental techniques for studying the ionosphere

Radio probing

- *Total reflection* at plasma frequency $f_p = \sqrt{(80.6\ N)}$ (typically HF/MF/LF in daytime F/E/D layers respectively). Used for routine sounding by ionosondes (which produce records known as ionograms). Cannot observe above F2 peak, except by topside sounders carried in satellites. Reflection is influenced by the geomagnetic field.
- *Coherent scatter* at frequencies >> f_p (HF, VHF, UHF). Scattering from irregularities of scale size $\lambda/2$ (λ = radar wavelength), which occur in the auroral oval and near the magnetic equator; also from meteor trails. Measures velocity by Doppler effect.
- *Incoherent scatter* at frequencies >> f_p (VHF, UHF). Thomson scattering from thermal random irregularities (ion-acoustic waves) of scale size $\lambda/2$. Very weak (depends on electron cross section, 10^{-28} m^2; the total target area is ~1 mm^2). Measures ion density, ion and electron temperatures, ion drift velocity, and indirectly some neutral atmosphere parameters.

Active remote sensing (using man-made signals traversing ionosphere)

- radio phase and polarisation (Faraday rotation)
- LF, VLF reflection, MF partial reflection in lower ionosphere
- radiowave absorption (signal strength)
- radar studies of meteor echoes (neutral winds, density)

Passive remote sensing from ground or space (using natural emissions); these techniques suffer from lack of height discrimination

- absorption of cosmic radiowaves; radio star refraction
- aurora – interferometry and imaging
- airglow – temperature and velocity of emitting gases

***In situ* measurements** (using instruments on rockets and satellites); good space and time resolution but no systematic coverage

- density, composition, temperature, velocity of ionised and neutral gas
- electric and magnetic fields
- tracking of objects and chemical trails released from spacecraft
- analysis of satellite orbits (giving air density)

Active experiments (e.g. r.f. heating, release of active chemicals)

of continuum radiation. The sun also emits energetic particles, but most of these have insufficient energy to ionise the atmosphere.

Sunspots themselves do not influence the ionosphere, but the strength of many solar emissions that do influence the ionosphere are linked to the sunspot number, a standard index of solar activity that rises and falls with the 11-year solar cycle; its values go back to 1749. Since 1948, the radio flux density of the sun at 10.7 cm wavelength has been routinely measured, as it also is found to be a good indicator of the ionising radiation (although the radio emission itself has no geophysical consequences).

Any particular radiation (wavelength λ) ionises gases for which the ionisation limit (79.6 nm for N_2, 91.1 nm for O and H, 102.7 nm for O_2 and 134 nm for NO) is longer than λ; the height to which the radiation penetrates depends on its absorption cross section σ (see section 15.3.1). The ionosphere is layered because σ is different for different radiations, the sources of the layers shown in Figure 15.1 being as follows (with some important lines indicated):

F1 layer (150–180 km): EUV 20–80 nm (He II 30.4 nm, He I 58.4 nm)
E layer (100–120 km): X-rays 1–20 nm, EUV 80–103 nm (Lyβ 102.6 nm)
D layer (70–90 km): Lyα 121.6 nm, X-rays 0.1–1 nm
C layer (50–70 km, an insignificant layer): MeV cosmic rays

No photon radiation has a large enough σ to be strongly absorbed at the height of the F2 layer; this layer is produced by the same band of radiation that produces the F1 layer. The electron density is greater in the F2 layer than in the F1 layer because of the smaller electron loss (Section 15.3.4). Charged particles can also ionise, at a height which depends on their energy.

15.2.2 Temperature at ionospheric heights

The ionospheric layers lie within the thermosphere, where the temperature T varies with height h. At the base, T is governed by the mesopause temperature, usually 150–180 K, but it increases rapidly upwards with a gradient dT/dh determined by the solar heat input. Higher up, the thermal conductivity of the air becomes large and the temperature gradient flattens out. Above about 300 km, the temperature tends to a limit known as the exospheric temperature T_∞ which varies considerably with local time (LT), latitude, season and the 11-year sunspot cycle. The temperature variations are mainly due to solar heating, which causes the atmosphere to expand on the daylit side of the earth, creating a high-pressure region sometimes called the daytime bulge. Typically at mid-latitudes T_∞ is 800 K at midday and 600 K at midnight at sunspot minimum, and 1400 K and 1100 K, respectively, at sunspot maximum.

From observations of satellite orbits, which yield extensive data on air density, it is deduced that T_∞ is highest around 1400 local time (LT) and lowest around 0300 LT; the maxima and minima are in low latitudes near equinox but migrate seasonally, as shown for example in Figure 15.2 which is a simplified map of T_∞ at the June solstice. Not shown in the Figure is the localised heating in the auroral ovals (Sections 15.2.5, 15.5.7), which is due to the energy deposited by energetic particles and electric currents from the magnetosphere.

15.2.3 Composition of the upper atmosphere

The formation of ionospheric layers depends on the density and composition of the ambient neutral air. From the ground up to 100 km, the major gases N_2 and O_2 (mean molar masses M of 28 and 32) are well mixed by winds and turbulence with a concentration ratio of about 4:1, giving a mean molar mass of about 29. The distributions of minor constituents H_2O, CO_2, NO and O_3 are controlled by

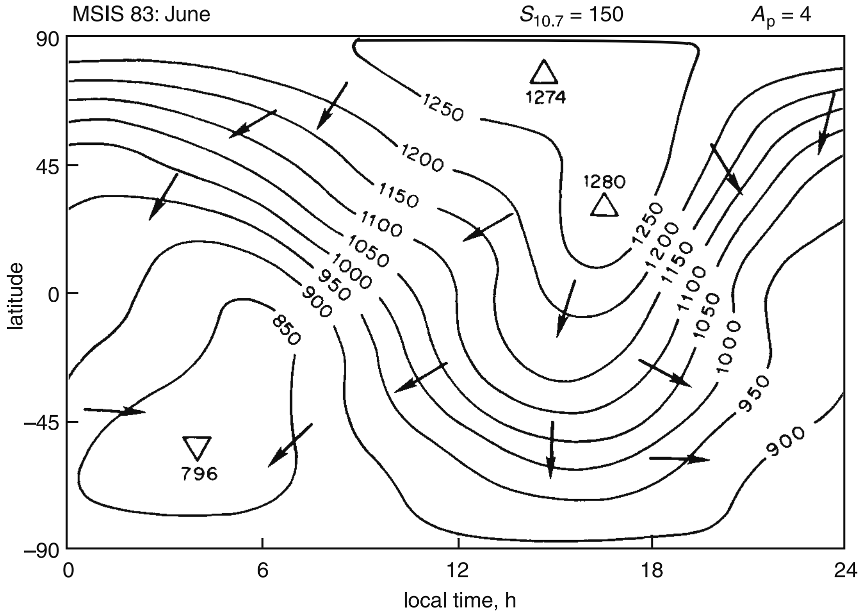


*Figure 15.2 Map of exospheric temperature at June solstice, for high solar activity and quiet geomagnetic conditions, according to the MSIS83 thermospheric model by A.E.Hedin, J. Geophys. Res., 1993, **88**, 10170; the arrows show approximate directions of the thermospheric winds (Section 15.4.6) (Rutherford Appleton Laboratory)*

chemical and transport processes, and do not follow the same trend as major gases. Above 100 km oxygen (but not nitrogen) is partly dissociated by the action of solar UV and, because there is little or no turbulence at these heights, the gases are diffusively separated under gravity so that each gas has its own scale height. Thus, atomic oxygen O (molar mass $M = 16$) progressively becomes more abundant than molecular N_2 and O_2. Higher still, in the exosphere, hydrogen and helium are dominant, but these gases are not important at ionospheric heights.

The composition varies with time and season. It is affected by the large-scale circulation in the thermosphere, which causes the atomic-to-molecular ratio above 200 km to be greater in winter than in summer (Section 15.4.7), and also by more localised factors such as heating in the auroral ovals. The composition changes profoundly affect the F2 layer because of the chemistry discussed in Section 15.3.5.

15.2.4 The scale height

The scale height H is the vertical distance in which the air pressure p, or the partial pressure of an individual gas, decreases by the exponential factor $e = 2.718\ldots$, thus:

$$(1/H) = -d(\ln p)/dh \tag{15.2}$$

The scale height may be deduced from the hydrostatic or barometric Eqn. 15.3 in which the vertical pressure gradient is balanced against gravity:

$$-dp/dh = \rho g = nmg \tag{15.3}$$

where h = height, ρ = density, n = concentration, g = acceleration due to gravity, m = particle mass (in kg, as compared to M which is the molar mass in atomic mass units). Using the perfect gas law in either of two forms:

$$p = nkT = \rho RT/M \tag{15.4}$$

where k = Boltzmann's constant, R = gas constant and T = temperature, to eliminate the density ρ, the scale height is found to be (again in either of two forms):

$$H = kT/mg = RT/Mg \tag{15.5}$$

H may be defined either for a single gas or for the air as a whole (in which case M is averaged over all the gases present). Integrating Eqn. 15.2 with respect to height gives the variation of pressure above any chosen base level h_0, at which $p = p_0$, *viz*.:

$$p = p_0\, e^{-z} \sim p_0 \exp\,((h_0 - h)/H) \tag{15.6}$$

where

$$z = \int (dh/H) \sim (h_0 - h)/H \tag{15.7}$$

is known as the reduced height, and is measured in units of H above the base level h_0. In Eqns. 15.6 and 15.7 the simplified expressions that follow the symbol ~ apply if H is constant with height – which is not really the case in the thermosphere, because T increases upwards and M decreases upwards, while g decreases upwards by about three per cent per 100 km. Whether or not H varies with height, it is easily shown that the total number of particles in a vertical column above any height h_0 is $Hn(h_0)$, so H may be regarded as the thickness of the atmosphere.

15.2.5 The geomagnetic field

The earth's magnetic field plays an important role in the ionosphere, because it largely governs the motions of ionospheric ions and electrons. It may be approximated by a dipole inclined at 12° to the earth's axis, slightly offset from the earth's centre in the direction of the Pacific Ocean. In low and middle latitudes the geomagnetic field lines are closed, but at higher latitudes they are linked to the magnetosphere, the tenuous region around the earth that is permeated by the geomagnetic field. The small daily variations of the geomagnetic field (of the order of 0.1 per cent) are produced by currents flowing in the ionosphere, generated by dynamo action of winds blowing across the geomagnetic field lines (Section 15.4.3).

The high latitude ionosphere is profoundly affected by charged particles and electric currents, originating in the magnetosphere, which cause the aurora and magnetic storms. The particles and currents mainly enter the atmosphere in the auroral ovals, rings 2000–3000 km in diameter that surround the magnetic poles (Section 15.5.7). In turn the magnetosphere is influenced by the stream of charged particles emitted by the sun (the solar wind) and the interplanetary magnetic field.

Magnetic storms tend to follow solar disturbances and the resulting strengthening of the solar wind (Section 15.6.2). During storms, the electric currents in the high latitude ionosphere may locally perturb the geomagnetic field by up to five per cent. The degree of geomagnetic disturbance is characterised by planetary geomagnetic indices (A_p and K_p), which are computed for every three hours of universal time from records made at several magnetic observatories.

15.3 Formation and photochemistry of the ionised layers

15.3.1 The Chapman formula for production of ionisation

The simple Chapman theory of the production of an ionised layer starts by assuming (*a*) monochromatic solar radiation acting on (*b*) a plane atmosphere containing (*c*) a single ionisable gas with (*d*) a constant scale height. All four simplifications (*a–d*) can be removed, but the equations then become more complicated.

Figure 15.3 shows the principles of the theory. Ionising radiation, initially of intensity I_∞, strikes the top of the atmosphere at zenith angle (obliquity) χ, and its intensity $I(h)$ decreases as it travels downwards, according to the equation:

$$dI/I = -\sigma n \, dh \sec \chi = -d\tau \tag{15.8}$$

where σ is the absorption cross section and τ is optical depth. By integrating Eqn. 15.8 along the path of the radiation, from the top of the atmosphere down to height h, and using the property that $\int n \, dh = n(h).H$ (as mentioned in Section 15.2.2), it is found that:

$$I(h) = I_\infty \exp(-n\sigma H \sec \chi) = I_\infty \exp(-\tau(h)) \tag{15.9}$$

From Eqn. 15.9 it may be seen that, for oblique sun, attenuation begins at a higher height (as shown by dots in Figure 15.3) than for overhead sun. The production rate q is found by multiplying the intensity of the radiation (dashed curve) by the gas concentration n (continuous curve) and by the cross section σ and ionising efficiency η. Thus:

$$q(h) = I(h)\, \eta\sigma n(h) = I_\infty\, \eta\sigma n(h) \exp(-\tau(h)) \tag{15.10}$$

On differentiating Eqn. 15.10, it is found that the peak value of q occurs at the height where $\tau = 1$ (unit optical depth). The height of peak production is given by:

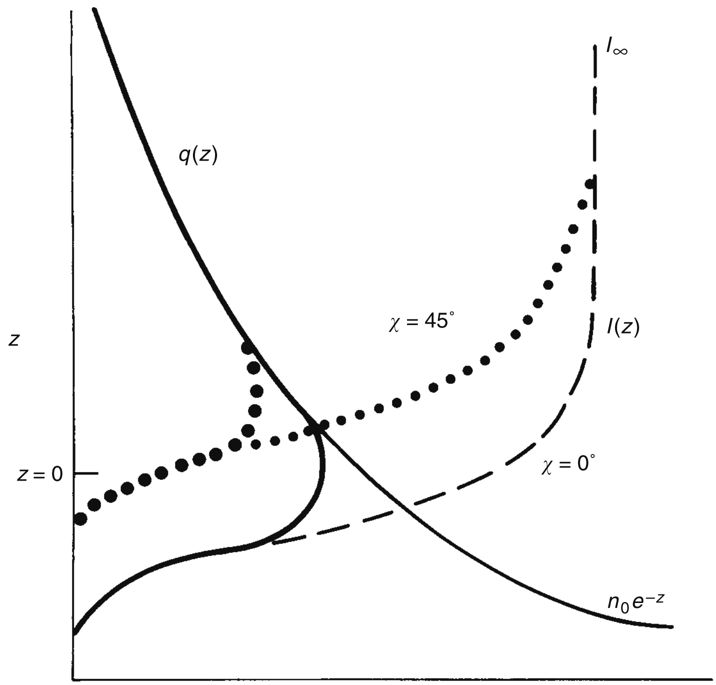


Figure 15.3 *Theoretical curves of ion production for overhead sun (full lines) and oblique sun (dotted lines), illustrating Eqns. 15.9 and 15.10 (Rutherford Appleton Laboratory)*

$$z_m = \ln \sec \chi \tag{15.11}$$

and the peak value of q is:

$$q_m = q_0 \cos \chi \text{ where } q_0 = (\eta I_\infty / eH) \tag{15.12}$$

Using the reduced height z as defined in Eqn. 15.7, q is given by the Chapman formula:

$$q(z, \chi) = q_0 \exp\,(1 - z - e^{-z} \sec \chi) \tag{15.13}$$

15.3.2 Chapman layers

For a simple Chapman layer in which electrons are produced according to Eqn. 15.13 and recombine at a rate proportional to N^2 (Section 15.3.5), the critical frequency depends on solar zenith angle χ and the mean sunspot number R according to the formula:

$$f_0 = A(1 + aR)\,(\cos \chi)^n \tag{15.14}$$

where $n = 0.25$ and A and a are constants. Data from ionosondes shows that the daytime E and F1 layers follow this formula quite well when χ varies with time of day, season, and latitude, although the index n actually varies between about 0.2 and 0.35. The incident solar flux I_∞ varies with the solar cycle and so f_0 varies regularly with the sunspot number. Figure 15.4 shows the regular seasonal and solar cycle variations of the E and F1 layers at Slough and Port Stanley, together with solar and interplanetary parameters. The normal D layer (although not possessing a measurable critical frequency) also varies quite regularly with cos χ. The F2 layer does not: careful examination of the seasonal ripple of Slough f_0 F2 shows that it is out of phase with those of f_0F1 and f_0E, which do conform to the variation of cos χ. At Port Stanley, however, the F2 layer has a predominantly semiannual variation, with maxima at the equinoxes.

15.3.3 The continuity equation

The electron density in the ionosphere is subject to the basic physical principles of conservation of mass, momentum and energy which, respectively, are expressed mathematically by the continuity equation for the electron density, the equation of motion for the electrons and ions, and equations expressing the heat balance (which is not discussed here). These equations may be formulated for any charged or neutral constituent of the atmosphere. Equations of motion are considered in Section 15.4.1. The present section deals with the continuity equation for the electrons and ions (the ionised plasma), which is:

$$\partial N/\partial t = q - L(N) - \mathrm{div}\,(N\mathbf{V}) \tag{15.15}$$

The loss term $L(N)$ can generally be simplified as follows:

$$\begin{gathered}\text{E \& F1 layers: } L(N) = \alpha N^2 \\ \text{F2 layer: } L(N) = \beta N\end{gathered} \tag{15.16}$$

where α, β are the square-law and linear loss coefficients (Section 15.3.5). The transport term, in which $\mathbf{V}$ is the plasma drift velocity, the drift velocity of the ionisation, is often simplified by omitting horizontal motion, which usually contributes little to div ($N\mathbf{V}$), and including only the vertical drift velocity, W, in which case:

$$\mathrm{div}\,(N\mathbf{V}) = \partial(NW)/\partial h \tag{15.17}$$

15.3.4 Ion chemistry

Reactions in the ionosphere have to conserve energy and momentum, besides satisfying the principles of quantum theory. One consequence is that atomic ions cannot recombine easily with electrons, unless a third body takes part in the reaction and absorbs the excess energy. This requires a three-body collision, which is only likely to occur below 100 km where the air density is sufficiently high. Molecular ions, however, recombine easily at any height, since they can dissociate to give two products as in Eqns. 15.24–15.26. Throughout the

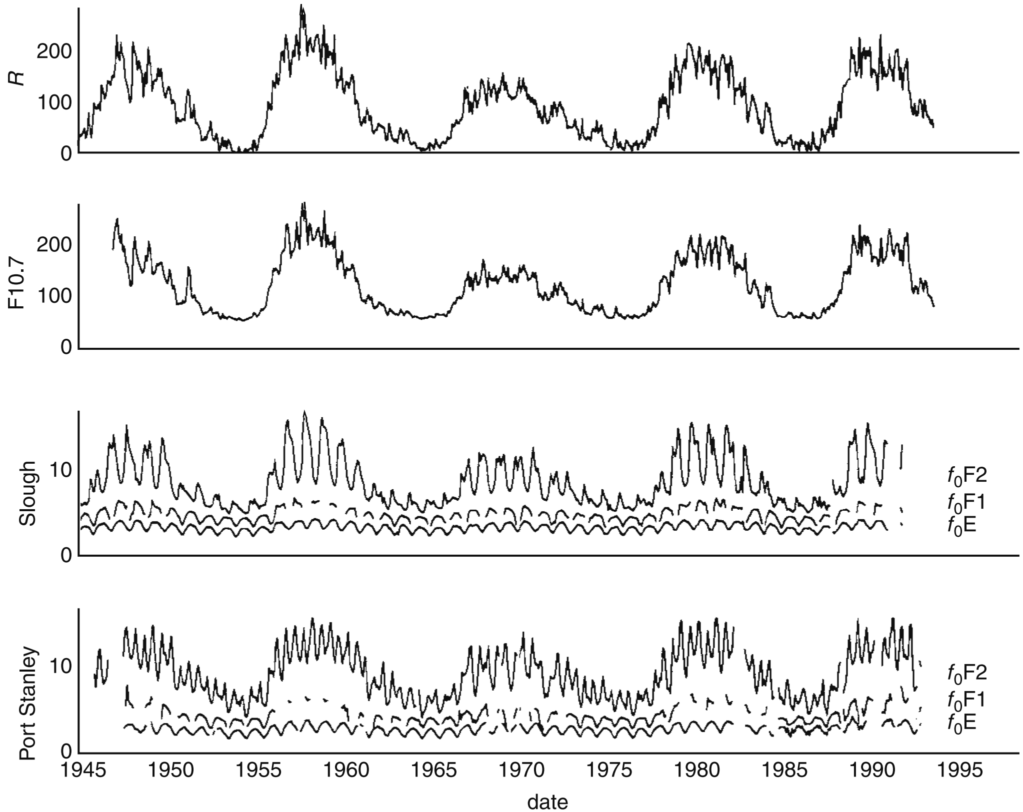


*Figure 15.4 Half a century of solar terrestrial data: monthly mean or median values of Zürich sunspot number R, solar 10.7 cm flux density, magnitude of the interplanetary magnetic field **B** near the earth, and noon critical frequencies of F2, F1 and E layers at Slough (latitude 52N) and Port Stanley (latitude 52S) (Rutherford Appleton Laboratory)*

ionosphere, reactions may leave atoms and molecules in excited states, which subsequently decay by emitting radiation. This is the origin of the upper atmosphere luminescence, the airglow.

In the D region, the atmosphere is relatively dense, so both two-body and three-body collisions occur and complicated chemical reactions take place. Electrons may become attached to atoms or molecules to form negative ions. Negative ions are more abundant than electrons at all heights below 90 km at night but, being destroyed by visible and ultraviolet sunlight, they are only abundant below 70 km by day. Many ions, both positive and negative, become attached to water molecules to form complex clusters.

In the E region, the atmosphere is more rarefied and only two-body collisions occur, so atomic ions cannot recombine easily with electrons, but are readily converted to molecular ions by reactions such as Eqns 15.21–15.23. The overall result is that the positive ions are mostly molecular, NO^- and O_2^+, and the loss coefficient α is of the square-law type in Eqn. 15.16. Some metal atoms (Fe^+, Mg^+, Ca^+), traces of which are deposited in the atmosphere by meteors, become ionised to produce long-lived metallic ions that are the main ionic component of mid-latitude sporadic E layers (Section 15.5.3). At greater heights, in the F region, the ions are mainly NO^+ and O_2^+ (F1 layer) and O^+ (F2 layer), as described in the following simplified summary of the chemistry.

15.3.5 The photochemical scheme of the ionosphere

Up to the F2 peak, the distribution of ions and electrons is essentially determined by photochemical processes of production and loss. As described in Section 15.2.1, the major gases in the thermosphere are ionised by EUV photons ($h\nu$), giving O^+, N_2^+ and some O_2^+ ions, plus electrons. For example:

$$O + h\nu \rightarrow O^+ + e^- \tag{15.18}$$

$$N_2 + h\nu \rightarrow N_2^+ + e^- \tag{15.19}$$

$$O_2 + h\nu \rightarrow O_2^+ + e^- \tag{15.20}$$

Ions collide with neutral molecules and undergo charge transfer or charge exchange reactions which produce molecular ions, for example:

$$O^+ + O_2 \rightarrow O_2^+ + O \tag{15.21}$$

$$O^+ + N_2 \rightarrow NO^+ + N \tag{15.22}$$

$$N_2^+ + O \rightarrow NO^+ + N \tag{15.23}$$

Molecular ions recombine with electrons and dissociate thus:

$$NO^+ + e^- \rightarrow N + O \tag{15.24}$$

$$O_2^+ + e^- \rightarrow O + O \tag{15.25}$$

$$N_2^+ + e^- \rightarrow N + N \tag{15.26}$$

This photochemical scheme has several consequences; the experimental evidence for (a)–(d) comes mainly from observations made by rocket-borne and satellite-borne instruments, such as mass spectrometers, and for (e)–(h) from the electron distributions measured by ionosondes:

(a) Despite the abundance of neutral N_2, very few N_2^+ ions are present, as they are rapidly destroyed *via* reaction 15.23 as well as by reaction 15.24.
(b) NO^+ ions are abundant because they are rapidly formed via reactions 15.22 and 15.23, even though neutral NO is hardly present at all.
(c) Below about 200 km molecular ions dominate, because the O^+ ions are rapidly converted to O_2^+ and NO^+ *via* reactions 15.21 and 15.22 by the relatively abundant O_2 and N_2.
(d) Above about 200 km, reactions 15.21 and 15.22 proceed very slowly because of the scarcity of N_2 and O_2 at these heights. Once formed, however, the NO^+ and O_2^+ ions rapidly recombine with electrons *via* reactions 15.24 and 15.25. As a result, O^+ is the dominant ion.
(e) In the E and F1 layers where NO^+ and O_2^+ ions dominate, the loss term in eqn. 15.15 takes the form aN^2, as in Eqn. 15.16.
(f) In the F2 layer, the loss term takes the form βN as in Eqn. 15.16, where

$$\beta = k'n[O_2] + k''n[N_2] \tag{15.27}$$

where k', k'' are the rate coefficients of reactions 15.21 and 15.22.
(g) The F1 layer is the transition between (e) and (f). It is a distinct feature with its own critical frequency at certain times (mainly summer and at sunspot minimum) but not at other times (especially winter). Its appearance is found to depend on the relative values of q, a and β.
(h) Above the F1 layer, the electron density increases upwards, because the loss coefficient β (which depends on the N_2 and O_2 concentrations) decreases upwards faster than the production rate q (which depends on the O concentration). The increase ends at the F2 peak (Section 15.4.8).

This photochemical scheme accounts for the general form of the $N(h)$ profiles as shown in Figure 15.5, up to the F2 peak which is discussed in Section 15.4.7, together with the sketches on the right of the Figure.

15.4 Dynamics of the ionosphere

15.4.1 Equation of motion of the charged particles

To understand the transport processes which determine the transport term in the continuity equation 15.15, one needs to understand how the ions and electrons move under the influence of the forces acting on them, namely gravity **g**, electrostatic field **E**, the geomagnetic field **B**, collisions with the neutral air (which moves with the wind velocity **U**) and gradients in their own partial pressures p_i and p_e (thus regarding them as an ion-electron gas or plasma). The

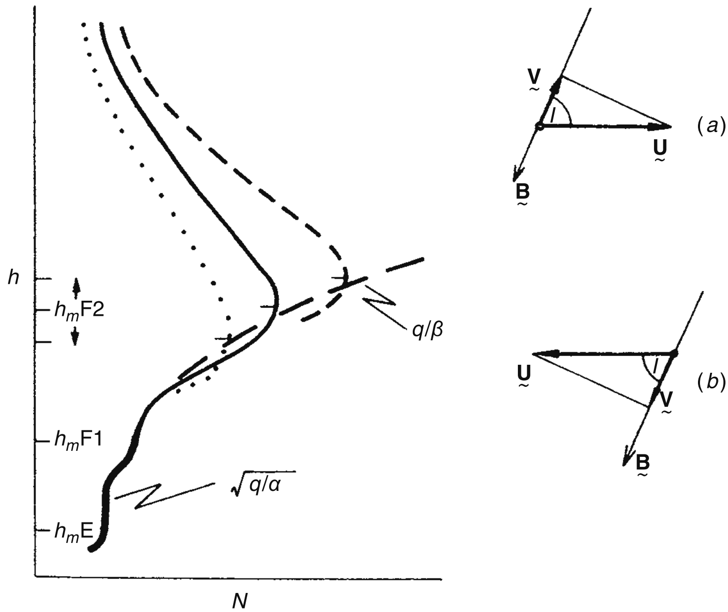


Figure 15.5 Idealised electron density profiles in the E and F1 layers, with square-law loss coefficient; and in the F2 layer, with linear loss coefficient. The sketches on the right show the plasma drift **V** *produced by a horizontal wind* **U** *in the neutral air, and* **B** *represents the geomagnetic field (dip angle I). Upward field-aligned drift is produced by a wind blowing towards the magnetic equator, as in sketch a; downward drift is produced by a wind towards the magnetic pole, as in sketch b (Contemp. Phys.* **14**, *244, 1973 (Taylor & Francis, London))*

simplified equations of motion for singly charged positive ions and electrons are:

$$m_i\, d\mathbf{V}_i/dt = m_i\mathbf{g} + e\mathbf{E} + e\mathbf{V}_i \times \mathbf{B} + m_i\nu_i(\mathbf{U} - \mathbf{V}_i) - \nabla p_i/N - 0 \qquad (15.28)$$

$$m_e\, d\mathbf{V}_e/dt = m_e\mathbf{g} - e\mathbf{E} - e\mathbf{V}_e \times \mathbf{B} + m_e\nu_e(\mathbf{U} - \mathbf{V}_e) - \nabla p_e/N - 0 \qquad (15.29)$$

The accelerations $d\mathbf{V}_i/dt$, $d\mathbf{V}_e/dt$ are negligible for the large-scale motions considered here, so the sum of the forces acting on the particles can be set equal to zero. The symbol e denotes the elementary positive charge (so that $-e$ is the charge on an electron), and ν_i, ν_e are the collision frequencies of ions and electrons with neutral particles.

15.4.2 Motion of charged particles due to winds and electric fields

Only a few special cases of Eqns. 15.28 and 15.29 need be discussed here. They apply to the low energy (essentially thermal) particles that make up the ionised layers, not the energetic particles in the magnetosphere and auroral regions, the motions of which are more complicated. They assume a single kind of positive ion and ignore negative ions (but can easily be generalized to remove these restrictions). Deleting the acceleration, gravitational and pressure gradient terms, Eqns. 15.28 and 15.29 can be solved to give the following conclusions.

For motion parallel to **B**: if the neutral air wind has a component parallel to **B**, it moves both ions and electrons freely in that direction. If the electric field has a component parallel to **B**, it drives electrons and ions in opposite directions, i.e. there is a high electrical conductivity.

For motion perpendicular to **B**, the direction of motion depends on the ratio of the collision frequency v (with neutral particles) to the magnetic gyrofrequency $\omega = \mathbf{B}e/m$. Since v decreases exponentially upwards, while ω is almost constant, the ratio v/ω falls off rapidly with height; it is different for ions and electrons. The important cases are:

(i) $v \gg \omega$ (up to 100 km for ions, 60 km for electrons). A wind drives the ions and electrons at its own speed, U, and the magnetic field has no effect. An electric field moves ions parallel and electrons antiparallel to itself, but only slowly because of the high collision frequency, and the conductivity is low.

(ii) $v \sim \omega$ (at about 125 km for ions, 75 km for electrons). A wind or an electric field drive ions and electrons in directions inclined to the applied force. The angle is given by arc tan (v/ω), and is thus 45° at the level where $v = \omega$.

(iii) $v \ll \omega$ (above 150 km for ions, 90 km for electrons). The particles move at 90° to the applied force. An electric field moves both ions and electrons at speed E/B in the same direction; this is the electromagnetic drift velocity (Hall drift), written vectorially as $\mathbf{E} \times \mathbf{B}/B^2$. Since electrons and ions move at almost exactly the same velocity, there is virtually no current. A neutral air wind **U** produces practically no motion of ions and electrons across the magnetic field.

Electric currents can flow at heights between 60 and 150 km, where the positive ions and electrons move in different directions because their ratios v/ω are different. Very little curent flows below 100 km, however, because the electron density N is usually so small. The current density is given by Ohm's law for the ionosphere which can be expressed in two equivalent ways:

$$\mathbf{j} = Ne(\mathbf{V}_i - \mathbf{V}_e) = \sigma \cdot \mathbf{E} \tag{15.30}$$

Eqn. 15.30 shows that the conductivity σ (a tensor quantity because **j** and **E** are usually not parallel) depends on the difference between the ion and electron velocities, as well as on N.

15.4.3 The ionospheric dynamo

The charge separation resulting from the different ion and electron motions is the basis of the ionospheric dynamo, which is driven by tidal winds (Section 15.4.4). The dynamo action was envisaged by Balfour Stuart in 1883, before its physical nature was known. The dynamo produces the ionospheric currents that cause the small daily variations in the geomagnetic field, first detected in the 18th century. Most of the dynamo current flows in the E layer where the conductivity σ is greatest. At night, the E layer current is very weak because N is small, although some current flows in the F layer. The electric field that arises from the charge separation causes drift motions of ions and electrons in the E and F layers, which contribute to the transport term in the continuity equation 15.15. Additional currents enter and leave the ionosphere in the auroral oval (Section 15.5.7), flowing along the lines of force of the earth's magnetic field. They are generated by processes that lead to charge separation in the magnetosphere.

For complicated geometrical reasons, charge separation in the ionosphere is particularly strong near the magnetic equator where a strong polarisation field develops in the E layer. It has the effect of creating high electric conductivity in the east–west direction, giving rise to a strong east–west electrojet current. The equatorial electrojet flows in a narrow belt extending about three degrees either side of the dip equator.

In addition to the wind-driven dynamo fields and currents in the ionosphere, electric fields and currents are generated in the magnetosphere and transmitted along geomagnetic field lines to the high latitude ionosphere and the auroral ovals, and spread to some extent to lower latitudes. They cause the strong localised east–west currents known as the auroral electrojets which, especially during magnetic storms, are major sources of heating at high latitudes mentioned in Section 15.2.2.

15.4.4 Atmospheric tides

Atmospheric tides form a complex system containing many components. They are forced (driven) by the heating effect of the sun or the gravitational attraction of the sun and moon. Unlike the case of marine tides (dominated by the moon's gravitational attraction) the atmospheric tidal motions are forced in three main ways: (1) the absorption of solar EUV and X-rays in the thermosphere; (2) the absorption of solar UV in the ozone layer; (3) the heating of the ground and lower atmosphere by solar visible and infra-red radiation. The tidal motions include diurnal (24-hour), semidiurnal (12-hour) and other components, plus a weak lunar gravitational (12.4-hour) component. The global winds described in Section 15.4.6 can be regarded as being a special type of diurnal tide.

15.4.5 Atmospheric waves

The upper atmosphere contains many kinds of wave. Besides the tides just mentioned, they include planetary waves (periods of days); gravity waves (period > 5 min); acoustic or compressional waves (period < 4 min); and infrasonic waves (periods < 1 s), some associated with aurora or with seismic activity.

Gravity waves are oscillations controlled by the buoyancy of the air, i.e. by gravity, and have a minimum period of about five minutes (the Brunt–Väisälä period) in the thermosphere. Unlike ocean waves which are virtually confined to the surface, gravity waves permeate the whole upper atmosphere. They have a complicated phase structure and, in addition to horizontal propagation, the phase velocity has a downward component which gives the appearance of a downward travelling wave on ionograms, although the group velocity and energy propagation is inclined slightly upwards.

Short and medium period gravity waves (5–30 min) are generated in the lower atmosphere by storms, by winds blowing over mountains and occasionally by man-made explosions. Long period gravity waves (periods 0.5–3 hours) originate from disturbances in the auroral oval; they often travel for thousands of kilometres, and are prevalent at times of geomagnetic disturbance. Such large-scale waves can have some effect on radio propagation, even at frequencies far above the ionospheric critical frequencies.

15.4.6 Thermospheric winds

The thermosphere is a vast heat engine driven by solar, magnetospheric and interplanetary sources, which produce the global temperature variations shown in Figure 15.2; these give rise to horizontal pressure gradients that drive winds in the directions shown by the arrows. The winds form a global circulation which carries heat away from the source regions and liberates it elsewhere. The velocity **U** is controlled by the Coriolis force due to the earth's rotation with angular velocity Ω, by the molecular viscosity of the air (coefficient μ), and by ion drag, the friction due to collisions between air molecules and ions. Ion drag exists because the ions are constrained by the geomagnetic field and cannot move freely with the wind (although they may be driven across the geomagnetic field by electric fields, Section 15.4.2). The equation of motion for the horizontal wind is:

$$d\mathbf{U}/dt = \mathbf{F} - 2\,\boldsymbol{\Omega} \times \mathbf{U} - KN(\mathbf{U} - \mathbf{V}) - (\mu/\rho)\ \nabla^2\mathbf{U} \qquad (15.31)$$

where **F** is the driving force due to the horizontal pressure gradients $\partial p/\partial x$, $\partial p/\partial y$. Its zonal (west-to-east) and meridional (south-to-north) components are, respectively:

$$F_x = -(1/\rho)\ \partial p/\partial x,\ F_y = -(1/\rho)\ \partial p/\partial y \qquad (15.32)$$

The vertical component of **F** is omitted, as it almost exactly balanced by gravity as in the hydrostatic equation, Eqn. 15.3. The product KN is the ion-neutral collision frequency. At great heights (small ρ) the kinematic viscosity (μ/ρ) is large and viscosity tends to smooth out any wind shears (variations of wind velocity with height).

Unlike winds in the lower atmosphere, thermospheric winds are described by the direction towards which they blow. Their direction depends on the ratio of Coriolis force to ion drag. Two limiting cases are:

$$U = F/(2\,\Omega \sin \varphi) \qquad (15.33)$$

$$U = F/(KN \sin I) \tag{15.34}$$

where φ is latitude. In the case of Eqn. 15.33, Coriolis force is dominant and ion drag is small, as at all heights below about 200 km, and the wind blows at right angles to the pressure gradient, as in the familiar weather maps of the lower atmosphere. But in the case of Eqn. 15.34, ion drag is strong and the wind is almost parallel to the pressure-gradient force, as in the daytime F layer. In the general case, both ion drag and Coriolis force are significant and the wind is inclined to **F**. Accordingly, the wind vectors in Figure 15.2 are nearly parallel to the temperature and pressure gradient by day, as in Eqn. 15.34, but are deflected by Coriolis force at night.

15.4.7 The vertical circulation

The pressure distribution and the wind velocity automatically adjust themselves to satisfy the continuity equation for the neutral air. This equation is analogous to Eqn. 15.15 except that production and loss processes are unimportant for the major constituents of the neutral air, and the continuity equation reduces to:

$$\partial n/\partial t = -\text{div}\,(n\mathbf{U}) \tag{15.35}$$

The pressure distribution and the wind velocity automatically adjust themselves to satisfy Eqn. 15.35. Thus, any divergence (convergence) of the horizontal winds is balanced by upward (downward) winds so that div ($n\mathbf{U}$) is small. Winds that blow along isobars, as in the lower atmosphere, are almost divergence free and so are ineffective in removing the horizontal pressure differences that drive them, which is why highs and lows in the lower atmosphere persist for days.

In the thermosphere the situation is different; the summer-to-winter winds illustrated in Figure 15.2 are accompanied by upward winds in equatorial and summer latitudes and downward winds at winter mid-latitudes. These vertical winds are small (only of order 1–10 m s^{-1}, as compared with 50–100 m s^{-1} for the horizontal winds) but they play an important role in the heat balance of the thermosphere. They produce changes in the molecular/atomic ratio in the neutral air which, through the photochemical scheme set out in Section 15.3.5 (point *h*), alter the balance between the production and loss processes in the F2 layer.

15.4.8 The F2 peak

Point *h* of Section 15.3.5 explained why the electron density N increases upwards from the F1 layer, because of the upward increase of the ratio q/β. Production and loss are roughly in balance at heights below the F2 peak. The peak occurs at the height where chemical control gives way to gravitational (or diffusive control), *viz*. where the transport terms in the continuity equation (Eqn. 15.15) become comparable to the production and loss terms. The $N(h)$ profile then looks roughly as in Figure 15.5.

Introducing the diffusion coefficient D for the ions, which is inversely proportional to the ion-neutral collision frequency, the F2 peak is governed by the relations (in which the suffix m denotes values at the peak height hmF2):

$$\beta_m \sim D_m/H^2 \tag{15.36}$$

$$N_m \sim q_m/\beta_m \sim I_\infty n[\mathrm{O}]/\{k' n[\mathrm{O_2}] + k'' n[\mathrm{N_2}]\} \tag{15.37}$$

Thus N_m (= NmF2) depends on the atomic/molecular ratio of the neutral air. The height hmF2 tends to lie at a fixed pressure level in the atmosphere, i.e. a fixed value of the reduced height z defined by Eqn. 15.7, unless shifted up or down by a neutral air wind or an electric field. A horizontal wind blowing towards the magnetic equator drives the ionization up magnetic field lines, raising the peak and increasing NmF2; opposite effects are produced by a poleward wind, see sketches *a* and *b* in Figure 15.5. Equatorward winds tend to occur at night, poleward winds by day. In addition, vertical drift is produced by east–west electric fields which produce upward/downward electromagnetic drift, through the vertical component of the drift velocity $\mathbf{E} \times \mathbf{B}/B^2$ (Section 15.4.2). This drift is effective at the magnetic equator, where it gives rise to the F2 layer equatorial anomaly (Section 15.5.5), but it is not very effective at mid-latitudes.

15.4.9 The topside ionosphere

At heights well above the F2 peak, the ion and electron distribution is diffusively controlled, gravity being balanced by the pressure gradient terms in Eqns. 15.28 and 15.29, much as is the neutral air in the hydrostatic equation, Eqn. 15.3. Above about 700 km there is a gradual transition from the O^+ ions of the F2 layer to the H^+ ions of the protonosphere, some He^+ ions being present also. An important feature of the topside F2 layer is the flux of plasma along the geomagnetic field lines to or from the overlying protonosphere, which involves the charge-exchange reaction:

$$\mathrm{O^+ + H \Leftrightarrow H^- + O} \tag{15.38}$$

By day the F2 layer supplies ions to the protonosphere through the forward reaction in Eqn. 15.38, but at night the reverse reaction, accompanied by downward flow, plays an important part in maintaining the O^+ content of the mid-latitude F2 layer, so the protonosphere acts as a reservoir of plasma. At higher latitudes, where the geomagnetic field lines are not closed, the ionosphere (particularly at auroral latitudes) acts as a source of energetic O^- ions to the magnetosphere.

15.5 Ionospheric phenomena

15.5.1 The D layer

The D layer plays an important part in LF/VLF propagation. At MF and HF it is important because of its absorbing properties, which stem from the relatively high air density and large electron collision frequency. Daytime absorption measurements show a fairly regular dependence of D layer ionisation on solar zenith angle. The attachment of electrons to form negative ions at sunset, and their detachment at sunrise, causes changes of LF/VLF reflection height. On some winter days the electron density is abnormally high, apparently because of enhanced concentrations of neutral nitric oxide which is ionised by solar Lyman α (Figure 15.1). D layer chemistry is extremely complex, and is not considered here.

Some D layer processes are linked with meteorological phenomena at lower heights. The stratospheric warmings that occur in winter at moderately high latitudes, representing major changes in the large-scale circulation, are accompanied by increases of radiowave absorption in the D layer. The auroral D layer is often enhanced by particle precipitation.

15.5.2 The E layer

The normal E layer follows the idealised Chapman layer behaviour of Eqn. 15.14 quite well, except that the index n is about 0.3 instead of 0.25, because of complications such as the vertical temperature gradient and drifts due to tidal electric fields. Its chemistry seems quite well explained by the photochemical scheme of Section 15.3.5.

15.5.3 Sporadic E

Sporadic E, also known as Es, consists of thin layers of enhanced ionisation around 100–120 km, which are often dense enough to affect radio propagation. Many types of Es have been defined, distinguished by their appearance on ionograms, but in reality there may be only three main types: mid-latitude Es layers, about 1 km thick, produced by wind shears (small-scale gradients of wind velocity) interacting with the geomagnetic field, which persist because they contain long-lived metal ions (Section 15.3.4), probably originating in meteors; equatorial Es, a plasma instability that accompanies the large electron drift velocities in the equatorial electrojet (Section 15.4.3); auroral Es, produced by precipitated kilovolt electrons.

15.5.4 F layer behaviour

The daytime F1 layer follows the simple Chapman formula of Eqn. 15.14 quite well, the F2 layer does not (Section 15.3.2 and Figure 15.4). Departures from purely solar-controlled layer behaviour are seen also in Figure 15.6 which shows the F2 layer variations at a mid-latitude station throughout a year, the data being arranged according to the solar 27-day rotation period, and in Figure 15.7 which is a latitude/local time map of noon F2 layer critical frequency for solar maximum in the American longitude sector.

The shape of the day/night variations of *Nm*F2 and *hm*F2 is largely due to

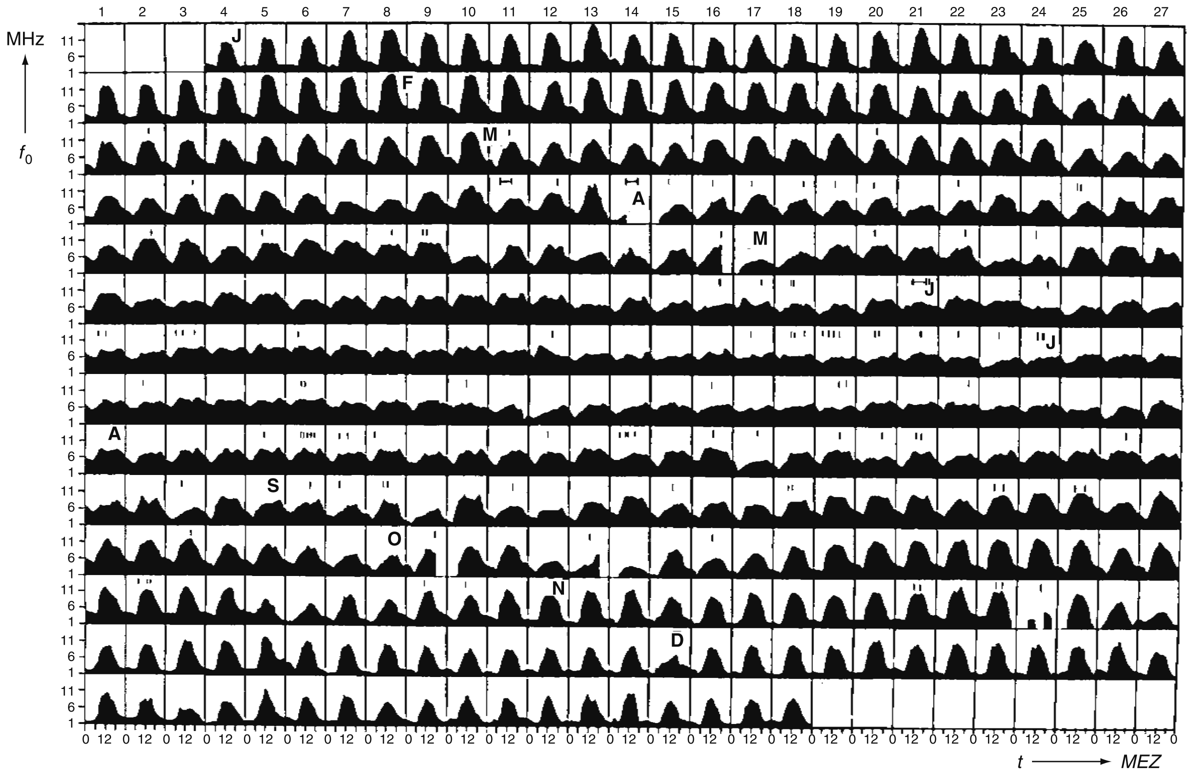


Figure 15.6 *Lindau mountain diagram showing the critical frequency foF2 at Lindau (latitude 52°N) throughout the post-sunspot maximum year 1960. Each box represents 00–24 LT and 1–16 MHz; the letters mark the first day of calendar months (Max-Planck-Institut für Aeronomie, Katlenburg-Lindau, Germany)*

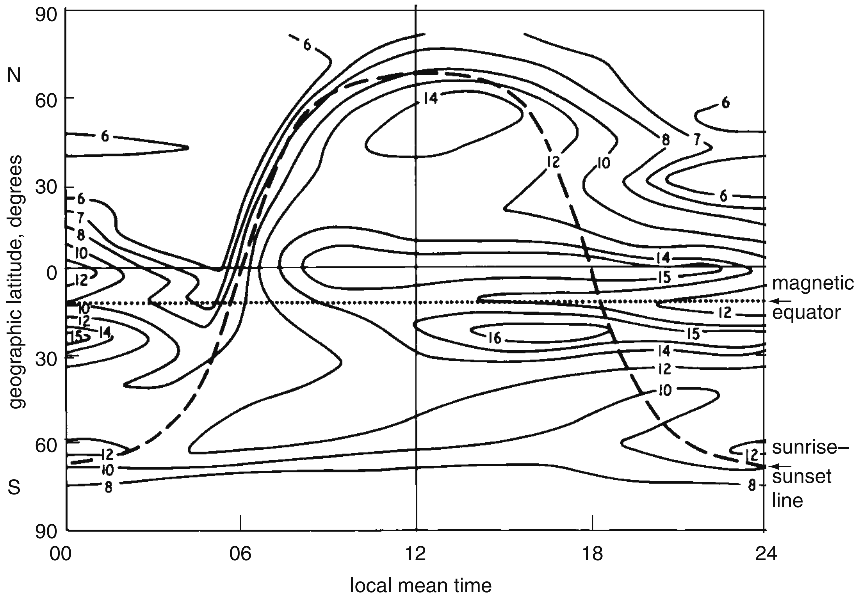


Figure 15.7 Prediction map of F2 layer critical frequency in the American longitude sector, for December 1957 (solar maximum) (US Department of Commerce, Central Radio Propagation Laboratory)

neutral air winds, and electric fields to a lesser extent. Around sunset, the meridional component of the wind changes from equatorward to poleward, and the resulting F2 layer drift changes from downward to upward (see sketches in Figure 15.5), which causes the commonly observed increase of *Nm*F2 in the late evening. The details depend on the local geometry of the geomagnetic field.

At night the F2 layer is maintained (*a*) by being raised by winds to heights where the loss coefficient β is small; (*b*) by inflow from the overlying protonosphere; (*c*) by weak EUV fluxes from the night sky; and, perhaps, (*d*) by energetic particles.

15.5.5 F2 layer anomalies

It is clear from Figures 15.4, 15.6 and 15.7 that the F2 layer has complicated variations. The variations throughout the year are sometimes described, mainly as a matter of convention, in terms of anomalies which are superimposed in different ways at different places. These anomalies are described as seasonal (tendency for maximum *Nm*F2 to occur in local winter); annual (tendency for maximum *Nm*F2 in January) and semiannual (tendency for maximum *Nm*F2 at equinox). Data from ionosonde stations in both hemispheres are needed to separate these anomalies, although it is not easy to find really comparable northern and southern stations, since they have to be matched both geographically and geomagnetically.

It is well established that the seasonal (summer/winter) anomaly seen in the Figures results from changes of atomic/molecular ratio of the neutral air produced by the global circulation (Section 15.4.6), which affect the production/loss ratio q/β as shown by Eqn. 15.37. The same is thought to be true of the semiannual anomaly; the data indicates that the relative importance of the seasonal (summer/winter) and semiannual variations, in different latitudes and longitudes, depends on the geometry of the earth's magnetic field. Detailed calculations can now account for the fact that the seasonal anomaly prevails in the North American/European sector and the Australasian sector, and the semiannual anomaly is most pronounced in low latitudes and the South Atlantic sector. The annual (nonseasonal) anomaly is that NmF2 worldwide is 15–20 per cent greater in January than in July. This is much more than would be expected from the 3 per cent annual difference in sun–earth distance, which causes a 6 per cent variation of the incoming flux of solar radiation, so another cause must act to reinforce it.

Near the geomagnetic equator, the main F2-layer feature is the daytime equatorial trough or Appleton anomaly, seen in Figure 15.7. It is produced by dynamo electric fields which cause a drift of plasma away from the magnetic equator.

15.5.6 Small-scale irregularities

The ionosphere is structured with a wide range of scale sizes. The small-scale irregularities at auroral and equatorial latitudes are due to plasma instabilities. At mid-latitudes, besides spread F and sporadic E, there is weaker irregular structure that causes radiowave fading and scintillation of radio signals traversing the ionosphere. Medium-scale structure, 10–1000 km in scale, is widely observed but not well explained.

15.5.7 The auroral ovals, polar cap and trough

At high latitudes the ionosphere is profoundly affected by charged particles and electric currents from the magnetosphere that cause aurora and magnetic storms. The particles and currents mainly enter in the auroral ovals, rings 2000–3000 km in diameter that surround the magnetic poles (Figure 15.8). In turn the magnetosphere is influenced by the solar wind and the interplanetary magnetic field which, by their interaction with the geomagnetic field, set up an electric field throughout the high latitude ionosphere. This electric field, pointing in the dawn-to-dusk direction as shown by the arrow in Figure 15.8, causes a fast day-to-night flow of plasma across the polar cap which maintains the F2 layer at polar latitudes in winter.

Much of the high latitude ionospheric structure is related to the auroral ovals and the magnetosphere, to which the ionosphere is linked by geomagnetic field lines. As mentioned previously, electric currents and energetic particles can flow into the ionosphere, causing heating and consequent expansion of the air, which

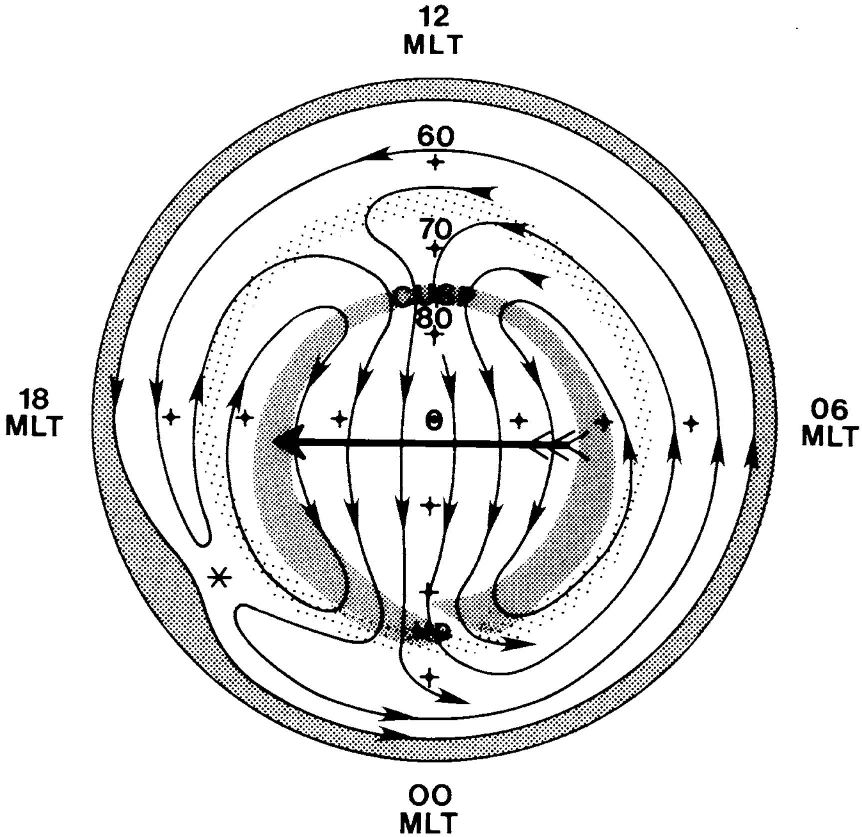


Figure 15.8 *The auroral oval and associated features. View of the north polar region centred on the magnetic pole. The bottom of the diagram represents magnetic midnight, 00 MLT; the top represents magnetic noon, 12 MLT; the outer boundary is at 50° magnetic latitude, the crosses being at magnetic latitudes 80°, 70°, 60° the star marks the centre of the trough region. The diagram corresponds roughly to moderate magnetic activity (Rutherford Appleton Laboratory)*

modifies the chemical composition and the local thermospheric winds. The auroral ovals are the site of many active phenomena, especially particle precipitation which maintains the E layer and affects all the ionospheric layers. Deep troughs exist on the low latitude side of the auroral ovals in the evening and night sector (around the star in Figure 15.8), where there are no significant local sources of ionisation and insufficient transport of plasma from elsewhere to maintain a dense F2 layer.

15.6 Solar–terrestrial relations and ionospheric storms

15.6.1 Solar flares and sudden ionospheric disturbances

The intense X-rays from solar flares produce strong enhancements of electron density in the D layer, giving rise to sudden ionospheric disturbances (SID), a term which embraces radio shortwave fadeouts, effects on LF/VLF propagation and magnetic perturbations. These occur in the sunlit hemisphere and typically last for half an hour. Soon afterwards MeV particles from the flare arrive and cause enhanced radio absorption at high latitudes (polar cap absorption, PCA).

15.6.2 Geomagnetic storms

Streams of plasma ejected from the solar corona travel outwards as gusts in the solar wind, taking about 24 hours to reach the earth. The impact of a plasma stream on the magnetosphere causes a magnetic storm, starting with a compression of the geomagnetic field observed at the ground as a storm sudden commencement (SSC). The interaction of the plasma stream and the magnetosphere sets up a ring current around the earth, mainly carried by keV ions, which causes a decrease of the geomagnetic field, greatest at 12–24 hours after the SSC and decaying in 2–3 days. The auroral oval expands equatorwards and associated ionospheric features move to lower latitudes during the storm.

15.6.3 Ionospheric storms

During a geomagnetic storm, particle precipitation at high latitudes produces disturbances at D layer heights, probably chemical in nature, such as through the production of extra nitric oxide which is then ionised by solar Lyman α. The effects are propagated to the mid-latitude D layer, where they last for several days. The F2 layer is often profoundly affected during magnetic storms, with severe effects on radio propagation. At mid-latitudes the F2 layer electron density initially increases, then decreases during the main phase, and recovers in 2–3 days. Examples of these negative storm effects can be seen in Figure 15.6. In winter, however, there is often a positive effect, i.e. an increase of F2 layer density in the main phase. Storm effects are very complicated, with many differences between individual events that depend on place, the time of year and the local time when the storm begins, so summarised descriptions can be misleading.

The F2 layer storm effects are largely due to changes of temperature and the horizontal and vertical thermospheric winds, driven by the enhanced heating in the auroral zone which modifies the quiet day wind system shown in Figure 15.2. Equatorward winds are enhanced, causing lifting of the F2 layer which affects NmF2 in the manner shown in Figure 15.5. The equatorward winds bring air with an enhanced molecular/atomic ratio from high latitudes to mid-latitudes, which in turn increases the loss coefficient β (see Eqn. 15.37) and hence decreases

*Nm*F2 at mid-latitudes; this is the main cause of negative storm effects at mid-latitudes. Positive F2 layer storm effects in winter and at low latitudes are due to decreases of molecular/atomic ratio, induced by increased downward air motions, or to lifting produced by winds or by enhanced electric fields.

15.6.4 Space weather

The ionosphere is part of the space environment, and space weather includes the short-term variability of the ionosphere, which is still not well understood. Its major aspects are the vagaries of sporadic E and of the F2 layer. At mid-latitudes, day-to-day variability of *Nm*F2 is typically 20 per cent by day and 30 per cent at night: the factors thought to cause it are geomagnetic disturbance, waves and tides transmitted from the lower atmosphere and variations of the incoming solar photon radiation, which is probably a minor factor as far as day-to-day variations are concerned.

15.6.5 Long-term change

This is a different kind of disturbance! Increasing concentrations of carbon dioxide and methane that cause global warming in the lower atmosphere result in global cooling in the upper atmosphere (because of the enhanced loss of heat to space by infrared radiation). If the mesospheric concentrations of these gases are doubled in, say, the next fifty years, as is anticipated, it is estimated that:

- the mean tropospheric temperature will increase by 1–2 K
- the mean mesospheric temperature will decrease by 5 K
- the mean exospheric temperature will decrease by 40 K

The resulting thermal contraction of the thermosphere will decrease the air density at 300 km by about 40 per cent, thereby prolonging satellite lifetimes, and lower the E layer by about 2 km and F2 layer by 20 km, with some change of MUF. There is experimental evidence that mesospheric temperature has decreased in recent decades, and lowering of *hm*F2 has been reported at some stations, although the observations do not give a clear-cut result, as there appear to be regional variations. Other possible long-term effects are changes in upper atmosphere tides, resulting from the depletion of the ozone layer where some tidal forcing originates, and chemical contamination by spacecraft launches.

15.7 Conclusions

Contemporary ionospheric science has many aspects: as a branch of solar-terrestrial physics and the earth's environment; the science of radiocommunications, e.g. predicting and forecasting; the mathematics of wave propagation in the ionosphere; man-made radio transmissions, and natural electromagnetic emissions, magnetic pulsations and other phenomena. As a natural plasma, often nonlinear and nonthermal, the ionosphere is a site for active experiments with radiowaves, powerful enough to produce observable ionospheric modifications, in the form of increased temperatures and plasma instabilities, and with chemical releases and particle beam experiments. Such ionospheric plasma physics experiments provide new physical insights, although they often raise more questions than they solve.

Problems remain at the upper and lower boundaries; i.e. the interactions with the lower atmosphere and the magnetosphere, and how their effects spread worldwide. Large electric fields generated in thunderstorms can affect the ionosphere. There are suggestions of topographic influences, e.g. of mountain ranges, on the ionosphere. The reverse question of ionospheric, geomagnetic and solar effects on weather and climate is contentious; possible relations between magnetic and solar disturbance and the weather have been found, but their physical mechanisms are not well established.

The basic mechanisms of the ionosphere seem to be known, but surprises are possible in any active area of science, and many questions remain. The everyday monitoring of this important part of the environment is as vital for progress in the future as it has been in the past.

15.8 Bibliography

JURSA, A.S. (Ed.): 'Handbook of geophysics and the space environment' (Air Force Geophysics Laboratory, Bedford, Mass., 1985)

HARGREAVES, J.K.: 'The upper atmosphere and solar-terrestrial environment' (Cambridge University Press, 1992)

RATCLIFFE, J.A.: 'Sun, earth and radio' (World University Library, 1970)

RISHBETH, H., and GARRIOTT, O.K: 'Introduction to ionospheric physics' (Academic, New York, 1969)

RUSSELL, C.T. and KIVELSON, M.: 'Introduction to space physics' (Cambridge University Press, 1995)

SCHUNK, R.W., and NAGY, A.F.: 'Ionospheres: physics, plasma physics and chemistry' (Cambridge University Press, 2000)

TASCIONE, T.F.: 'Introduction to the space environment' (Orbit, 1988)

Chapter 16

Ionospheric propagation

Paul S. Cannon and Peter A. Bradley

16.1 A systems perspective

The ionosphere, which extends from ~60 km to ~1000 km, significantly affects the propagation of high-frequency (HF) to ultrahigh-frequency (UHF) signals which pass through it. The effects are varied but include refraction, retardation and scintillation. Ground–ground HF communications systems, ground–space communications systems, single frequency GPS (global positioning system), HF over-the-horizon radars, satellite altimeters and space-based radars [1] are examples of radio systems constrained by this medium.

HF communications and radar systems are of course dependent on the ionosphere for long distance operation but most systems are degraded by the ionosphere. Loss of phase lock and range errors in GPS are examples of such deleterious effects.

As with all radio systems the design of ionospheric systems must address the issues of signal-to-noise plus interference ratio (SNR), multipath and Doppler shift and spread. For many systems, however, signal group delay, phase scintillation and other issues are also important. Table 16.1 provides quantitative estimates of some of these effects.

For ionospheric signals the SNR is determined by a number of factors. For HF signals whether the signal is actually reflected by the ionosphere is critical. All transionospheric signals also experience some excess attenuation over free space but, as we will see, because this is frequency dependent, the effects at higher frequencies are generally negligible. In the HF band interference dominates over the noise but further discussion falls outside of the remit of this Chapter. The reader is referred to Reference [2] for further discussion.

Multipath arises from various sources. A transmitted HF signal can be reflected from more than one of the several layers in the ionosphere. The transmission of a single pulse of energy is consequently received as a number of

Table 16.1 *The effects of ionospheric variability on system operation*

	Communications	*Surveillance*	*Navigation*
Systems	HF communications and broadcasting	HF OTH radar HF electronic warfare	GPS GLONASS GNSS
	VLF-MF communications and broadcasting		ground-based systems
	UHF/SHF satellites	UHF/SHF radars spaced-based SAR geolocation	
	LEO and MEO SATCOM		
Effects	loss of communications changes in area of coverage low signal power fading error rate changes ground-wave/sky-wave interference	range and bearing errors skip zone changes multiple targets generated loss signal power clutter target masking	range errors loss of phase lock and data loss
	loss of communications fading error rate changes	range and bearing errors multiple targets loss of target discrimination loss of phase coherence across SAR aperture	

Severity	>20 dB fades at HF even no signal	>20 dB fades at HF even no signal	single frequency position errors up to 35 metres	loss of position update
	intersymbol interference	100s km range errors 100 deg bearing errors		
	30 dB fades at UHF 20 dB fades at L-band	over 500 m at UHF 1000s of erroneous returns clutter		
Causes	changes in E and F region electron density profile D region absorption irregularities multipath	changes in E and F region electron density profile D region absorption irregularities particle precipitation	total electron content variations	irregularities
	irregularities	total electron content variations irregularities particle precipitation		

pulses which may be distinct or which may overlap. This situation is further complicated because the signals can also bounce off the ionosphere more than once, having been reflected from the ground in between. The earth's magnetic field also splits an arbitrary signal into two orthogonal polarisations which travel at a different speed and follow a slightly different path. There sometimes exist the so-called high and low-angle paths. At VHF and above the multipath manifests itself as rapid phase and amplitude scintillation of the signal [3, 4].

The ionosphere is in constant turmoil; more so at some latitudes than at others. This introduces Doppler shifts and spreads to the signals.

16.2 Ionospheric morphology

The ionosphere is a lightly ionised region of the atmosphere lying mainly in the altitude range 60 km to 1000 km, Figure 16.1. This ionisation is caused by several mechanisms. The most important of these, at nonauroral latitudes, are the sun's extreme ultra violet (EUV), X-ray and Lyman α radiation together with solar cosmic rays. At high latitudes, particularly during magnetically active periods, the effects of energetic particles and electric fields are also important.

The rates of ionisation at any altitude depend on the atmospheric composition as well as the characteristics of the incident radiation at that height. As the solar radiation propagates down through the neutral atmosphere the various frequency (energy) bands of this radiation are attenuated by different amounts. At the same time the composition of the atmosphere alters with altitude. Consequently, different ionisation processes become predominant at different heights resulting in a layered structure. The principal layers are designated D, E and F (Figure 16.1), each being characterised by a different set of ionisation

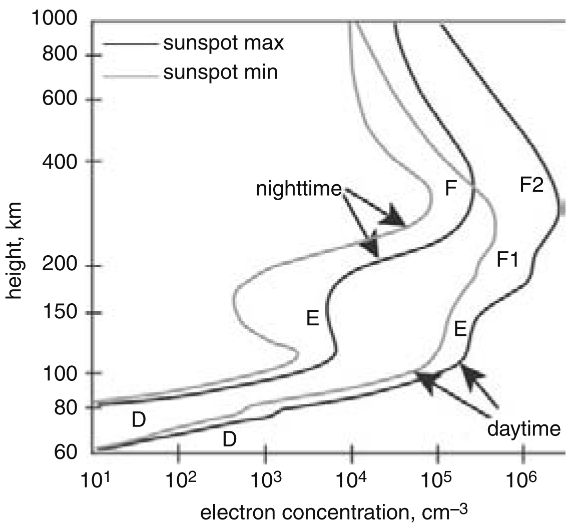


Figure 16.1 The ionosphere

processes. These regions are themselves layered or structured into the E, Es, F1 and F2 regions. The number of layers, their heights and their ionisation density vary with time and in space. The high latitude ionosphere is particularly complicated and will not be dealt with here, but is described in Reference [5].

16.3 Theory of propagation in the ionosphere

16.3.1 Vertical propagation – no collisions

The ionosphere is a dispersive medium, that is to say the refractive index of the medium varies as a function of frequency. As a consequence, different frequencies travel with different speeds causing spreading of the pulse. The ionosphere is also layered, as we have described, and this leads to multipath propagation. The earth's magnetic field introduces a further complication in as much as it renders the ionosphere anisotropic. This means that the incident ray will be split into two rays on entering the ionosphere in a manner similar to that of crystal optics. The two differently polarised rays are referred to as ordinary (O) and extraordinary (X). One result of this splitting is the phenomenon of Faraday rotation of the plane of polarisation as it traverses the ionosphere. The ionosphere, however, is above all a region where the refractive index is less than unity resulting in a group velocity which is less than that of light in free space. A consequence of this is that vertically launched rays can be reflected and oblique rays are progressively bent away from the vertical. Obliquely launched HF rays can in fact be turned back towards the ground providing the phenomenon of sky-wave communications.

A fundamental ionospheric propagation equation is the Appleton-Lassen formula, which describes the complex refractive index, n^2. In particular, the equation describes the propagation of HF waves. When collisions between the electrons and neutrals are negligible (in the E and F regions) the real part of the refractive index μ can be determined from a simplified form of the Appleton-Lassen formula:

$$\mu^2 = 1 - \frac{2X(1-X)}{2(1-X) - Y_T^2 \pm [Y_T^4 + 4(1-X)^2 Y_L^2]^{1/2}} \tag{16.1}$$

where

$$X = \frac{Ne^2}{\varepsilon_0 m\omega^2}, \; Y_L = \frac{eB_L}{m\omega} \text{ and } Y_T = \frac{eB_T}{m\omega} \tag{16.2}$$

The subscripts T and L refer to transverse and longitudinal components of the imposed geomagnetic flux (B) with reference to the direction of the wave normal of propagation. N is the electron density, e is the charge on the electron, m is the mass of the electron, ε_0 is the permittivity of free space and ω is the angular wave frequency. The plus sign above refers to the ordinary wave and the minus sign refers to the extraordinary wave.

If, the magnetic field is ignored the above reduces to:

$$\mu^2 = 1 - X = 1 - \left(\frac{f_N}{f}\right)^2 = 1 - k\frac{N}{f^2} \tag{16.3}$$

where

$$f_N = \left(\frac{Ne^2}{\varepsilon_0 m}\right)^{0.5} \tag{16.4}$$

This parameter, f_N, is known as the plasma or critical frequency. The constant

$$k = 80.5 \tag{16.5}$$

when N is in electrons per cubic metre and f is in Hz.

Reflection at vertical incidence occurs when $\mu = 0$, i.e. when $f = f_N$. At higher frequencies, where $f > f_N$, the refractive index never drops to zero, reflection never occurs and the vertical ray passes through the ionosphere into space. This is the normal condition for signals above the HF band.

16.3.2 Group path and phase path

Two important concepts in ionospheric radio propagation are those of group path and phase path. Due to the embedded geomagnetic field the ionosphere is anisotropic, that is the phase velocity, v, of the wave is dependent on direction. As a consequence the phase paths and group paths differ. It can be shown that the angle α between the wave normal and the ray direction is given by:

$$\tan \alpha = -\frac{1}{v}\frac{dv}{d\theta} = +\frac{1}{\mu}\frac{d\mu}{d\theta} \tag{16.6}$$

where θ is the angle at which the wave normal direction cuts a reference axis. The phase path in an anisotropic medium is:

$$P = \int_s \mu \cos \alpha \, ds \tag{16.7}$$

integrated over the ray path s. The corresponding group path, P' is:

$$P' = \int_s \mu' \cos \alpha \, ds \tag{16.8}$$

where μ' is the group refractive index such that:

$$\mu' = \frac{c}{\mu} \tag{16.9}$$

16.3.3 Oblique propagation

Obliquely launched HF rays can also return to the ground, and this can even occur for frequencies above the plasma frequency. At HF and above the medium is slowly varying and changes little within one wavelength. As such, ray theory can be applied and reflection can be described in terms of Snell's law within a stack of thin slabs (each several wavelengths thick). Since the plasma frequency increases with height the refractive index decreases with height (Eqns. 16.4 and 16.3), the ray bends round and is eventually reflected. If, however, the frequency is above $\sim f_N/\cos i_0$ the signal is not reflected and continues into space, albeit after some refraction. This is generally the case for VHF signals and above.

A modified version of Snell's Law, known as Bouguer's law, is used to account for the earth's curvature; the ionosphere is assumed to be concentric with the earth and to have no embedded geomagnetic field. It enables the angle of incidence at a slab of refractive index μ and height h to be determined for a ray launched with elevation angle $\Delta_{\hat{u}}$ relative to the earth of radius R (Figure 16.2). The law gives:

$$R \cos \Delta_{\hat{u}} = \mu(R + h) \sin i \tag{16.10}$$

Application of Eqns. 16.7 and 16.8 yields:

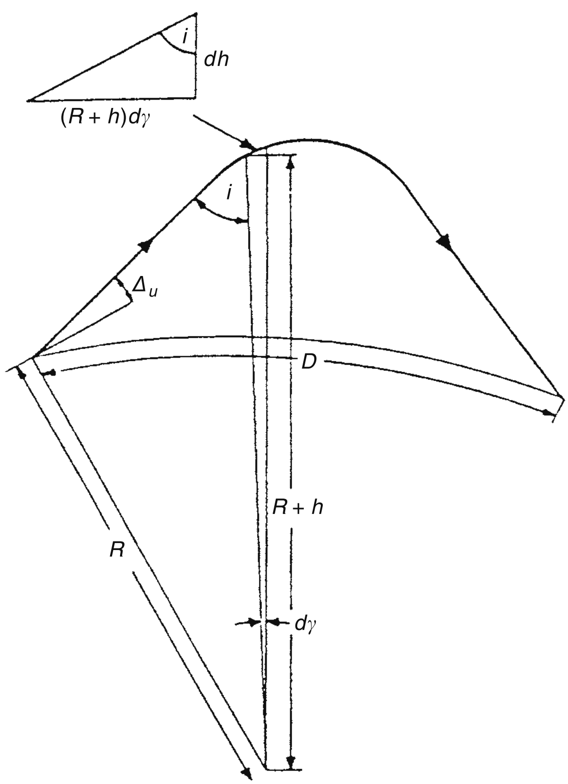


Figure 16.2 Ray path geometry for spherically stratified ionosphere

$$P = \int_0^h \frac{\mu \, dh}{\cos i} = \int_0^h \frac{\mu^2 \, dh}{\sqrt{\mu^2 - \left(\frac{R}{R+h}\right)^2 \cos^2 \Delta_u}} \tag{16.11}$$

$$P = \int_0^h \frac{dh}{\sqrt{\mu^2 - \left(\frac{R}{R+h}\right)^2 \cos^2 \Delta_u}} \tag{16.12}$$

and

$$D = R\int_0^\gamma d\gamma = R\int_0^h \frac{\tan i \, dh}{R+h} = R^2 \cos \Delta_{\mathrm{u}} \int_0^h \frac{dh}{(R+h)^2 \sqrt{\mu^2 - \left(\frac{R}{R+h}\right)^2 \cos^2 \Delta_u}} \tag{16.13}$$

In reality of course the ray paths travel in an ionosphere which is not concentric and which has a magnetic field. Often, we find, however, that approximations such as the one just described provide adequate accuracy. This is particularly true at mid-latitudes.

16.3.4 Absorption

If the effects of collisions are included then Eqn. 16.1 becomes more complicated *via* the introduction of a term $Z = v/\omega$, where v is the electron-neutral collision frequency:

$$n^2 = (\mu - i\chi)^2 = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \sqrt{\frac{Y_T^4}{4(1 - X - iZ)^2} + Y_z^2}} \tag{16.14}$$

For a wave travelling in the z direction at time, t:

$$E = E_0 \exp i\left(\omega t - \frac{w}{c} nz\right) \tag{16.15}$$

where, c is the velocity of light, or:

$$E = E_0 \exp\left(\frac{-\omega}{c} \chi z\right) \exp i\left(\omega t - \frac{w}{c} \mu z\right) \qquad 16.16)$$

If χ, the imaginary term in the refractive index, is nonzero the quantity $(\omega\chi/c)$ is a measure of the decay of amplitude per unit distance and is called the absorption coefficient, κ (kappa):

$$\kappa = \frac{\omega\chi}{c} \tag{16.17}$$

In the absence of a magnetic field the absorption, in nepers per metre (1 neper = 8.69 dB) is given by:

$$\kappa = \frac{\omega XZ}{2c\mu(1+Z^2)} = \frac{e^2}{2\varepsilon_0\, mc\mu}\frac{N\nu}{(\omega^2+\nu^2)} \tag{16.18}$$

When ω is much greater than ν and μ is close to unity Eqn. 16.18 gives:

$$\kappa = \frac{e^2 N\nu}{2\varepsilon_0\, mc\omega^2} \tag{16.19}$$

This is called nondeviative absorption and arises primarily in the D region. On the other hand, near reflection, when μ becomes small:

$$\kappa = \frac{\nu}{2c}\left(\frac{1}{\mu} - \mu\right) \tag{16.20}$$

and the absorption is deviative since it occurs in a region where considerable ray deviation takes place.

16.4 Ray tracing

16.4.1 Introduction

It is often necessary to evaluate the point where an HF ray launched upward towards the ionosphere returns again to the ground and this requires a ray tracing procedure. If the ray is of sufficiently high frequency it will penetrate the ionosphere providing a satellite to ground or ground to satellite path. Again, detailed ray trajectories may be required but more likely calculations such as group delay will then be needed.

16.4.2 Virtual techniques

The simplest HF ray tracing technique (Figure 16.3) is known as virtual ray tracing (often called sky billiards) and it assumes that the actual propagation can be approximated by reflection from a simple horizontal mirror at an appropriate height. These concepts are formalised in the secant law, Breit and Tuve's theorem and Martyn's equivalent path theorem, e.g. Reference [6]. This approach is implemented in most HF prediction codes. The major advantage of this approach is that it is computationally efficient. Its major disadvantage is that it cannot easily deal with a horizontally nonstratified ionosphere. It may, therefore, be considered an approximate technique and more complicated ray tracing methods are necessary in certain situations. Virtual techniques are, however, well matched to a median ionospheric database and simple ionospheric profiles. Indeed, in most circumstances the use of a more complicated model is not justified given the accuracy of the ionospheric electron density models generally used.

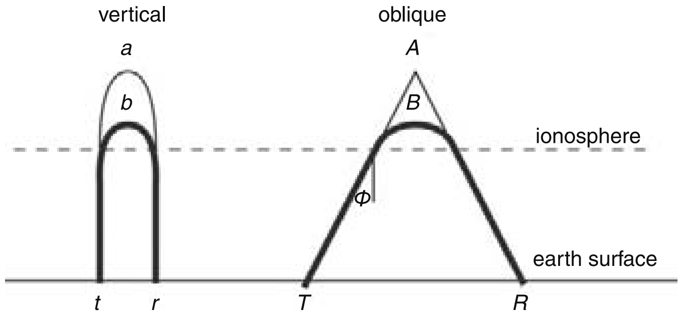


Figure 16.3 Geometry for ray propagation; earth and ionosphere are plane surfaces; the transmitter and receiver are at T (t) and R (r) respectively, the true point of reflection is at B (b) and the virtual point of reflection is at A (a)

16.4.2.1 Simple transionospheric models

A radio signal, which penetrates the anisotropic ionosphere, is modified in a number of ways. Both large-scale changes due to variations in the electron density and small-scale irregularities affect the signal. The effects include scintillation, absorption, variations in the direction of arrival, group path delay, dispersion, Doppler shift, polarisation rotation, refraction and phase advances. All of these effects are proportional to the total electron content, N_T, or its time derivative and inversely proportional to frequency to the appropriate power. (The total electron content (TEC) is the total number of electrons in a column of unit area, along the path from the transmitter to the receiver.) For frequencies well above the critical frequency the following simple relationships can be used:

RF carrier phase advance:

$$\Delta\Phi = \frac{8.44 \times 10^{-13}}{f} N_T \text{ (radians)} \tag{16.21}$$

where N_T is measured in electrons.m^{-2} and the frequency f is measured in Hz.

Group path delay: the excess time delay, over the free space transit time is given by:

$$\Delta t = \frac{40.3}{cf^2} N_T \text{ (s)} \tag{16.22}$$

Distortion of pulse waveforms: the ionospheric dispersion causes pulse broadening by virtue of the differential time delay across the signal bandwidth, Δf according to:

$$\Delta t = \frac{80.6 \times 10^{6}}{cf^3} \Delta f N_T \text{(s)} \tag{16.23}$$

Faraday rotation: At frequencies above 100 MHz this may be approximated by:

$$\Omega = \frac{2.365 \times 10^{-5}}{f^2} B_L N_T \text{ (radians)} \tag{16.24}$$

where B_L is the geomagnetic field (nT) component parallel to the wave direction taken at the mean ionospheric height.

16.4.3 Numerical ray tracing

Numerical ray tracing is important whenever very accurate ray tracing is required, i.e. the precise assessment of magnetic-ionic effects. The standard approach for numerical ray tracing is due to Jones and Stephenson [7]. The program is based on the solutions of a set of six coupled, first-order, non-linear differential equations, in spherical coordinates (providing that the time dependencies of the ionosphere are neglected).

Numerical ray tracing techniques are as accurate as the ionospheric model allows but they are computationally intensive to run. In many applications the issue of ray homing is critical and several techniques for solving this problem have been reported.

16.4.4 Analytic ray tracing

A technique intermediate between the simple but less accurate virtual technique and the complicated and accurate numerical technique also exists. The analytic technique relies on describing the ionosphere by functions that can be integrated. The technique was first pioneered by Croft and Hoogasian [8] for realistic spherical earth models and was extended by Milsom [9]. In its simplest form it requires the use of quasiparabolic (QP) and linear ionospheric segments to approximate the true electron density profile.

Multi-QP (MQP) [10, 11] techniques provide a good compromise between computational speed and performance for many applications. More recently, techniques have been developed which allow accurate ray tracing to take place even when there are strong electron density gradients along the path [12, 13]. The MQP analytic approach has a further significant advantage over virtual techniques; it provides an analytic calculation of the received signal power. The MQP does, however, ignore magneto-ionic effects although an estimate for their inclusion has been developed by Bennett *et al.* [14]. It has been estimated that neglecting the magnetic field can, in extreme circumstances, give rise to a 15 per cent error in ground range at 10 MHz [15]. At higher frequencies the error from neglecting the magnetic field diminishes.

16.5 The basic MUF, multipath and other HF issues

Figure 16.4 shows the calculated ray paths passing through a simple single-layer ionosphere at three separate frequencies, launched with a series of different

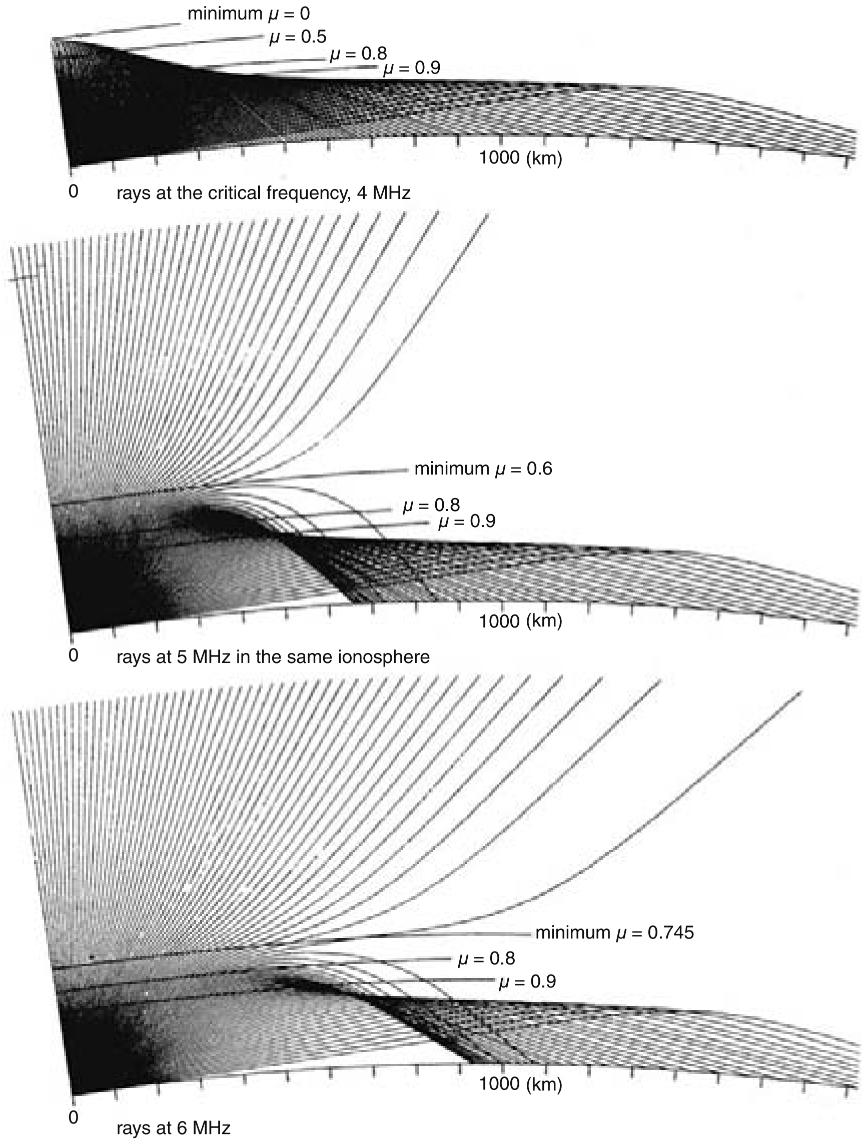


Figure 16.4 Ray paths for propagation at three frequencies via a simple Chapman model ionosphere of critical frequency 4 MHz, height of maximum electron density 300 km and scale height 100 km

elevation angles from a ground-based transmitter. A number of features are apparent:

- at the lowest frequency there is sufficient ionisation present to reflect the waves at all elevation angles, including the vertical; at the higher

frequencies, rays launched with an elevation angle greater than some critical value escape;
- waves launched more obliquely, generally, travel to greater ranges
- waves suffer more refraction at greater heights
- waves of higher frequencies are reflected from a greater height
- waves launched more obliquely are reflected from a lower height

The maximum range attainable after one ionospheric reflection arises for rays launched at grazing incidence and this depends primarily on the height of maximum electron density. For typical E, F1 and F2 layers, the maximum range is 2000, 3400 and 4000 km, respectively.

For a given ionosphere there will be some limiting upper frequency reflected vertically at the height of maximum electron density. At frequencies above this critical frequency there is a ground distance out from the transmitter at points along which illumination is not possible by waves reflected from the ionosphere. This distance is known as the skip distance. The skip distance increases as the wave frequency increases and in the limit for a very high frequency can extend to the maximum ground range possible for rays launched at grazing incidence; in that case all rays escape into space. It follows that there is a maximum frequency for which waves can be reflected to a fixed point of reception. This is the frequency making the distance from the transmitter to the point equal to the skip distance. The frequency is known as the basic maximum usable frequency (BMUF). The BMUF is defined in Reference [16] as the highest frequency that can propagate between ground-based terminals on a specified occasion by ionospheric refraction alone.

The BMUF increases with ground distance and depends also on the amount of ionisation present. It depends too on the height of the ionosphere since the determining factor as to whether reflection or transmission occurs is the angle of incidence at the layer. The greater the layer height, the steeper the angle of incidence to achieve propagation to a fixed range, and therefore the lower the BMUF. This means that although the critical frequency of the E layer is less than that of the F1 layer, which in turn is less than that of the F2 layer, sometimes the E-BMUF can be the greatest of the three separate layer BMUFs. This is most likely to be the case in the summer daytime at low solar epochs (when the ratio of E to F2 critical frequencies is greatest) over path ranges of 1000–2000 km.

Since the earth's field leads to the production of *O* and *X* waves which follow different ray paths, these waves also have differing BMUFs. The *O* wave is refracted less than the *X* wave, becomes reflected from a greater height and so has a lower critical frequency and BMUF. For propagation between a pair of fixed terminals the path BMUF is the greatest of the individual BMUFs for reflection from the different layers. This frequency undergoes systematic variations with the time of day, season and solar epoch as the electron density and layer heights vary; there are also large day-to-day changes that create problems for modelling.

Now consider propagation to some point beyond the skip distance. Figure 16.4 shows that, as elevation angle is increased, at a fixed frequency, rays travel to shorter ground ranges until the skip distance is reached. Rays of slightly larger elevation angle do not penetrate the ionosphere into space because, contrary to a popular misconception, ray apogee at the BMUF is below the height of maximum electron density, except in the limiting case of vertical incidence. These larger elevation rays are reflected from a greater height, and they travel back to the ground at increased range by virtue of having a significant length of near-horizontal path close to apogee. In principle, such so-called high-angle or Pedersen rays can exist out to a limiting ground range where ionospheric reflection is from the layer maximum. This limiting range can exceed that of the low-angle ray and may well be in excess of 7500 km in temperate regions and 10 000 km in equatorial regions [17]. The band of elevation angles providing high-angle rays is usually only a few degrees. There is thus a range of ground distances along which there are both low- and high-angle rays. The path length through the ionosphere of the high-angle ray exceeds that of the low-angle ray by an amount which increases when moving out from the skip distance. Therefore, the strength of the high-angle ray tends to be less than that of the low-angle ray both because of increased spatial attenuation and also, particularly in the case of reflection from the E layer, because of increased ionospheric absorption.

The presence of two rays with different group-path lengths is a disadvantage for it gives rise to signal distortions. Since the low- and high-angle rays merge at the BMUF, this frequency is sometimes known as the junction frequency (JF). Both the *O* and *X* waves have their own separate families of high-angle rays and associated JFs. Figure 16.5 shows an oblique-incidence ionogram recorded over a path from Norway to the UK where the propagation time is displayed as a function of wave frequency. The separate traces are associated with signals successively reflected once, twice and three times from the F2 region, being

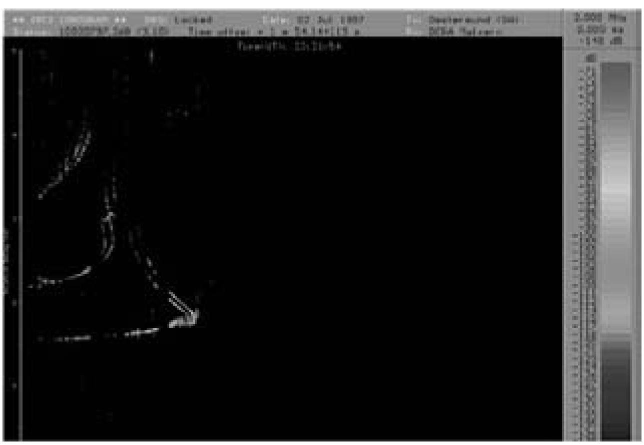


Figure 16.5 Sample oblique-incidence ionogram showing multihop propagation and O/X

sustained by intermediate ground reflections. The corresponding junction frequencies, together with the high-angle rays, can be seen.

Aside from signal-strength considerations, for a particular mode to be present the wave frequency must be below the BMUF and, for F modes, the lower ionosphere must not screen or blanket it. Screening of the 1F2 mode, but not of the 2F2 mode because of the lesser path obliquity, is a common summer daytime occurrence at certain frequencies. The strongest or dominant mode on a long path is usually the lowest possible order F2 mode unless the antennas discriminate against this. Higher-order F2 modes traverse the ionosphere a greater number of times and also experience more ground reflections, so that they tend to be weaker. Fewer F than E hops can span a given range. Modes involving more than two reflections from the E layer are rarely of importance. Reflections from the F1 layer arise only under restricted conditions and the 1F1 mode is less common than the 1E and 1F2 modes. The 1F1 mode can be important at ranges of 2000–3400 km, particularly at high latitudes. Multiple-hop F1 modes are very rare in practice because the necessary ionospheric conditions to support an F1 layer reflection do not occur simultaneously at separated positions.

Geographical changes in ionisation cause so-called mixed modes with successive reflection from different layers. Mixed modes are a common feature of transequatorial paths and east–west paths across a daylight–darkness boundary. Other more complex examples of mixed modes are those involving upward reflection from the E layer between two F reflections, known as *M* modes. Changes in ionisation of a smaller-scale size influence ray paths on single hops. These are variously referred to as ionisation gradients, horizontal gradients or ionospheric tilts. They cause the upward and downward legs of a hop to differ in length and direction.

Longitudinal tilts produce differences in the elevation angles on the two legs; lateral tilts create off-great-circle paths. Longitudinal tilts are usually the more important because they can give rise to changes in propagation modes. Lateral deviations are generally small in comparison with antenna beamwidths. An exception occurs when the transmitter and receiver are almost antipodal where ionospheric tilts lead to marked departures from the great-circle path. Simultaneous propagation may then take place in several directions and the dominant mode direction may vary with time of day, season and frequency. An effective tilt may result from geographical changes in either electron density or layer height. It follows that longitudinal tilts modify the BMUF over a fixed path length. On long paths with low elevation angles these longitudinal tilts can give rise to modes involving multiple reflection from the ionosphere without intermediate ground reflection. In such cases, if ray perigee at the middle of the path is within the ionosphere and above the D and lower E regions, there is little resulting absorption so that received signals are relatively strong. These so-called perigee modes can be particularly important across the equator and at high latitudes where significant ionisation gradients commonly exist. Associated with perigee modes are ground dead zones, additional to the skip zone, for which ray path illumination is not possible.

As well as propagation modes resulting from ionospheric reflections, there are others associated with scattering and ducting. Various mechanisms are believed to be involved, and so it is not surprising that there are uncertainties in the interpretation of particular observational data and therefore in assessing the relative importance of the different phenomena. These create modelling difficulties. Signals are scattered by ionospheric irregularities in the D, E and F regions – patches of varying electron density such as those that give rise to the phenomena observed on vertical-incidence ionograms known as sporadic E (Es) and spread F. The scattering may result in onwards propagation (forward scatter), deviation out of the great circle (sidescatter) or return along the same path (backscatter). Ionospheric scatter modes are usually weaker than the corresponding reflected modes and they tend to fade more. However, they are important at the higher frequencies of the HF band since they enhance the practical (operational) MUF (referred to simply as the MUF [16]) so that it exceeds the basic MUF. The geographical and temporal occurrence is governed by the incidence of the irregularities. Sporadic E is most prevalent at low latitudes in the daytime and at auroral latitudes by night. It tends to be opaque to the lower HF waves and partially reflecting at the higher frequencies. F region irregularities can exist simultaneously over a wide range of heights. They are found at all latitudes, but are particularly common at low latitudes in the evenings where their occurrence is related to rapid changes in the height of the F region. Forward scatter modes associated with the spread F are important on long trans-equatorial paths. F region irregularities are field aligned and sidescatter from these has been observed on paths at high and low latitudes; in some instances the received signals were incident simultaneously from a range of directions.

Normal ground terrain is sufficiently rough that it too scatters significant signal power out of the great-circle direction. Ground sidescatter and backscatter result. Since sidescatter paths are longer than the more direct routes, they tend to have correspondingly greater MUFs. There is some practical evidence supporting a dependence of signal intensity on scattering angle and whether sea or land is involved. The backscatter mechanism is of value in providing a means of remote probing (e.g. studying the state of the sea) or for monitoring ionospheric conditions. Special backscatter sounders can be used to determine the skip distance and are sometimes deployed in support of systems operation.

It is believed that another mechanism for wave propagation in the ionosphere concerns channelling as in a waveguide. This waveguide may be formed within the F layer and has an upper but no lower boundary, being sustained by the concave ionosphere, or it may be a double-walled duct in the electron-density minimum between the E and F regions. The waveguide is sometimes known as the whispering gallery. Signal coupling into the waveguide is assumed to involve ionospheric tilts like those which develop in the twilight periods or to be caused by the existence of ionisation irregularities such as Es or those responsible for spread F.

Mention has been made of ionospheric absorption. For propagation along the direction of the earth's magnetic field (and approximately over a wide range of wave directions) the absorption in decibels $L(f_v)$ at vertical incidence, in traversing a height region h at a wave frequency f_v, is given by:

$$L(f_v) = K \int_h \frac{N\nu}{\mu} \frac{dh}{(f_v \pm f_L) + \dfrac{\nu^2}{4\pi^2}} \tag{16.25}$$

where K is a constant of proportionality and f_L is the electron gyrofrequency about the component of the earth's magnetic field along the direction of propagation. The positive sign applies for the O wave and the negative sign for the X wave. For ground-based reflection the limits of integration are from the base of the ionosphere to the height of wave reflection. For propagation at oblique incidence the absorption is proportionately increased because of the greater lengths of path traversed. Inspection of Eqn. 16.25 shows that:

- the absorption in a given slab of ionosphere is proportional to the product of electron density and collision frequency: electron density increases with increase of height whereas the collision frequency for electrons, which is proportional to the atmospheric pressure, decreases; hence, the absorption reaches a maximum in the lower E region with most of the contribution to the total absorption occurring in the D region;
- large amounts of additional deviative absorption arise near the height of reflection where μ is small;
- absorption decreases with increase of frequency;
- the O wave absorption is less than that of the X wave and differences are accentuated the lower the frequency, provided that the first term of the denominator of Eqn. 16.25 remains dominant.

The absorption is low at night time because of the reduced D and E region ionisation. The nondeviative absorption reaches a maximum around local noon in the summer, but the influence of deviative absorption can modify the resultant seasonal variation. Ionospheric absorption is one of the most important factors influencing received sky-wave signal strengths at MF and HF so that accurate methods of modelling it are needed. There are particular difficulties at MF because ray-path reflection heights of around 85–90 km are common and much of the absorption is deviative occurring within 2–3 km of ray apogee. Such electron density data as exist for these heights display considerable irregular variations.

16.6 Fading and Doppler effects

If the ionosphere were unchanging, the signal amplitude over a fixed path would be constant. In practice, however, fading arises as a consequence of variations in propagation path, brought about by transmitter or receiver movement or fluctuations in ionisation which in turn result in changes of the path refractive index. The principal causes of fading are:

- variations in absorption
- changes of path length
- changes of polarisation, such as for example due to Faraday rotation
- movements of small scale irregularities to produce scintillation.

These various causes lead to different depths of fading and a range of fading rates. The slowest fades are usually those due to absorption changes that have a period of about ten minutes. The deepest and most rapid fading occurs from the beating between two signal components of comparable amplitude propagated along different paths. A regularly reflected signal together with a signal scattered from spread F irregularities can give rise to so-called flutter fading, with fading rates of about 10 Hz.

Amplitude fading is accompanied by associated fluctuations in group path and phase path, giving rise to time- and frequency-dispersed signals. Signals propagated simultaneously *via* different ionospheric paths are usually received with differing frequency shifts. Frequency shifts and spreads for reflections from the regular layers are usually less than 1 Hz, but shifts and spreads of up to 20 Hz have been reported for scatter-mode signals at low and high latitudes.

16.7 Modem requirements on SNR, Doppler and multipath

Any practical modem can only work within a specified range of SNR, multipath and Doppler. Under conditions of no multipath and Doppler the operating SNR, in additive Gaussian noise, is given by well known equations [18] for a specified symbol rate and modulation scheme. When multipath and Doppler are present they also constrain the achievable data rate *via* their impact on the channel coherency bandwidth and time, respectively. The ideal communications channel exhibits slow Doppler induced fading and has a channel frequency spectrum that does not, as a result of multipath, vary across the operating bandwidth. Under these dual conditions intersymbol interference (ISI) is minimised. The HF communications channel is generally one with time-variant frequency-selective fading. Watterson, Angling *et al.*, Mastrangelo *et al.* and others [19–21] have modelled it with a wide sense stationary uncorrelated scattering channel model [22]. In this model discrete signals are modelled each with an individual delay spread and Doppler spread.

For fading to be considered slow it is necessary that $T_s \ll t_0$, where T_s is the transmitted symbol period and t_0 is the coherency time. As such, the signal characteristics will remain stable within a symbol period permitting demodulation of the signal. Equivalently $W \gg f_d$, where W is the signal bandwidth and f_d is the Doppler bandwidth. HF communication systems generally operate with signal rates between a few tens of Hz and a few thousand Hz. Doppler spreads have been measured throughout the world. Figure 16.6 shows the channel

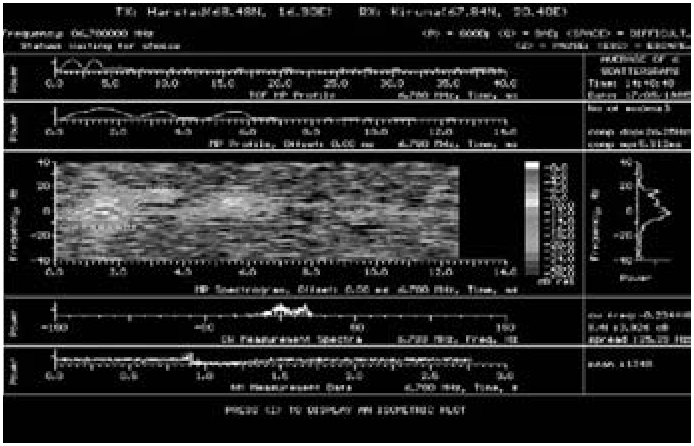


Figure 16.6 Doppler and multipath characteristic for a high-latitude path from Harstad to Kiruna; the centre panel shows the channel scattering function

scattering function for an auroral path where Doppler spreads of several Hertz to several tens of Hertz are often measured. Table 16.2 provides typical values of both f_d and W and evaluates the inequality given above.

Similar considerations apply to coherency bandwidth (Table 16.3) which is controlled by the multipath spread. In order for the channel to introduce negligible

Table 16.2 Considerations of coherency time – a tick indicates where the inequality is met

	Typical Doppler bandwidth (f_d)		
Signal	*mid-latitudes*	*high/low latitudes*	
bandwith (W)	*0.1 Hz*	*5 Hz*	*20 Hz*
50 Hz	√	√	x
300 Hz	√	√	√
3000 Hz	√	√	√

Table 16.3 Considerations of coherency bandwidth – a tick indicates where the inequality is met

	Typical multipath spread (σ_t)		
Signal	*0 ms*	*1 ms*	*3 ms*
bandwith (W)	$f_0 = \infty$ Hz	*160 Hz*	*53 Hz*
50 Hz	√	√	x
300 Hz	√	x	x
3000 Hz	√	x	x

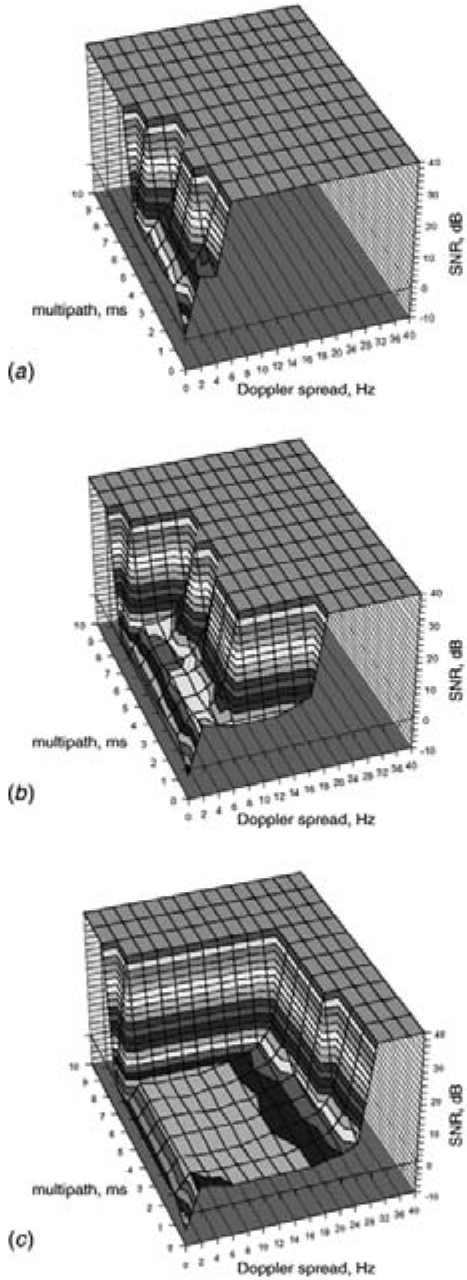


Figure 16.7 Performance surfaces of Mil Std 188–110A modems for data rates of 2400, 1200 and 300 bit/s

(a) Mil-Std-188-110A serial, short interleaver, 2400bps
(b) Mil-Std-188-110A serial, short interleaver, 1200bps
(c) Mil-Std-188-110A serial, short interleaver, 300bps

channel distortion it is necessary for $f_0 >> W$, where f_0 is the channel coherency bandwidth. In the ionospheric community we define a root mean squared (RMS) delay spread, σ_t and write $f_0 = 1/2\pi\sigma_t$. Equivalently, we therefore require $1/2\pi\sigma_t >> W$.

Table 16.3 shows that it is difficult to achieve flat fading at anything more than moderate symbol rates and only low rate modems (50 to 100 baud) can operate without significant ISI. Conversely, Table 16.2 has demonstrated that these low rate modems are susceptible to Doppler effects at high and low latitudes.

Historically, the issues of coherency bandwidth and coherency time limited reliable HF digital data transmissions to mid-latitudes and even then at low symbol rates. However, for many applications these rates were simply unacceptable. In order to overcome the coherency bandwidth problem multicarrier modems using a number of orthogonal carriers within the operating bandwidth were developed, thereby keeping the symbol rate low. Adaptive channel equalisers have also been developed.

Figure 16.7 shows the performance of a standard single carrier modem operating at different bit rates, where the surface is that combination of SNR, multipath and Doppler spread required to provide a BER of 1 in 10^3 when the Doppler shift is zero Hz. The improved operating range at lower data rates is clear.

16.8 References

1 GOODMAN, J.M., and AARONS, J.: 'Ionospheric effects on modern electronic systems,' *Proc. IEEE*, pp. 512–528, **78**, 1990.

2 GOTT, G.F., CHAN, S.K., PANTJIAROS, C.A., and LAYCOCK, P.J.: 'High frequency spectral occupancy at the solstices', *IEE Proc. Commun.* 1997, **144**

3 KNEPP, D.L., and HOUPIS, H.L.F.: 'Altair VHF/UHF observations of multipath and backscatter enhancement', *IEEE Trans. Antennas Propag.*, 1991, **39**, 528–534

4 BISHOP, G.J., MAZZELLA, A.J., HOLLAND, A., and RAO, S.: 'An overview of ionospheric effects and mitigation in RF communication, navigation and surveillance', *Ionospheric effects* symposium, Alexandria, USA, 1996 (JMG Associates Ltd, Alexandria, VA.)

5 CANNON, P.S.: 'Morphology of the high latitude ionosphere and its implications for HF communications systems,' *IEE Proc.* I, 1989, **136**, pp.1–10

6 NcNAMARA, L.F.: 'The ionosphere, communications, surveillance and direction finding' (Krieger Publishing Company, Malabar, Florida; 1991)

7 JONES, R.M., and STEPHENSON, J.J.: 'A three-dimensional ray tracing computer program for radio waves in the ionosphere'. US Dept. of Commerce, Office of Telecommunications OT report 75–76, 1975

8 CROFT, T.A. and HOOGASIAN, H.: 'Exact ray calculations in a quasi-parabolic ionosphere', *Radio Sci.* 1968, **3**, pp.69–74

9 MILSOM, J.D.: 'Exact ray path calculations in a modified Bradley-

Dudeney model ionosphere', *IEE Proc. H, Microw. Antennas Propag.* 1985, *132*, 33–38
10 DYSON, P. and BENNETT, J.A.: 'A model of the vertical distribution of the electron concentration in the Ionosphere and its application to oblique propagation studies', *J. Atmos. Terr. Phys.* 1980, **50**, pp.251–262
11 BAKER, D.C., and LAMBERT, S.: 'Range estimation for SSL HFDF by means of a multi-quasi-parabolic ionospheric model', *IEE Proc. H*, 1989, **136**, pp.120–125
12 NORMAN, R.J., and CANNON, P.S.: 'An evaluation of a new 2-D analytic ionospheric ray tracing technique – SMART', *Radio Sci*. 1999, **34**, 489–499
13 NORMAN, R.J., and CANNON, P.S.: Cannon, 'A two-dimensional analytic ray tracing technique accommodating horizontal gradients', *Radio Sci*., 1997, **32**, pp.387–396
14 BENNETT, J.A., CHEN, J., and DYSON, P.L.: 'Analytic ray tracing for the study of HF magneto-ionic radio propagation in the ionosphere', *Appl. Comput. Electromagn. Soc. J.*, 1991, **6**, pp.192–210
15 GOODMAN, J.M.: 'HF communications science and technology' (Van Nostrand Reinhold, 1992)
16 ITU-R Recommendation P.373–7: 'Definitions of maximum and minimum transmission frequencies'. International Telecommunication Union, Geneva, 1995
17 CCIR Report 250–6: 'Long-distance ionospheric propagation without intermediate ground reflection', International Telecommunication Union, Geneva, 1990
18 PROAKIS, J.G.: 'Digital communications' (McGraw-Hill, 1989, 2nd edn.)
19 WATTERSON, C.C., JUROSHEK, J.J., and BENSEMA, W.D.: 'Experimental confirmation of an HF channel model', *IEEE Trans. Commun. Technol.* 1970, **18**, pp.792–803
20 ANGLING, M.J., CANNON, P.C., DAVIES, N.C., WILLINK, T.J., JODALEN, V., and LUNDBORG, B.: 'Measurements of doppler and multipath spread on oblique high-latitude HF paths and their use in characterising data modem performance', *Radio Sci.* 1998, **33**, pp.97–107
21 MASTRANGELO, J.F., LEMMON, J.E., VOGLER, L.E., HOFFMEYER, J.A., PRATT, L.E., and BEHM, C.J.: 'A new wideband high frequency channel simulation system', *IEEE Trans. Commun.* 1997, **45**, pp.26–34
22 BELLO, P.A.: 'Characterization of randomly time variant linear channels', *IEEE Trans. Commun. Systems*, 1963, **11**, pp.360–393

Chapter 17

Ionospheric prediction methods and models

Paul S. Cannon

17.1 Introduction

The ionosphere, the ionised region of the atmosphere between about 60 km and 1000 km, affects all radio signals below ~5 GHz which pass through or travel *via* it. Some of the associated radio systems can only operate because of the ionosphere, but it can also degrade radio system operation. Sometimes the effects of the ionosphere are highly significant, sometimes they can all but be ignored. For high frequency (HF) sky-wave propagation the ionosphere is, of course, a prerequisite.

If the ionosphere were stable in time and constant in space it would be relatively easy to determine the effect of the ionisation on the radio propagation and hence on the radio system. Unfortunately, stability is not the norm, particularly in the high-latitude regions [1], and as a consequence ionospheric prediction methods and models are required to support system design, service planning and frequency management. These methods and models characterise the medium and estimate the system performance. Some operate in real, or close to real, time and they may even directly advise the radio operator on a course of action which would improve system performance.

Ionospheric prediction methods can be considered to incorporate one or more of a number of elements:

- an ionospheric model
- a ray tracing method
- system factors including antennas, noise etc.
- a calculation of the parameter(s) of interest, e.g. SNR

17.1.1 System design and service planning

A prediction method containing all four elements noted above is often used for long-term prediction of system performance. In such cases it is usual to provide monthly median estimates of system performance at specific hours. In HF models ray-path and launch angle information can be obtained from the model for antenna optimisation. The main use of such models is, however, the calculation of signal field strength and signal-to-noise ratio (SNR) at a number of frequencies. This is usually accompanied by a statistical quantification of day-to-day variability and the probability that an acceptable quality-of-service (QOS) will be met. At higher frequencies the parameter of interest might well be the signal group delay to support the development of spaceborne navigation systems or it might be scintillation depth to support the development of UHF satellite communications modems.

17.1.2 Short-term adaptation to the propagation environment

Frequency management involves real-time adaptation of the operating frequency to optimise system operation. In many modern point-to-point HF systems this is achieved using automatic link establishment (ALE) systems such as MilStd 188–141A [2] but this is not possible for broadcast systems. Further, ALE takes no account of interference, both within a network and on other users. Therefore, some HF users need a real-time specification of the ionosphere or system performance or even a forecast of the likely situation. Likewise, a number of UHF systems require an up-to-date assessment of the ionosphere. A good example of this is single-frequency operation of the global positioning system (GPS) when high accuracy is required. Short-term adaptation is, of course, far more difficult to achieve than the median prediction since it requires up-to-date ionospheric measurements and strategies for using these data.

17.2. Ionospheric models

17.2.1 Introduction

The most difficult aspect of propagation modelling is the generation of the ionospheric model of the vertical distribution of electron concentration in the E and F regions. The efficacy of a prediction or forecasting model is no better than the knowledge of the ionospheric electron density profile. Fortunately, there have been substantial improvements in ionospheric modelling capabilities over the past few years, both as a result of improved computing facilities and as a result of improved ground-based sensors which can be used to provide real-time updates.

Bilitza [3,4] provides excellent summaries of the various ionospheric models. These cover electron density, temperature, ion composition, ionospheric electric field, auroral precipitation and conductivity models. We restrict our discussions to electron density models only.

Both statistically-based and physically-based electron density models exist and these have been reviewed by References [5] and [6]. Unfortunately, the computer run times for the physical models are high; typically, several hours are needed on a supercomputer to specify the global electron density. As a consequence of these long run times the physical models are not appropriate for real-time or even near real-time applications.

17.2.2 Empirical models

17.2.2.1 Electron density profile models

Empirical models of the electron density height profile currently enjoy the greatest operational use. Empirical models are essentially climatological models for specified solar and/or magnetic activity levels, season, time of day, geographical area etc. Such models are based upon a statistical database of ionospheric characteristics derived from a world network of ionosondes. Scaled anchor points are used to generate a stylised model of the ionosphere based upon a straight-line, parabolic or polynomial function. Typical anchor points derived from the ionospheric data base include *fo*F2, *fo*F1, *fo*E and *M*(3000). (*fo*F2 corresponds to the highest ordinary polarisation signal returned from the ionospheric F2 layer at vertical incidence. *fo*F1 and *fo*E are similar parameters appropriate to the F1 and E regions. *M*(3000) = *MUF*(3000)/*fo*F2 provides a measure of the F region peak altitude.)

Figure 17.1 illustrates a stylised electron density profile as a function of height

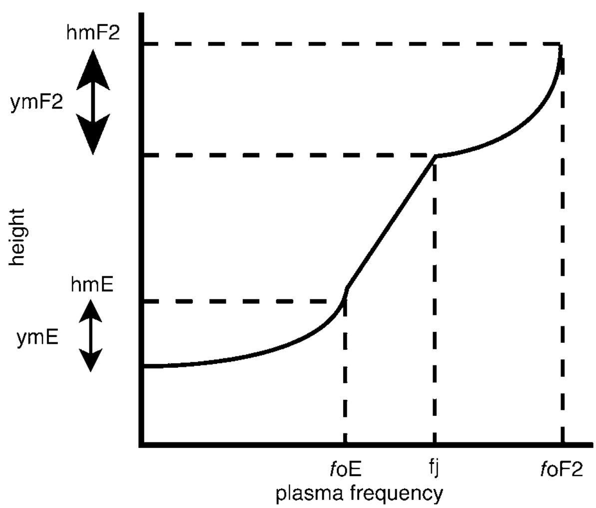


Figure 17.1 Bradley–Dudeney approximate model of the ionospheric electron density profile

[7] adopted as a component of the ITU-R recommended model. This model cannot be used for numerical ray tracing because of the gradient discontinuities. To overcome this problem the Dudeney model [8] was introduced. The international reference ionosphere (IRI) model [9] provides a further level of detail in the electron density profile. The value of a number of empirical models for radio communications purpose is described by Reference [10].

17.2.2.2 ITU database of ionospheric coefficients

The most commonly used numerical representation of the ionosphere is the ITU atlas of ionospheric characteristics [11]. Values from this database are used in all of the models described above. The atlas gives maps of the F2 layer peak critical frequency, *fo*F2 and the $M(3000)$ factor.

The E and F1 layer characteristics are given by a series of formulae that assume a variation with latitude, time of day and season depending on the solar zenith angle. The contours of *fo*F2 are, however, based on spherical harmonic and Fourier functions for every month of two reference years with an assumed linear dependence on R_{12} (the smoothed sunspot number) for intermediate solar epochs. The whole model consists of 34 296 coefficients. Figure 17.2 shows an example of the *fo*F2 contours for December, 1900 UT, with a sunspot number of 20. It is important to note that the model lacks detail at high latitudes.

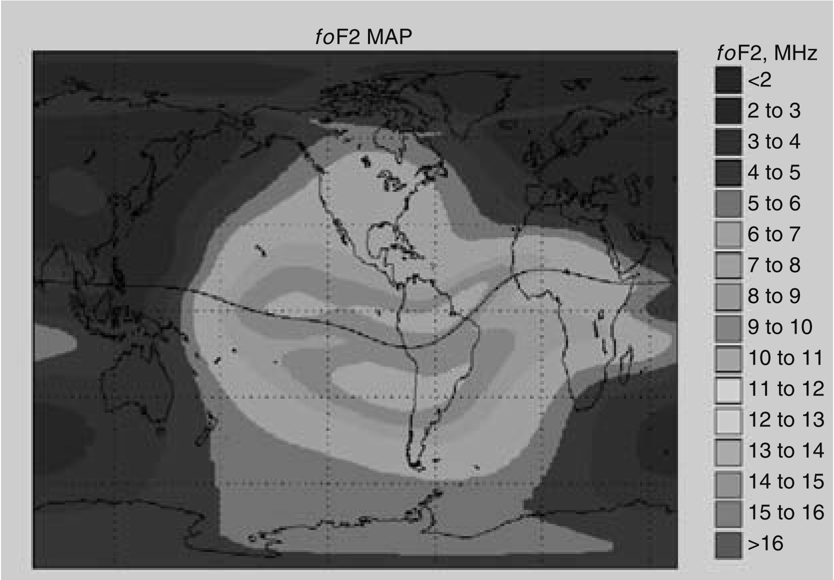


Figure 17.2 foF2 contours, December, 1900 UT, SSN = 20

17.2.2.3 URSI database of ionospheric coefficients

It has long been known that the ITU model is inaccurate above oceans and in the southern hemisphere where ionosonde measurements are scarce or do not exist. Rush *et al.* [12,13] suggested a framework for introducing theoretical values in regions of no measurements and this led to a new model which was adopted by the International Union of Radio Science (URSI) [14].

17.2.3 Physical ionospheric models

Physical models are based on self-consistent theoretical simulations of the ionosphere *via* a solution of the coupled continuity, energy and momentum equations for electrons and ions. Typical input parameters required to drive the model are the solar EUV radiation, the auroral particle precipitation (together with the interaction cross sections) and the atmospheric and magnetospheric boundary conditions. From those inputs and equations the plasma densities, temperatures and drifts are obtained numerically from the nonlinear coupled system of equations. The value of physical models is that they can be used to understand the complex processes which lead to the changes in the ionosphere. For instance the coupled thermosphere ionosphere models [15,16] have been used extensively to untangle the complex response of the ionosphere to geomagnetic storms [17,18].

17.2.4 Parameterised models

The statistical (or climatological) models and the physical models each have advantages and disadvantages for operational use. The accuracy of the former is limited by averaging and the rules adopted for spatial extrapolation. The second is more intellectually appealing but the computational demands are high and the model drivers are in reality not sufficiently well known.

A third approach is the parameterised version of the physical model. The parameterised model begins with a physical model which gives a more realistic representation of the spatial structure of the ionosphere than the statistical model can provide. (Any statistical database inevitably averages over similar geophysical conditions resulting in a smoothed model which is unrepresentative of the instantaneous ionosphere.) The physical model is then parameterised in terms of solar and geographical parameters to provide an easy to use and quick to run ionospheric model.

17.2.4.1 PIM

The PIM (parameterised ionospheric model) [19] uses either the *fo*F2 ITU coefficients for normalisation of the electron density profiles or it uses coefficients produced by a Utah State University model. The calculation of the electron density profile is based upon the underlying neutral composition, temperature and wind together with electric field distributions, auroral precipitation and the solar EUV spectrum. Inputs to the model consist of universal time and the daily value of $F_{10.7}$ (the solar flux at a wavelength of 10.7 cm) or the solar sunspot number. The current value of *Kp* (the index of geomagnetic activity) is also required together with the sign (positive or negative) of the interplanetary magnetic field *y* component. This latter factor is used to select one, of two, solar-wind-driven transport (convection) patterns in the high latitude region. The PIM model is at its most powerful, however, when it is adjusted by real time input data.

17.3 Prediction methods

17.3.1 HF

Prediction method development has largely followed the evolution of the ionospheric model and more recently the development of analytic ray tracing. HF propagation methods are in fact not new, but date back to World War II. Through the years a large number of methods have been developed and are still being developed. Many are, however, not generally available and, of course, a number have fallen into disuse because of the availability of the newer models. As a consequence this Chapter will restrict itself to a small number of the most popular and accessible models:

IONCAP [20]: this very popular program developed by the Institute for Telecommunications Sciences (ITS) in the USA will run on a variety of computer platforms. A number of user friendly front ends have also been developed, for example the derivative VOACAP [21] which also makes a number of small changes to the calculations used in IONCAP. IONCAP and its derivatives use virtual mirror techniques.

ITU REC533 [22]: this is the method recommended by the International Telecommunication Union. It is a relatively simple technique but tests show that its accuracy is comparable to that provided by more complex methods.

ICEPAC [23]: the ionospheric communications enhanced profile analysis and circuit (ICEPAC) prediction method is the latest program from ITS. It uses the ICED electron density model which should give improved performance in the high latitude regions.

17.3.1.1 ICEPAC

ICEPAC represents one of the newest of the HF prediction decision aids and as such it is worthy of more detailed discussion. Its outputs are very similar to IONCAP and it provides in total 29 different output options. The BMUF is assumed to follow a given distribution so that it is possible to determine, for each wave frequency and mode, the fraction of days for which that mode exists – this is known as the availability. Reference values also exist for the upper and lower standard deviations of the day-to-day signal variability to estimate the probability that the signal strength will exceed a prescribed threshold.

Ionospheric descriptions, methods 1 or 2
Methods 1 and 2 provide specialist output relating to the ionospheric parameters.

MUF output options, methods 3 to 12 and 25 to 29
These methods include listings and graphical output of the basic maximum usable frequency (MUF) and frequency of optimum transmission (FOT) for the path and for individual modes E, Es, F1 and F2. They also provide information on angle of take off. Table 17.1 illustrates typical method 10 output for a path from the southern UK to Ottawa, Canada, for January 1994.

Table 17.1 ICEPAC method 10 output

```
CCIR COEFFICIENTS          METHOD 10   ICEPAC  Version ITS.01

  JAN    1994          SSN =  62.      Qeff= 3.0
  Cobbett Hill, UK    CRC, Ottawa, Canada   AZIMUTHS          N. MI.      KM
  51.27 N     .63 E - 45.40 N   75.92 W    294.97   53.87    2934.6   5434.4
  MINIMUM ANGLE   3.00  DEGREES


  XMTR  2-30 CCIR.000   ISOTROPE +    .0 dBi  Azim=295.0 OFFaz=360.0     .500kW
  RCVR  2-30 CCIR.000   CCIR  /ISOTROPE       Azim=    .0 OFFaz= 53.9


           MUF(....)        FOT(XXXX)       ANG(++++)

   00  02  04  06  08  10  12  14  16  18  20  22  00
 MHZ+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+MHZ
40-                                                   -40
  -                                                   -
38-                                                   -38
  -                                                   -
36-                                                   -36  GMT   MUF   FOT   ANG
  -                                                   -
34-                                                   -34   1.0   9.0   7.5   9.6
  -                                                   -     2.0   8.7   7.3   9.7
32-                                                   -32   3.0   8.9   7.3   9.7
  -                                                   -     4.0   9.1   7.5   9.7
30-                                                   -30   5.0   8.8   7.2   9.6
  -                                                   -     6.0   7.4   6.1   9.3
28-                                                   -28   7.0   7.2   5.9   8.4
  -                                                   -     8.0   7.9   6.6   8.9
26-                               .                   -26   9.0   9.0   7.1   7.9
  -                             .   .                 -    10.0  10.7   8.4   7.1
24-                                                   -24  11.0  13.7  10.8   6.3
  -                                   .               -    12.0  18.1  14.3   5.7
22-                           .   X                   -22  13.0  22.4  16.8   5.6
  -                                 X                 -    14.0  25.2  18.9   5.6
20-                                                   -20  15.0  26.0  21.8   5.7
  -                             X     X .             -    16.0  24.6  20.7   6.4
18-                         .                         -18  17.0  22.7  19.1   6.5
  -                           X                       -    18.0  19.3  16.2   7.0
16-                                       X .         -16  19.0  15.9  12.4   7.3
  -                                                   -    20.0  13.2  10.3   8.1
14-                       . X                         -14  21.0  11.6   9.1   8.5
  -                                         .         -    22.0  10.5   8.2   8.8
12-                                       X   .       -12  23.0   9.6   8.1   9.5
  -                     . X                           -    24.0   9.2   7.8   9.7
10- + + + + + +                             X   . . + -10
  - . . . . . . +   + .                       X + + . -
08- X X           + . + X                   + + X X X -08
  -     X X X X . . X X +               + +           -
06-             X X       + + + + + + +               -06
  -                                                   -
04-                                                   -04
  -                                                   -
02-                                                   -02
  -                                                   -
 MHZ+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+MHZ
   00  02  04  06  08  10  12  14  16  18  20  22  00
                 UNIVERSAL TIME
```

System performance options, methods 16 to 24

The system performance outputs are largely tabular and provide system performance data. Method 16 (Table 17.2) illustrates typical output from ICEPAC for the path from southern England to Ottawa, Canada. Listed at each time (UT) and frequency (FREQ) are the most reliable propagating modes (MODE), arrival angles (ANGLE), the path delay (DELAY), the virtual height (V HITE), the probability that the operating frequency will be below the predicted MUF

Table 17.2 ICEPAC method 16 output

```
  CCIR COEFFICIENTS         METHOD 16   ICEPAC  Version ITS.01  PAGE   1

JAN    1994           SSN =  62.      Qeff= 3.0
Cobbett Hill, UK    CRC, Ottawa, Canada    AZIMUTHS          N. MI.      KM
51.27 N      .63 E - 45.40 N   75.92 W    294.97   53.87    2934.6   5434.4
MINIMUM ANGLE   3.00  DEGREES
XMTR  2-30 IONCAP     Const 10dB           Azim=295.0 OFFaz=360.0  10.000kW
RCVR  2-30 CCIR.000   CCIR  /ISOTROPE      Azim=    .0 OFFaz= 53.9
3 MHZ NOISE = -136.0 DBW     REQ. REL = .90     REQ. SNR = 44.0 DB
MULTIPATH POWER TOLERANCE = 10.0 DB   MULTIPATH DELAY TOLERANCE =   .850 MS
```

14.0	25.2	2.0	4.0	6.0	7.0	9.0	11.0	13.0	15.0	17.0	19.0	21.0	FREQ
	2F2	5 E	4 E	4 E	4F2	3F2	3F2	2F2	2F2	2F2	2F2	2F2	MODE
	5.6	5.0	3.2	3.7	19.4	13.2	11.1	4.8	4.0	4.0	4.1	4.3	ANGLE
	18.9	18.4	18.3	18.4	20.0	19.4	19.1	18.8	18.7	18.7	18.7	18.8	DELAY
	290	71	75	81	287	288	251	269	248	248	251	254	V HITE
	.50	1.00	1.00	1.00	1.00	1.00	1.00	.99	.98	.95	.90	.80	F DAYS
	162	291	241	214	177	161	153	147	144	143	144	146	LOSS
	13	-138	-82	-27	-10	8	17	25	28	29	29	28	DBU
	-122	-251	-201	-149	-134	-118	-111	-105	-103	-102	-104	-106	S DBW
	-162	-131	-139	-144	-146	-149	-152	-154	-155	-157	-158	-159	N DBW
	40	-120	-61	-5	12	31	41	49	53	54	55	53	SNR
	30	176	118	63	46	27	20	11	10	11	15	16	RPWRG
	.41	.00	.00	.00	.00	.07	.35	.66	.73	.73	.70	.68	REL
	.00	.00	.00	.00	.00	.00	.09	.00	.00	.00	.00	.00	MPROB
15.0	26.0	2.0	4.0	6.0	7.0	9.0	11.0	13.0	15.0	17.0	19.0	21.0	FREQ
	2F2	5 E	4 E	4 E	4F2	3F2	3F2	2F2	2F2	2F2	2F2	2F2	MODE
	5.7	5.0	3.2	3.9	18.3	11.9	10.9	4.0	4.0	4.0	4.1	4.3	ANGLE
	18.9	18.4	18.3	18.4	19.9	19.2	19.1	18.7	18.7	18.7	18.7	18.8	DELAY
	291	71	75	83	271	265	248	249	247	249	251	256	V HITE
	.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.98	.94	F DAYS
	162	301	250	222	180	163	154	147	144	143	142	142	LOSS
	14	-147	-91	-30	-13	6	16	25	29	31	32	32	DBU
	-122	-261	-210	-153	-137	-120	-112	-104	-102	-101	-101	-102	S DBW
	-162	-131	-139	-144	-146	-149	-152	-154	-155	-157	-158	-159	N DBW
	40	-130	-71	-9	9	29	40	49	53	56	57	57	SNR
	29	183	124	62	44	24	14	5	1	0	0	5	RPWRG
	.42	.00	.00	.00	.00	.05	.32	.75	.87	.91	.89	.82	REL
	.00	.00	.00	.00	.00	.00	.08	.38	.51	.00	.00	.00	MPROB
16.0	24.6	2.0	4.0	6.0	7.0	9.0	11.0	13.0	15.0	17.0	19.0	21.0	FREQ
	2F2	5 E	4 E	4F2	3F2	3F2	2F2	2F2	2F2	2F2	2F2	2F2	MODE
	6.4	5.0	3.4	18.3	13.4	11.0	4.0	3.9	4.0	4.0	4.2	4.6	ANGLE
	19.0	18.4	18.3	19.9	19.4	19.1	18.7	18.7	18.7	18.7	18.8	18.8	DELAY
	309	71	77	271	292	249	248	245	248	248	254	263	V HITE
	.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	.97	.88	F DAYS
	160	284	242	187	174	160	150	146	143	142	141	142	LOSS
	15	-130	-75	-22	-7	9	21	27	30	32	32	32	DBU
	-120	-243	-194	-144	-131	-117	-107	-103	-101	-100	-101	-102	S DBW
	-161	-131	-139	-144	-146	-149	-152	-154	-155	-157	-158	-159	N DBW
	41	-112	-55	0	15	32	44	51	55	57	58	57	SNR
	29	164	105	51	35	19	6	1	-3	-4	-2	5	RPWRG
	.44	.00	.00	.00	.00	.08	.52	.88	.96	.96	.94	.82	REL
	.00	.00	.00	.00	.00	.01	.21	.46	.51	.00	.00	.00	MPROB

(F DAYS), the median system loss in decibels (LOSS), field strength in dBμV m^{-1} (DBU), median signal and noise powers at the receiver input terminals in dBw (S DBW and N DBW), signal-to-noise ratio (SNR), the required combination of transmitter power and antenna gains needed to achieve the required reliability in decibels (RPWG), reliability, i.e. the probability that the SNR exceeds the required SNR (REL), and the probability of an additional mode within the multipath tolerances (MPROB).

Antenna output options, methods 13 to 15
ICEPAC includes the antenna gain description of a number of antennas that may be specified at the transmitter and receiver. These antennas include the isotropic reference radiator, dipoles, rhombics and sloping-V antennas. The radiation pattern of the antennas can be plotted and the effects of the antennas on the system performance evaluated *via* methods 16 to 24.

17.3.2 Computerised models of ionospheric scintillation

The regions of ionospheric scintillation consist of a belt 20° wide centred on the geomagnetic equator and in the polar regions above 55° to 60° latitude. Within the equatorial scintillation belt, Rayleigh distributed scintillation is often experienced at frequencies in the VHF and UHF bands during the late evening and early morning hours from about 2000 to 0200 local time (LT). In the polar regions the scintillation is more irregular and can occur any time of day and night although at lower levels than equatorial scintillation.

The climatology of scintillation in the equatorial regions is reasonably well established and long-term average scintillation conditions in particular geographic regions can be modelled using the WBMOD computer programme [25, 26]. This model also includes a climatological model of the high latitude; however this is less rigorous.

The electron density irregularities model, EDIM, within WBMOD is a collection of empirical models which describe the geometry, orientation, strength and motion of the irregularities as a function of latitude, date, time of day, solar activity and geomagnetic activity.

One of the basic parameters generated by the model is the height-integrated electron density irregularity strength, *CkL*, which is a measure of the vertical height integrated power in the irregularities. It is this parameter which is most dynamic and controls the worldwide scintillation.

The propagation model implemented in WBMOD is that described by Rino [27]. This model assumes that all propagation is line-of-sight, and that the scintillation effects can be calculated as though the effect of the ionospheric irregularities can be ascribed to an infinitely thin, phase-changing screen set at some altitude within the irregularity layer. A propagating ray passing through the screen will have its phase changed according to the spectral characteristics of the irregularities, and the resulting intensity and phase scintillation will develop as the ray propagates beyond the screen to the receiver. The irregularities are described in terms of their power-density spectrum (PDS) and their shape and orientation with respect to the local geomagnetic field direction.

Among other outputs WBMOD will predict the probability of scintillation occurring with a severity quantified by the S_4 index [28], Figure 17.3. S_4 is defined as the standard deviation of received power divided by the mean value of the received power (Eqn. 17.1).

$$S_4 = \sqrt{\frac{\langle I^2 \rangle - \langle I \rangle^2}{\langle I \rangle^2}} \tag{17.1}$$

where I = carrier intensity, and $\langle\rangle$ denotes the sample mean.

In the ionosphere S_4 varies from 0 to about 1.2. For a quiet nonfading channel, $S_4 = 0$, an $S_4 < 0.6$ denotes weak fading and an $S_4 \sim 1$ denotes a Rayleigh fading channel (severe fading). From an engineering perspective the fade depth is perhaps a more useful parameter than S_4 but conversion from one to the other requires knowledge of the fading cumulative distribution function (CDF). Chytil [29] has shown that:

$$S_4 = \left(\frac{1}{m}\right)^{\frac{1}{2}} \text{ for } m \geq 0.5 \tag{17.2}$$

where m is defined in the Nakagami m-distribution [30]. Consequently, evaluation of the m-distribution CDF for the appropriate value of S_4 gives the fade depth [31].

Aarons [32] details the equivalence of selected values of S_4 with the fade depth at the 99.9 per cent level and these are repeated in Table 17.3.

Table 17.3 Equivalence of S_4 and fading intensity

S_4	dB
0.075	1
0.17	3
0.3	6
0.45	10

Phase stability may be a more important parameter than amplitude stability, and to this end WBMOD will also provide an estimate of the standard deviation of phase, σ_ϕ given by:

$$\sigma_\phi = \sqrt{\langle \delta\phi^2 \rangle - \langle \delta\phi \rangle^2} \tag{17.3}$$

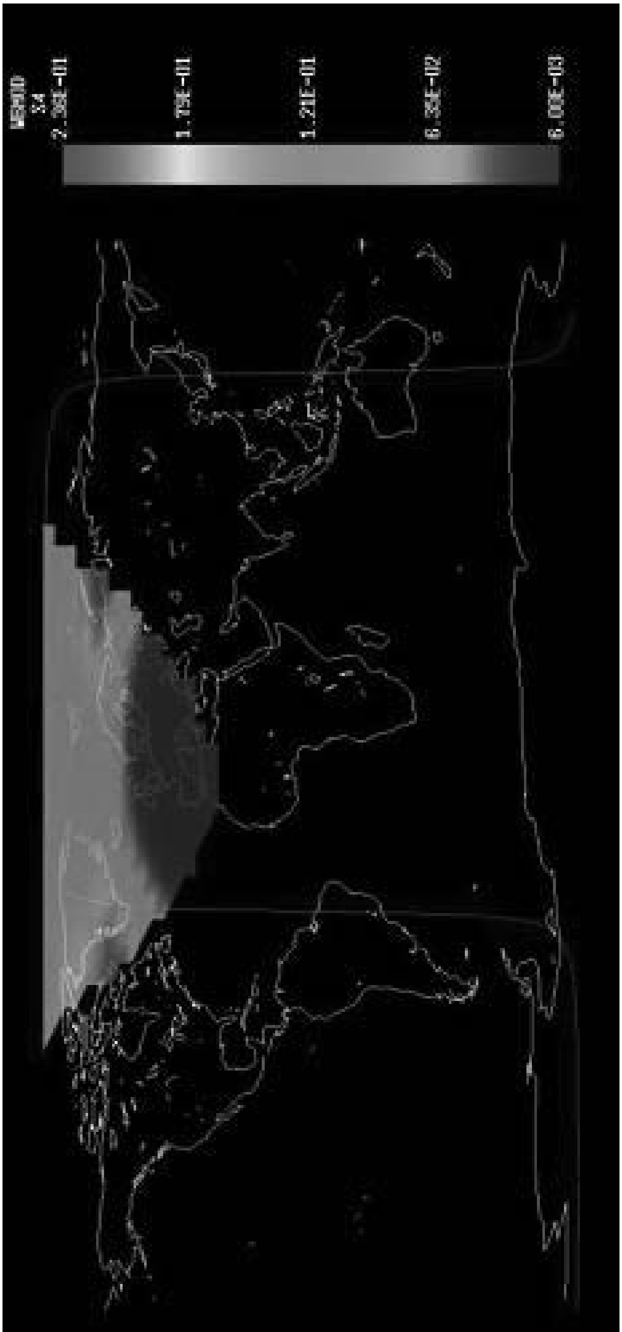


Figure 17.3 *WBMOD S_4 images for 1990, day 260, 21UT, Kp = 3, SSN = 137, freq = 1500 MHz, satellite altitude = 1000 km, integration period = 0.05s [33]*

17.4 Application specific models

The aforementioned propagation models exhibit a general architecture which makes them useful in many applications. A number of prediction methods have, however, been developed which take these generalised architectures and develop them for specific applications.

In military HF, low probability of interception/jamming (LPI/LPJ) is important and is generally based on techniques such as spread spectrum, cryptography and antenna nulling. Even with such sophisticated strategies the signals can still propagate to unwanted or unauthorised receivers causing interference or resulting in the transmission of intelligence. An alternative technique, based on the tactical use of signal propagation prediction [34,35], may then be advantageous. The tactical technique exploits detailed knowledge of the ionosphere and ray tracing to minimise the signal coverage and thereby deny, or minimise, the hostile receiver's access to the radiated electromagnetic energy. The technique can be used in isolation or in conjunction with more sophisticated LPI/LPJ techniques.

Although this tactical technique could be performed using any of a number of current prediction programs it is operator intensive, specialist data (e.g. location of interceptors and frequency plans) are required and the contribution from the ground wave may not be assessed.

To overcome these deficiencies tactical HF decision aids have been developed. An example is the HF electromagnetic environmental modelling system (HF–EEMS)[36]. The aid is designed to enable both HF experts and inexperienced operators to predict those communications system parameters such as frequency, receiver station and radiated power which optimise the HF system. The decision aid uses a modified version of the REC533 prediction program [21]. HF–EEMS V1.1 also contains a simple ground-wave propagation model for line-of-sight (LOS < 200 km) systems. The ground-wave propagation characteristics are calculated by a modified version of GRWAVE [37], which computes ground wave field strengths over a smooth, curved, homogeneous earth.

The decision aid displays data to the operator in two formats: signal coverage maps and frequency tables. In Figure 17.4 both the sky-wave and ground-wave signals (the latter as concentric circles) are shown on the same map, along with friendly and hostile station locations. These monthly median maps enable users to inspect communications link configurations and evaluate the effectiveness of broadcast transmissions.

Via a table, the decision aid also recommends the best operating frequencies to minimise hostile interception. There are two approaches which can be employed, one based on LUF (lowest usable frequency) considerations and the other based upon the MUF. In the former the aim is to operate on a frequency which is below the LUF to the hostile receivers but above the LUF to the friendly receivers. The latter approach aims to choose a frequency which is above the MUF to the hostile receivers but below the MUF to the friendly receivers. It is to be noted that the LUF technique is more problematical since it relies on

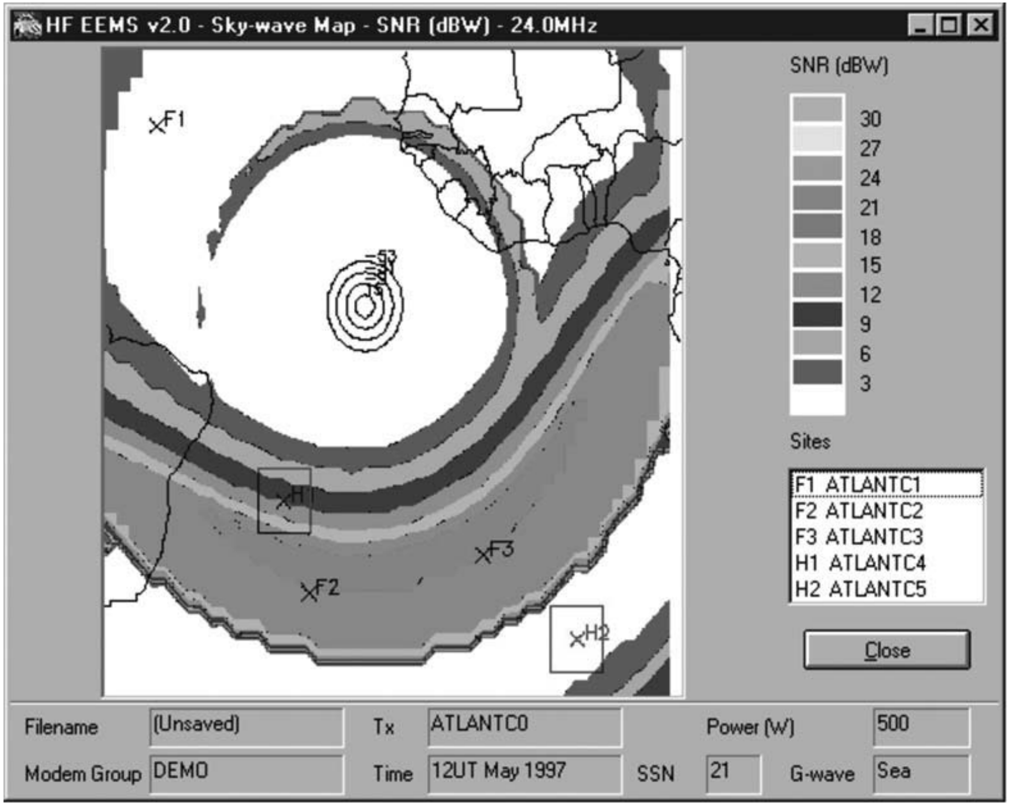


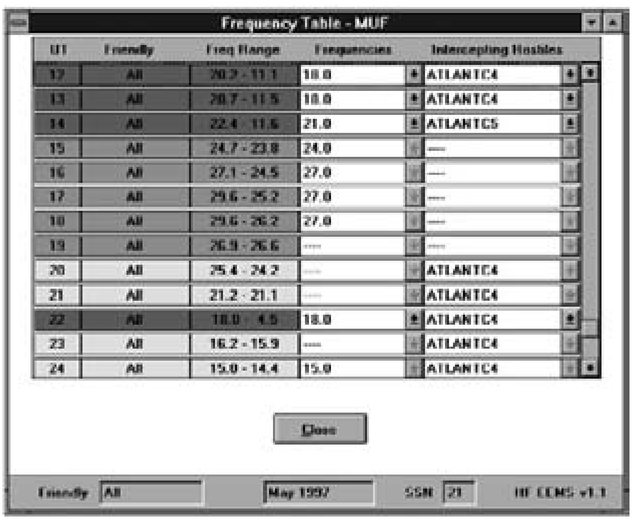
Figure 17.4 *Signal coverage map at 24 MHz with three friendly stations (+) and two hostile stations (×)*

signal strength estimates which are more difficult to achieve than MUF predictions.

Figure 17.5 is an extract from a 24-hour frequency table, based on MUF predictions, for the broadcast scenario shown in Figure 17.4. If mode support is predicted the table gives a frequency range within which all friendly stations will receive the signal, while minimising hostile interception. The frequencies in the assigned frequency plan which fall within this frequency range are also shown. The column on the far right of Figure 17.5 shows those hostile receivers predicted to intercept the broadcast transmissions. The hostile displayed is that most likely to intercept the signal; additional hostiles are listed in decreasing order of interception probability in drop down boxes (not shown). The operational versions of these tables are colour coded. Frequency ranges not intercepted by hostiles are coloured green (e.g. 15–19 UT). Those intercepted by at least one hostile (e.g. 20–21 UT) are coloured yellow to indicate that caution should be employed. If all hostiles intercept the signal (e.g. 12–14) then the frequency range is coloured red. As such, transmissions should be avoided unless vital.

17.5 Ionospheric sounders for HF frequency management

Ionospheric sounders (ionosondes) are essentially radars which measure the group delay of signals propagating *via* the ionosphere. Due to the structure of the ionosphere multiple radar returns can be obtained at a single frequency. Ionospheric sounders can be monostatic with the receiver and transmitter


Frequency Table - MUF

UT	Friendly	Freq Range	Frequencies	Intercepting Hostiles
17	All	20.2 - 11.1	18.0	ATLANTC4
13	All	20.7 - 11.5	18.0	ATLANTC4
14	All	22.4 - 11.6	21.0	ATLANTC5
15	All	24.7 - 23.8	24.0	----
16	All	27.1 - 24.5	27.0	----
17	All	29.6 - 25.2	27.0	----
10	All	29.6 - 26.2	27.0	----
19	All	26.9 - 26.6	----	----
20	All	25.4 - 24.2	----	ATLANTC4
21	All	21.2 - 21.1	----	ATLANTC4
22	All	18.0 - 4.5	18.0	ATLANTC4
23	All	16.2 - 15.9	----	ATLANTC4
24	All	15.0 - 14.4	15.0	ATLANTC4

Figure 17.5 Frequency ranges which minimise hostile signal interception during signal broadcasts using the MUF algorithm

colocated or they can be bistatic with the transmitter and receiver separated by many tens, hundreds or thousands of kilometres. The former generates vertical ionograms and the latter oblique ionograms.

The oblique sounder provides ionospheric information such as propagating frequencies and multipath over the path from the sounder transmitter to the sounder receiver. If the HF communications is over the same path as the sounder, the latter provides an invaluable decision aid to choose the correct operating frequency.

17.6 Real-time updates to models, ionospheric specification and forecasting

17.6.1 Introduction

The performance of a wide variety of communications, surveillance, tracking and navigation systems is controlled by the ionosphere. Empirical models of the nonauroral, quiet ionosphere typically describe the median monthly or seasonal conditions at every hour. Day to day *fo*F2 deviations from these median values can range from 10 to 30 per cent even for quiet magnetic conditions and will be higher still during ionospheric storms. Furthermore, there are short-term variations due to atmospheric gravity waves which typically have periods of 10–120 minutes. The difference between monthly median predicted and measured signal strengths can be up to 40 dB. These variations average out over a period of a month or so and consequently we can use the median ionospheric models for planning and certain other purposes.

However, there are circumstances where it is necessary to be able to correct for these temporal variations in real time. Empirical or climatological models (such as those used in the aforementioned REC533 and ICEPAC) are inadequate for the specifications and forecasts required by many modern operational systems. We need both an ionospheric model capable of assimilating the necessary data and also the appropriate data streams.

The most important developments in this field derive from the USA in support of the Air Force Space Forecast Center (AFSFC). Both the ICED and PIM models were conceived to allow for real-time updates of the input parameters from a number of sensors. ICED updates are based only on changes of an effective sunspot number SSN_{eff} together with the auroral Q index whereas real-time control of PIM, *via* the model known as PRISM (parameterised real-time ionospheric specification model), is more sophisticated.

17.6.2 PRISM

An example of this emerging capability is the on-going development of the parameterised real-time ionospheric specification model (PRISM), now operational at the U.S. Air Force Space Forecast Centre [19]. It is a parameterised model that can be driven in real-time by ground- and space-based sensor data, to provide electron density and ion profiles globally, every two degrees in latitude and every five degrees in longitude, from 90–1600 km altitude. PRISM provides specifications of the auroral boundary, the high latitude ionospheric trough, and the enhanced densities associated with the Appleton equatorial anomaly. These outputs can be adjusted on the basis of near real-time sensor data, including: bottomside electron-density profiles from digital ionospheric sounders; total-electron-content (TEC) data from GPS ground-receiving sites; energetic particle fluxes and auroral boundary data from SSJ/4 satellite sensors; 840 km altitude *in situ* electron/ion densities; temperatures and velocities from the SSIES satellite.

Vertical-looking ionograms (Figure 17.6) are one of the principal methods used to derive the electron density directly. Electron density information can also be obtained from oblique ionogram traces although this approach is less accurate. The oblique ionograms must be scaled to obtain a trace representing group path *versus* frequency information over the path. This is a difficult process. The scaled trace is then converted to an equivalent vertical ionogram trace [38–40] from which the electron density can be extracted using the technique appropriate to vertical ionograms.

The data streams described above can be supplemented by ultraviolet (UV) data from the Defense Meteorological Satellite Program (DMSP). The capacity and requirement for monitoring and sensing of the environment is rapidly increasing. Further down stream other enhancements, such as measurements of the solar EUV flux, thermospheric winds, equatorial and polar drift velocities and parameters to specify accurately geomagnetic activity are expected. As the variety and quantity of such sensor data become available, it is anticipated that

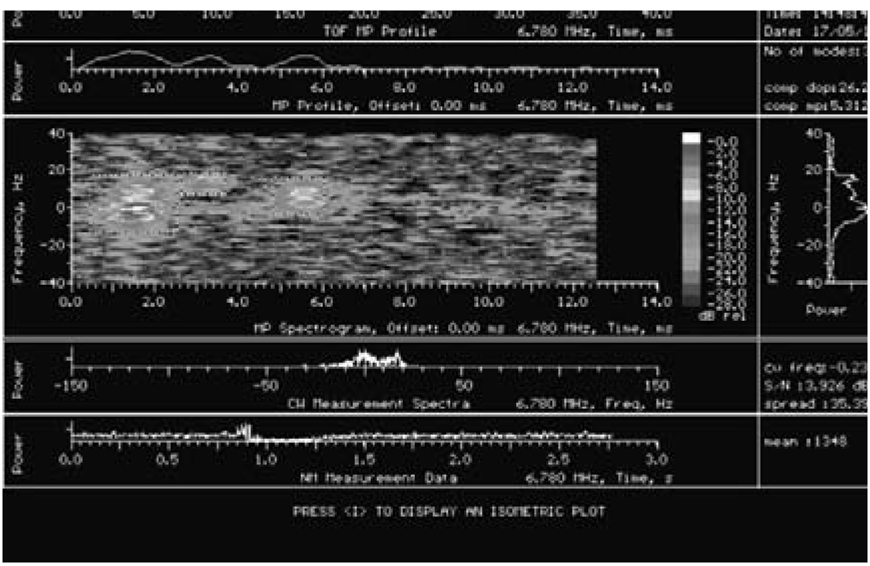


Figure 17.6 Vertical ionogram from Tromsø, Norway

Table 17.4 SESC ionospheric forecast

```
SUBJ:  HF RADIO PROPAGATION REPORT
JOINT USAF/NOAA BULLETIN PREPARED AT THE AIR FORCE SPACE
FORECAST CENTER, FALCON AFB, COLORADO.
PRIMARY HF RADIO PROPAGATION REPORT ISSUED AT 28/0525Z JAN 94.

PART I.    SUMMARY   28/0000Z  TO  28/0600Z  JAN  94/
                 FORECAST   28/0600Z  TO  28/1200Z  JAN  94.
                                      QUADRANT
                              I          II          III          IV
                        0 TO 90W     90W TO 180  180 TO 90E    90E TO 0
 REGION          POLAR        N4         N4          N5           N5
                 AURORAL      N3         N3          N4           N4
                 MIDDLE       N6         N6          N7           N7
                 LOW          N7         N7          N7           N7
                 EQUATORIAL   N7         N7          N6           N7

PART II.
GENERAL DESCRIPTION OF HF RADIO PROPAGATION CONDITIONS OBSERVED DURING THE 24
HOUR PERIOD ENDING 27/2400Z, AND FORECAST CONDITIONS FOR THE NEXT 24 HOURS.
CONDITIONS WERE GENERALLY NORMAL, EXCEPT FOR SOME DEGRADATIONS REPORTED IN
THE TRANSITION SECTORS OF THE AURORAL ZONE.  THE DEGRADATIONS WERE MOSTLY
SPREAD-F AND NON-DEVIATIVE ABSORPTION.
FORECAST:  EXPECT MOSTLY NORMAL CONDITIONS EXCEPT FOR THE AURORAL SECTORS,
ESPECIALLY THE TRANSITION SECTORS, WERE MINOR PROBLEMS SUCH AS INCREASED
SPREAD-F AND NON-DEVIATIVE ABSORPTION CAN BE ANTICIPATED.

PART III.    SUMMARY OF SOLAR FLARE INCLUDED IONOSPHERIC DISTURBANCES WHICH
MAY HAVE CAUSED SHORT WAVE FADES IN THE SUNLIT HEMISPHERE DURING THE 24 HOUR
PERIOD ENDING 27/2400Z JAN 94  . . . .
               START                        END                       CONFIRMED
FREQS AFFECTED
               0508Z                       0531Z                             NO
UP TO 13 MHZ
PROBABILITY FOR THE NEXT 24 HOURS  . . . . SLIGHT

PART IV.   OBSERVED/FORECAST  10.7 CM FLUX AND K/AP.
THE  OBSERVED 10.7 CM FLUX FOR 27 JAN 94 WAS 120.
THE FORECAST  10.7 CM FLUX FOR 28, 29, AND 30 JAN 94
ARE 120, 115, AND 110.
THE OBSERVED K/AP VALUE FOR 27 JAN 94 WAS  03/16.
THE FORECAST K/AP VALUES FOR 28, 29, AND 30 JAN 94
ARE 03/18,  02/10,  AND 03/15.
SATELLITE X-RAY BACKGROUND:  B3.5  (3.5 E MINUS 04 ERGS/CM S Q/SEC).
THE EFFECTIVE SUNSPOT NUMBER FOR 27 JAN 94 WAS 062.0.
```

the errors in the specification of the global ionosphere will decrease from ~35 per cent to below 5 per cent. This will lead to significant reductions in errors and outages associated with a number of systems.

17.6.3 Ionospheric forecasting

As well as real-time specification of the ionosphere (often called a nowcast), it is sometimes necessary to have available a forecast of the ionospheric environment. Ionospheric forecasts are based on data from a wide variety of sources, both ground based and satellite based. Time series analysis plus forecaster experience are then often combined to make the forecast. More recently,

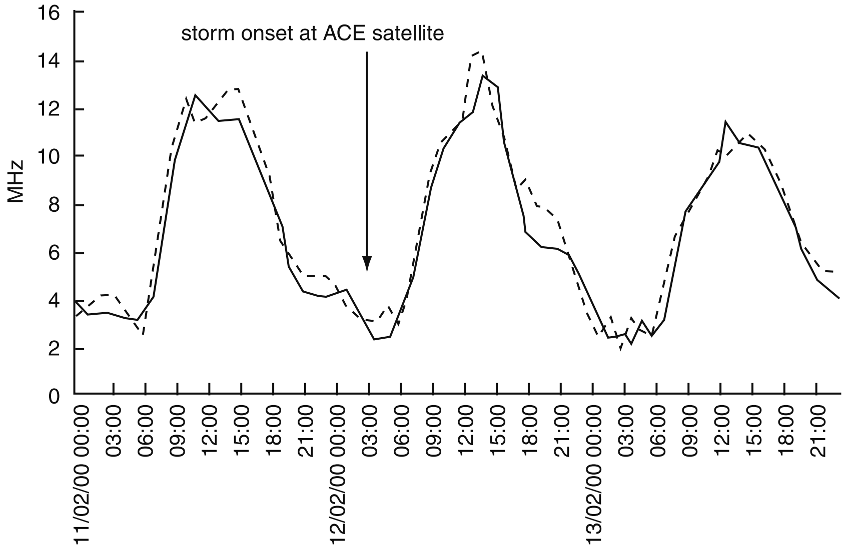


Figure 17.7 Time series of actual foF2 (broken line) together with predicted values (unbroken line) before and after a solar storm

nonlinear dynamical modelling has been used to improve the forecast quality [41,42] without recourse to heuristic approaches.

Ionospheric forecasts are communicated by e-mail, fax, telex and even by dedicated satellite links. Typically, the products cover geomagnetic activity, sunspot number and other ionospheric information for the immediate past, current and immediate future epochs. In addition to forecasts of the ionospheric electron density, forecasts of HF propagation conditions are often provided. Table 17.4 is a typical forecast for 28–30 January 1994. It is to be noted that this forecast provides an indication of the effective sunspot number specifically required by the ICED model. Figure 17.7 shows a further example of ionospheric forecasting of the parameter *fo*F2. In this example a solar storm has occurred and the forecast model has succeeded in tracking the resultant dip in *fo*F2.

17.7 References

1 CANNON, P.S.: 'Morphology of the high latitude ionosphere and its implications for HF communications systems', *IEE Proc. I*, 1989, **136**, pp.1–10
2 'Interoperability and performance standards for medium and high frequency radio equipment, MIL-STD-188–141A, Appendix A'. U.S. Department of Defense, 1988
3 BILITZA, D.: 'The world-wide ionospheric data base', National Space Science Center/World Data Center A for Rockets and Satellites, Greenbelt, Maryland, USA, report, NSSDC/WDC-A-R&S 89-03, 1989

4 BILITZA, D.: 'Solar-terrestrial models and application software', National Space Sciences Data Center, World Data Center A for Rockets and Satellites, Goddard Space Flight Center, Greenbelt, Maryland, USA, report NSSDC/WDC-A-R&S 90-19, 1990

5 RUSH, C.M.: 'Ionospheric radio propagation models and predictions – a mini-review', *IEEE Trans. Antennas and Propag.*, 1986, **34**, pp.1163–1170

6 SCHUNK, R.W. and SZUSZCZEWICZ, E.P.: 'First principal and empirical modelling of the global scale ionosphere', *Ann. Geophys.*, 1988, **6**, pp.19–30

7 BRADLEY, P.A. and DUDENEY, J.R.: 'A simplified model of the vertical distribution of electron concentration in the ionosphere', *J. Atmos. Terr. Phys.*, 1973, **35**, pp.2131–2146

8 DUDENEY, J.R.: 'An improved model of electron concentration with height in the ionosphere', *J. Atmos. Terr. Phys.*, 1983, **45**, pp.629–640

9 RAWER, K. and BILITZA, D.: 'International reference ionosphere – plasma densities: status 1988', *Adv. Space Res.*, 1990, *10*, pp.5–14

10 DUDENEY, J.R. and KRESSMAN, R.I.: 'Empirical models of the electron concentration of the ionosphere and their value for radio communications purposes', *Radio Sci.*, 1986, **21**, pp.319–330

11 ITU-R Recommendation P.1239: 'ITU-R reference ionospheric characteristics and methods for basic MUF, operational MUF and ray-path predictions', 1997

12 RUSH, C.M., POKEMPNER, M., ANDERSON, DN., STEWART, F.G., and PERRY, J.: 'Improving ionospheric maps using theoretically derived values of foF2', *Radio Sci.*, 1983, **18**, pp.95–107

13 RUSH, C.M., POKEMPNER, M., ANDERSON, D.N., STEWART, F.G., and REASONER, R.: 'Maps of FoF2 derived from observations and theoretical data', *Radio Sci.* 1984, **19**, pp.1083–1097

14 FOX, M.W. and MCNAMARA, L.F.: 'Improved empirical world maps of FoF2:1. the method', Ionospheric Prediction Service, Sydney, Australia Technical Report IPS-TR-86-03, 1986

15 FULLER-ROWELL, T.J., REES, D., QUEGAN, S., MOFFETT, R.J., CODRESCU, M.V., and MILLWARD, G.H.: 'A coupled thermosphere ionosphere model (CTIM)', *in* SCHUNK, R.W. (Ed.): 'Handbook of ionospheric models', STEP Report, 1996

16 ROBLE, R.G.: 'The NCAR thermosphere-ionosphere-mesosphere-electrodynamics general circulation model', *in* SCHUNK, R.W. 'Handbook of ionospheric models', STEP Report, 1996

17 BURNS, A.G., KILLEEN, T.L., DENG, W., and CARIGNAN, D.R.G.: 'Geomagnetic storm effects in the low to mid latitude upper thermosphere', *J. Geophys. Res.*, 1995, **100**, pp.14673–14691

18 FULLER-ROWELL, T.J., CODRESCU, M.V., MOFFETT, R.J., and QUEGAN, S.: 'On the seasonal response of the thermosphere and ionosphere to geomagnetic storms', *J. Geophys. Res*, 1996, **101**, pp.2343–2353

19 DANIELL, R.E., BROWN, L.D., ANDERSON, D.N., FOX, M.W., DOHERTY, P.H., DECKER, D.T., SOJKA, J.J., and SCHUNK, R.W.: 'PRISM: a global ionospheric parameterisation based on first principles models', *Radio Sci*, 1995, **30**, p.1499

20 TETERS, L.R., LLOYD, J.L., HAYDON, G.W., and LUCUS, D.L:

'Estimating the performance of telecommunications systems using the ionospheric transmission channel: ionospheric communications analysis and prediction program (IONCAP) user's manual', NTIA, Springfield, VA, USA, 1993, report 83-127, NTIS order no N70-24144
21 LANE, G., RHOADS, F.J., and DEBLASIO, L.: 'Voice of America coverage analysis program, (VOACAP). A guide to VOACAP'. US Information Agency, Bureau of Broadcasting, Washington, D.C. 20547-0001, 1993, B/ESA technical report 01-93
22 ITU-R Recommendation P.533–6: 'HF Propagation Prediction Method', 1999
23 STEWART, F.G. and HAND, G.: 'Technical description of the ICEPAC propagation prediction program'. Institution of Telecommunications Sciences, Boulder, Colorado, USA, 1994, private communication
24 TASCIONE, T.F., KROEHL, H.W., CREIGER, R., J.W. FREEMAN JNR, WOLF, R.A., SPIRO, R.W., HILMER, R.V., SHADE, J.W., and HAUSMAN, B.A.: 'New ionospheric and magnetospheric specification models', *Radio Sci.*, 1988, **23**, pp.211–222
25 SECAN, J.A., BUSSEY, R.M., and FREMOUW, E.J.: 'High latitude upgrade to the Wideband ionospheric scintillation model', *Radio Sci.*, 1997, **32**, pp.1567–1574
26 SECAN, J.A., BUSSEY, R.M., FREMOUW, E.J., and BASU, S.: 'An improved model of equatorial scintillation', *Radio Sci.*, 1995, **30**, pp.607–617
27 RINO, C.L.: 'A power law phase screen model for ionospheric scintillation, 1 Weak scatter', *Radio Sci.*, **14**, pp.1135–1145
28 BRIGGS, B.H. and PARKIN, I.A.: 'On the variation of radio star and satellite scintillations with zenith angle', *J. Atmos. Terr. Phys.*, 1963, *25*, pp.339–366
29 CHYTIL, B.: 'The distribution of amplitude scintillation and the conversion of scintillation indices', *J. Atmos. Terr. Phys.*, 1967, **29**, pp.1175–1177
30 NAKAGAMI, M.: 'Statistical methods', *in* 'Radio wave propagation', HOFFMAN, W.C. (Ed.) (Pergamon, 1960) pp.3–36
31 WHITNEY, H.E., AARONS, J., ALLEN, R.S., and SEEMAN, D.R.: 'Estimation of the cumulative amplitude probability distribution function of ionospheric scintillations', *Radio Sci.* 1972, **7**, pp.1095–1104
32 AARONS, J.: 'Global morphology of ionospheric scintillation', *Proc. IEEE*, 1982, **70**, pp.360–377
33 HILMER, R.V.: 'AF-GEOSpace user's manual, Version 1.4 and 1.4P' Tech. report AFRL-VS-TR-1999-1551. Hanscom AFB, MA, USA: AirForce Research Laboratories, 1999
34 ARGO and ROTHMULLER: 'PROPHET; an application of propagation forecasting principles'. Presented at *Solar-terrestrial predictions* workshop proc., 1979
35 GOODMAN, J.M., UFFELMAN, D.R., and FAHLSING, F.: 'The role of the propagation environment in HF electronic warfare'. Presented at *Propagation effects on ECM-resistant systems in communications and navigation*, NATO-AGARD Electromagnetic Wave Propagation Panel, Conference No. CP331, London, UK, 1982
36 SHUKLA, A.K., CANNON, P.S., ROBERTS, S., and LYNCH, D.: 'A tactical HF decision aid for inexperienced operators and automated

HF systems', Presented at IEE 7th international conference on HF, radio systems and techniques, Nottingham, UK, 1997, pp.383–387
37 ITU-R Recommendation P.386: 'Groundwave propagation curves for frequencies between 10 kHz and 30 MHz', 1999
38 REILLY, M.H.: 'Ionospheric true height profiles from oblique ionograms', *Radio Sci*. **20**, p.280–286
39 REILLY, M.H.: 'A method for real height analysis of oblique ionograms', *Radio Sci*. 1989, **24**, pp.575–583
40 HEATON, J.A., CANNON, P.S., ROGERS, N.C., MITCHELL, C.N., and KERSLEY, L.: 'Validation of electron density profiles derived from oblique ionograms over the UK', *Radio Sci*., 36.5, p.149, 2001
41 FRANCIS, N.M., CANNON, P.S., BROWN, A.G., and BROOMHEAD, D.S.: 'Nonlinear prediction of the ionospheric parameter foF2 on hourly, daily and monthly timescales', *J. Geophys. Res.*, 2000, **105**, pp.12839–12849
FRANCIS, N.M., BROWN, A.G., CANNON, P.S., and BROOMHEAD, D.S.: 'Non-linear prediction of the hourly foF2 time series in conjunction with the interpolation of missing data points', *Phys. Chem. Earth*, 2000, **25**, pp.261–265

Chapter 18

Surface waves, and sky waves below 2 MHz

John Milsom

18.1 Introduction

The principal modes of radiowave propagation at frequencies below 2 MHz are the surface wave and sky wave. In this Chapter these two modes are introduced and described. Rather than concentrate on the details of elaborate path-loss prediction theories, they are merely introduced and the discussion then concentrates on their application by the planning engineer.

The sky-wave propagation prediction methods described in this Chapter are only applicable at frequencies below 2 MHz. However, the surface-wave models are based on more general theories and can also be applied in the HF band.

Antenna and external-noise aspects of system planning are discussed.

Throughout, readers are directed towards relevant data sources, prediction procedures and computer programs so that they might apply the planning methods described.

18.2 Applications

The radio spectrum below 2 MHz is allocated to a variety of radio systems:

aeronautical radionavigation
amateur
fixed services
land mobile
maritime mobile coastal radiotelegraphy
maritime radionavigation (radio beacons)
mobile (distress and calling)
radiolocation
radionavigation

sound broadcasting
standard-frequency and time services

Some of these applications are discussed in more detail in Chapter 1.

18.3 Surface-wave propagation

18.3.1 What is the surface wave?

Consider the case of a transmitting antenna, T, above a perfectly conducting flat ground as depicted in Figure 18.1. The voltage, V, induced in the receiving antenna, at an arbitrary receiving position, R, might be expressed as a vector sum of direct and ground-reflected components:

$$V = QI\left\{Q_1 \frac{\exp(-jkr_1)}{r_1} + Q_2 R \frac{\exp(-jkr_2)}{r_2}\right\} \tag{18.1}$$

where I is the current in the transmitting antenna, Q is a constant, Q_1 and Q_2 take account of the transmitting- and receiving-antenna polar diagrams and R is the appropriate reflection coefficient (see Chapter 6). Other terms are defined in Figure 18.1.

In many cases, especially where the radiated frequency is in the VHF or higher frequency bands, the above calculation will give a perfectly acceptable result for practical applications. However, it transpires that a complete description of the field at R requires an additional contribution to the resultant:

$$V = QI\left\{Q_1 \frac{\exp(-jkr_1)}{r_1} + Q_2 R \frac{\exp(-jkr_2)}{r_2} + S \frac{\exp(-jkr_2)}{r_2}\right\} \tag{18.2}$$

where S is a complicated factor which depends on the electrical properties of the ground, transmitted polarisation, frequency and the terminal locations. When

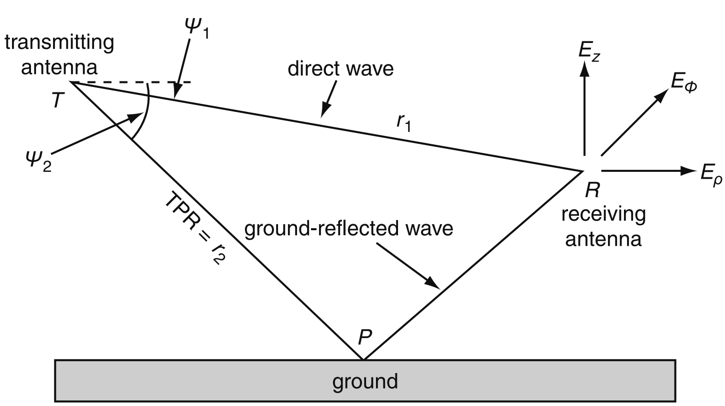


Figure 18.1 Geometry of direct and ground-reflected waves

introduced in this way, as an apparent afterthought to make up the numbers and satisfy Maxwell's equations, one is tempted to regard it as a minor contribution of interest primarily to the mathematical physicist. In fact, this third term represents the surface wave and it is a propagation mode of great practical value to radio systems operating in the HF and lower frequency bands. We will see later that, when the points T and R are close to the ground, the direct and ground-reflected waves act to cancel each other, leaving only the surface wave. Therefore, for example, during the daytime when ionospheric absorption smothers MF sky-wave modes, the surface wave is the carrier of all the signals which occupy the medium-wave broadcasting band. Surface waves also support the operations of LF broadcasting, VLF/LF communication and navigation systems, HF short-range communication and some classes of HF radar.

The surface wave propagates by virtue of currents which flow in the ground and does not depend for its existence on the atmosphere. Horizontally polarised surface waves are very heavily attenuated and have little or no practical worth. All the applications mentioned earlier utilise vertically polarised surface waves.

At this stage it is appropriate to introduce some new terms and clarify the relationship between surface waves, space waves and ground waves:

$$\text{ground wave} = \underbrace{\text{direct wave} + \text{reflected wave}}_{\text{space wave}} + \text{surface wave}$$

Unlike ionospherically propagated signals, the surface wave suffers negligible dispersion so that, in principle, wideband signals can be transmitted when the surface wave alone is active.

Fading only occurs when there is some temporal variation in the propagation path. Overland ground waves are stable signals. Oversea ground waves can be subject to slow fading due to changing tidal effects and shifts in the seawave characteristics.

18.3.2 Theory for a homogeneous smooth earth

18.3.2.1 Plane finitely conducting earth

Sommerfeld–Norton flat-earth theory: after Sommerfeld [1], Norton [2, 3] derived expressions for the ground-wave field-strength components above a finitely conducting plane earth due to a short vertical current element. In its full form Eqn. 18.2 becomes:

$$E_z = j30kIdl\left[\left\{\cos^2\psi_1 \frac{\exp(-jkr_1)}{r_1} + \cos^2\psi_2 R_v \frac{\exp(-jkr_2)}{r_2}\right\} + \right.$$
$$\left.\left\{(1 - R_v)(1 - u^2 + u^4\cos^2\psi_2)F\frac{\exp(-jkr_2)}{r_2}\right\}\right] \qquad (18.3)$$

$$E_p = -j30kIdl\left[\sin\psi_1\cos\psi_1\frac{\exp(-jkr_1)}{r_1} + \sin\psi_2\cos\psi_2 R_v\frac{\exp(-jkr_2)}{r_2} - \right.$$

$$\left.\cos\psi_2(1-R_v)u\sqrt{(1-u^2\cos^2\psi_2)}\left\{1-\frac{u^2}{2}(1-u^2\cos^2\psi_2)+\frac{\sin^2\psi_2}{2}\right\}F\frac{\exp(-jkr_2)}{r_2}\right] \quad (18.4)$$

where ψ_1 and ψ_2 are defined in Figure 18.1, j is the square root of -1, k is the radio wavenumber = $2\pi/\lambda$, Idl is the product of source current and length – the 'dipole moment', R_v is the plane-wave Fresnel reflection coefficient for vertical polarisation and F is an attenuation function which depends on ground type and path length. F is given by the expression:

$$F = [1 - j\sqrt{(\pi w)}\exp(-w)\{\mathrm{erfc}(j\sqrt{w})\}] \quad (18.5)$$

erfc denotes the complementary error function (Abramowitz and Stegun) and

$$w = \frac{-j2kr_2u_2(1-u^2\cos^2\psi_2)}{(1-R_v)} \quad (18.6)$$

$$u^2 = \frac{1}{(\varepsilon r - jx)} \quad (18.7)$$

and

$$x = \frac{\sigma}{(\omega\varepsilon_0)} = 1.8\times 10^4\frac{\sigma}{f_{MHz}} \quad (18.8)$$

σ is the conductivity of the earth in S m^{-1}, $\varepsilon_r = \varepsilon/\varepsilon_0$ is the relative permittivity of the earth and f_{MHz} is the frequency in MHz.

Note that Eqns. 18.3 and 18.4 represent field components in the vertical and radial directions of a cylindrical co-ordinate system.

Special case of ground-based terminals: when the points T and R are both at the ground so that $R_v = -1$ (see Chapter 6) and $\psi_1 = \psi_2 = 0$, the direct and ground-reflected waves act in opposition and sum to zero. Such circumstances will prevail in many practical applications. When this happens the surface wave dominates and may be described by somewhat simplified forms of Eqns. 18.3 and 18.4, thus:

$$E_z = j60kIdl(1-u^2+u^4)F\frac{\exp(-jkr)}{r} \quad (18.9)$$

$$E_p = j30kIdl\{u\sqrt{(1-u^2)}(2-u^2+u^4)\}F\frac{\exp(-jkr)}{r} \quad (18.10)$$

In the pure surface wave the vertical and radial components of the electric field are still present. In physical terms this means that the propagating wavefront is tilted. The radial component given by Eqn. 18.10 is small relative to the vertical component described by Eqn. 18.9. The phase relationship is such that the modest wavefront tilt is in the direction of propagation. The degree of tilt depends on ground conductivity and frequency. Measurements of wave tilt can be used to infer the electrical properties of the local ground. Because E_p is finite and the magnetic field component is horizontal, there exists a downward component of the Poynting vector and energy is lost from the horizontally propagating wave. In this way attenuation occurs in addition to that due to ordinary inverse-square-law spreading. Responsibility for describing this extra attenuation within the Sommerfeld–Norton theory falls to the term F which appears in the earlier expressions.

For ground-based terminals the attenuation factor, F, introduced in Eqn. 18.5 still appears, but w simplifies to become:

$$w = \frac{-jkru^2}{2}(1 - u^2) \tag{18.11}$$

Having survived this short excursion into the realm of the theoretical physicist it is time to retreat and attempt an engineering interpretation of the above results.

Interpretation and key results: radio coverage predictions are almost invariably conducted in terms of electric field strengths. This approach also prevails in LF and MF broadcasting even though most domestic receivers now incorporate ferrite rod antennas which are sensitive to the radio frequency magnetic field. The propagating surface wave contains a horizontal magnetic component, H_ϕ, which is approximately related to the major electric component *via* the expression:

$$H_\phi = -\frac{E}{Z_0} \tag{18.12}$$

where Z_0 is the intrinsic impedance of free space ($\approx$ 377 Ω). It is sufficient therefore to design in terms of the electric field strength.

Attenuation of the surface wave arises through the forward tilt of its electric field. The rate of attenuation becomes more marked as the tilt angle increases. By combining Eqns. 18.9 and 18.10, it is possible to show the ratio of electric field components to be simply related by:

$$\frac{E_\rho}{E_z} \approx u = \frac{1}{\sqrt{K_r}} \tag{18.13}$$

K_r is the complex dielectric permittivity of the ground. It varies with frequency and the electrical properties of the ground.

Some representative values are presented in Table 18.1.

Large values of K_r correspond, according to Eqn. 18.13, to low degrees of forward tilt and therefore attenuation. We can conclude that surface-wave attenuation is greatest over ground of low conductivity and at high radio frequencies. Sea water has an outstandingly high conductivity and the surface wave, with a near-vertical electric field, propagates over it with relatively low

Table 18.1 Typical values of complex dielectric permittivity for different ground types and frequencies

Ground type	Frequency (kHz) 200 (LF)	1000 (MF)
Sea (σ =5 S m^{-1}, ε_r = 70)	70 – *j*450000	70 – *j*90000
Good ground (σ = 10^{-2} S m^{-1}, ε_r = 10)	10 – *j*900	10 – *j*180
Poor ground (σ = 10^{-3} S m^{-1}, ε_r = 4)	4 – *j*90	4 – *j*18

attenuation. This conclusion will be shown more graphically in Section 18.3.4.

In Eqn. 18.9 the factor $(1 - u^2 + u^4)$ is close to unity for all practical situations. The amplitude of the vertical component of electric field is therefore given by:

$$|E_z| = \frac{300}{r}\sqrt{P}\,|F| \tag{18.14}$$

where P is the total radiated power from the Hertzian-dipole current element, expressed in kW, r is the path length in km and E_z is the electric field strength in mV/m.

All the interesting effects are associated with $|F|$. Within a few wavelengths of the signal source $|F|$ is approximately unity. In this regime the field strength varies as $1/r$, i.e. in inverse-distance fashion. At sufficiently large distances $|F|$ makes a transition to become inversely proportional to distance so that the field strength varies as $1/r^2$. This long-range behaviour will persist for as long as the Sommerfeld assumption of a plane earth remains valid.

Usually it is necessary to compute fields from more practical antennas. This merely entails substituting a different constant on the right-hand side of Eqn. 18.14. Chapter 2 gives some useful factors.

18.3.2.2 Spherical finitely conducting earth

The next stage of refinement in our efforts to devise a realistic propagation model involves substituting a spherical earth shape for Sommerfeld's plane model. At short ranges Sommerfeld's ground-wave model can be applied without adaption. At longer ranges, and when the two terminals are beyond line of sight, it is necessary to compute fields with proper regard for diffraction over the curved earth [4–8].

It would serve no useful purpose to detail the associated theory in this overview of surface-wave propagation. In any case the mathematics is complicated and not easily understood. Bremmer [9] gives a good account of the theory.

It transpires that the curved earth introduces a third range regime, normally beyond that where inverse-square-law field variation occurs, in which the decrease in field strength becomes exponential. The starting distance of this exponential behaviour can be estimated by the expression:

$$\frac{80}{\sqrt{f_{MHz}}}\ \text{km} \tag{18.15}$$

Apart from this new far-range behaviour, most of the other characteristics of the surface wave above a spherical earth are identical to those deduced from Sommerfeld's plane earth model.

18.3.3 Atmospheric effects

Terrestrial surface waves would propagate in the total absence of an atmosphere. All the theoretical work of Sommerfeld, Norton, Van der Pol and Bremmer reported above has ignored atmospheric effects. They have assumed that a wave propagating in space above the ground would travel in a straight line. In practice, the earth's atmosphere is stratified and possesses a refractive index which normally decreases with height. On average the height variation is exponential. Near the ground it is sometimes sufficient to assume a linearly decreasing height profile.

In any atmosphere where the refractive index decreases with height, a radiowave will be refracted towards the ground. If the profile is linear, a remarkably simple change to the Van der Pol/Bremmer/Norton theory enables the atmospheric refraction to be accommodated. It is only necessary to increase artificially the earth radius above its true value of some 6371 km. Such a trick is commonly applied in the VHF and higher frequency bands (see Chapter 7). An effective earth radius should only be used when both terminals are near to the ground and at frequencies greater than 10 MHz.

Rotheram [10–12] has explored the behaviour of the effective earth radius multiplying factor in the frequency band where surface waves are of practical importance. Above 30 MHz a factor of 1.374 is appropriate. At frequencies below 10 kHz the atmosphere has negligible effect and the factor tends to unity. In the neighbourhood of the MF broadcasting band the factor lies in the range 1.20–1.25 for most classes of ground. Rotheram's results have been computed for average atmospheric conditions. During times of abnormal atmospheric conditions, effective earth radius factors outside the range 1–1.374 may be required to simulate the prevailing propagation effects.

When problems arise where one or both terminals are elevated, energy propagating between the two encounters a refractivity/height profile which is approximately exponential and clearly nonlinear. Such paths cannot be modelled using an effective earth radius factor and an atmosphere-free propagation theory. The nonlinear refractivity profile becomes significant at all frequencies when a terminal is elevated above 1 km and at frequencies below 10 MHz even for terminals on the ground [12].

Happily, these doubts about the validity of atmosphere-free theories and the earth radius factor have been overcome by the work of Rotheram. Rotheram has developed a general-purpose ground-wave prediction method and an associated computer program. The method incorporates an exponential atmospheric refractivity profile. It is now recommended by the ITU-R for system planning and has been adopted by many agencies. Because of its practical worth, Rotheram's program, GRWAVE, is described in fair detail in Section 18.3.4.

18.3.4 ITU-R recommended prediction method

Rotheram describes three methods of predicting space-wave and surface-wave fields over a smooth homogeneous earth surrounded by a uniform atmosphere which exhibits an exponential refractivity/height profile. The methods cater for elevated terminals and a very wide frequency band. No single method is effective for all path geometries but, using appropriate numerical techniques, it is possible to establish whether or not one method is working and, if it is not, switch to a better approach. The methods and their approximate regions of validity are:

(i) Residue series (mode summation): used at the farthest distances, for elevated terminals this is beyond the radio horizon. For terminals near the earth's surface it is for distances greater than approximately 10 $\lambda^{1/3}$ km where λ is the radio wavelength in metres.
(ii) Extended Sommerfeld flat-earth theory: an extended Sommerfeld theory can be applied at short ranges and small heights. These restrictions turn out to be ranges less than approximately 10 $\lambda^{1/3}$ km and heights below 35 $\lambda^{2/3}$ m.
(iii) Geometrical optics (ray theory): this final method is applied within the radio horizon when the terminal heights are above that which can be handled by the Sommerfeld approach. It involves calculating the phase and amplitude of the direct and ground-reflected paths with due regard for the atmospheric refraction.

Fortunately, the three methods are able to deal with all reasonable geometries so that inelegant interpolation between two inappropriate results is unnecessary. Under circumstances where two methods are simultaneously valid the results are found to be in good agreement.

Rotheram has written a Fortran computer program to compute ground-wave fields using these theories. This program, GRWAVE, has been used by the ITU-R to produce a series of curves which show how vertically polarised electrical field strength varies as a function of range, ground type and frequency (10 kHz to 30 MHz). In doing this the ITU-R has elected to adopt a global average refractivity profile given by:

$$n = 1 + (n_s - 1)\exp(-h/h_s) \tag{18.16}$$

where n_s = surface refractivity = 1.000315 and h_s = refractivity profile scale height = 7.35 km. These curves are to be found in Recommendation ITU-R P.368 [13].

Although the ITU-R curves are comprehensive, it is sometimes useful to use GRWAVE program itself.

Figure 18.2 shows some example ITU-R ground-wave curves for sea, land (σ = 0.03 S m^{-1}) and very dry ground (σ = 0.000 1 S m^{-1}). The latter two represent some extreme ground conditions. Sea is best regarded as being in a class of its own. For these curves both terminals are on the ground so that, in their computation, the geometrical optics parts of GRWAVE will not have been invoked. A close inspection of Figure 18.2 will reveal the inverse-distance,

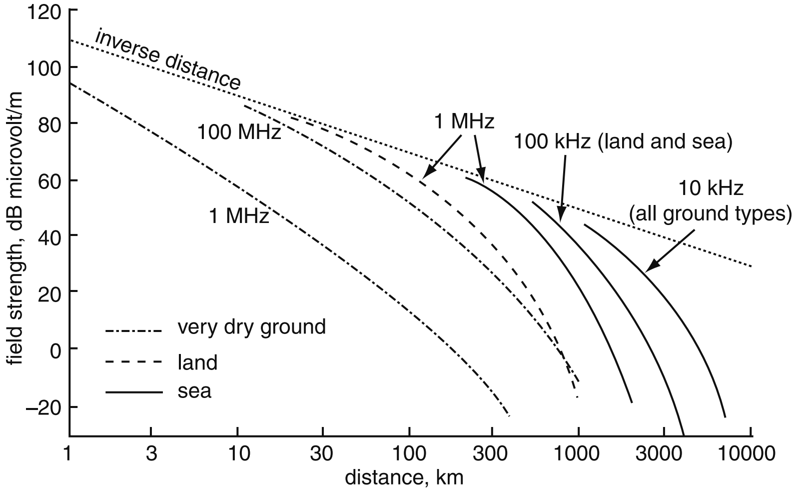


Figure 18.2 ITU-R surface-wave curves

inverse-square-distance and exponential range attenuation regimes which were inferred in earlier Sections. Also note the very strong frequency and ground-conductivity dependence of surface-wave attenuation over dry ground.

Figure 18.2 shows the relatively low attenuation experienced by the surface wave over sea. In the frequency bands below approximately 3 MHz the wave passes over the first 100 km of sea with an inverse-distance attenuation rate, i.e. as if the ground was perfectly conducting. However, by a range of approximately 400 km even a VLF transmission at 10 kHz begins to suffer losses beyond that given by the inverse-distance line.

In applying these curves for system planning purposes it is essential to have a clear understanding of the reference radiator used in their calculation. For the ITU-R curves the transmitting antenna is a Hertzian vertical dipole with a current length product (dipole moment) of $5\lambda/2\pi$. This moment has been carefully selected so that the characteristics of the dipole are identical to those of a short vertical monopole radiating 1 kW, a configuration which is easier to visualise.

Such a monopole, located over a perfectly conducting plane, will establish a field of 300 mV/m at a distance of 1 km along the plane. This factor appears explicitly in Eqn. 18.14 where the same reference antenna was used. It is a trivial matter to adjust the curves so that actual radiated power and antenna gain, relative to the 1 kW short monopole, are incorporated.

For some applications it is convenient (or conventional) to work in terms of transmission loss rather than field strength. Radar systems are one such case. Transmission losses are defined in terms of ratios of power transmitted to power received, but there are several ways of constructing this type of ratio and care is required (see Chapter 2).

When the transmitting and receiving terminals are on the ground, as assumed in Figure 18.2, it is easy to relate ITU-R field strength curves to basic transmission loss:

$$L_b = 142.0 + 20 \log_{10} f_{MHz} - E \tag{18.17}$$

where E is the field strength, in dB (μV m^{-1}) for the ITU-R reference radiator.

The frequency term in the above equation arises because the collecting aperture of a receiving antenna, and therefore the power available, depends on radio frequency.

ITU-R Recommendation P.341 [14] and Chapter 2 discuss the use of L_b in system design calculations. This will not be reiterated here. It is, however, important to address one key point and this relates to the antenna gain definition which must be used when L_b is involved. In the system calculation, actual antenna gains must be introduced as relative to an isotropic antenna at the same location. This is not the convention commonly used by antenna manufacturers in their sales literature. They tend to use dB relative to a truly isotropic antenna in free space.

This apparently trivial matter of antenna gain in ground and surface-wave path-loss calculations can lead to great confusion and, worse, an incorrect system design. The problem is compounded by the existence of several different (but internally consistent) sets of definitions and papers in the open literature which present vague or incorrect accounts of antenna gain.

18.3.5 Ground conductivity maps

In practice, when the propagation engineer has been asked to compute the viability of a link or the coverage of a broadcasting station, one of the most difficult stages is to acquire a proper description of the ground conductivity along the path or in the area of interest. A useful source of data lies in ITU-R Recommendation P.832 [15]. Here, geographic maps are presented on which areas of differing electrical properties are delineated. The maps are a conglomerate of information collected over many years, and the degree of detail varies from region to region.

Once the ITU-R maps have been consulted, it is often worth seeking supplementary local information about the ground types present. Broadcasting authorities sometimes have details of ground conductivity in their areas which may not have found its way into Recommendation P.832.

When the ITU-R and local information sources both prove to be inadequate, it is sometimes instructive to consult geological maps of the district and attempt an association of mineral type with electrical properties.

18.3.6 Smooth earth of mixed conductivity

So far we have discussed increasingly elaborate models of surface-wave propagation and the ways in which they might be applied. However, all of the models only treat the case of a homogeneous smooth earth. In practice, it is often necessary to solve planning problems which involve changes in ground type along the propagation path.

Suppose we have a situation where a surface-wave link must be established between two terminals which are located on ground of different electrical properties. At some point along the smooth propagation path a transition occurs between the two ground types. The upper half of Figure 18.3 shows the situation in schematic form. How do we compute the electric field strength at R due to the transmission from T?

Eckersley suggested that this might be done by using sections of the surface-wave attenuation curves appropriate to the radio frequency and different ground types. Figure 18.3 shows the idea in graphical form. The Eckersley construction can be made using the curves for homogeneous ground such as those published by the ITU-R. Intuitively, this method appears correct. However, it produces results in poor agreement with experiment and, furthermore, violates the need to have reciprocity on the path. Reciprocity demands that, if the transmitter and receiver were transposed, so that the wave encountered the two types of ground in reverse order, then the field strength at the receiver would be unchanged.

Millington, in a classic paper on ground-wave propagation, presented a simple but effective method of solving the problem depicted in Figure 18.3. His work was done at a time when reliable measurements of field strength changes in the neighbourhood of a conductivity transition were scarce. Others were developing analytic solutions to the problem but their results were complicated and unsuitable for practical application. Millington's approach was a blend of known theory and physical intuition. The argument proceeded along the following lines.

Suppose T in Figure 18.3 is well removed from the conductivity transition at X. In this case the surface wave which is launched in the direction of R will have a rate of attenuation with horizontal distance and a variation with height above ground which is characteristic of ground type 1. The field can be computed using a model for the homogeneous earth with little loss of accuracy. Similarly, if R is well removed from X then the surface-wave field strength will vary with horizontal distance and height in a way which is largely dictated by the ground type 2. In effect, it might appear as if the signal had travelled from T to R over homogeneous ground of type 2. The only residual evidence of its passage over the type 1 ground will be a shift in the absolute signal level. Homogeneous earth propagation models cannot be applied directly to compute the absolute field strength on the receiver side of X.

Consider now what might be happening close to the conductivity transition point X. Millington argued that some sort of distortion or disturbance of the surface wave must occur. On approaching the transition from T the field will probably be affected even before X itself is encountered. However, this must be modest in comparison with the trauma undergone by the field on the receiver side of the transition.

Millington then elaborated the discussion and sought to establish a prediction method which would give a consistent change in the height profile of the field

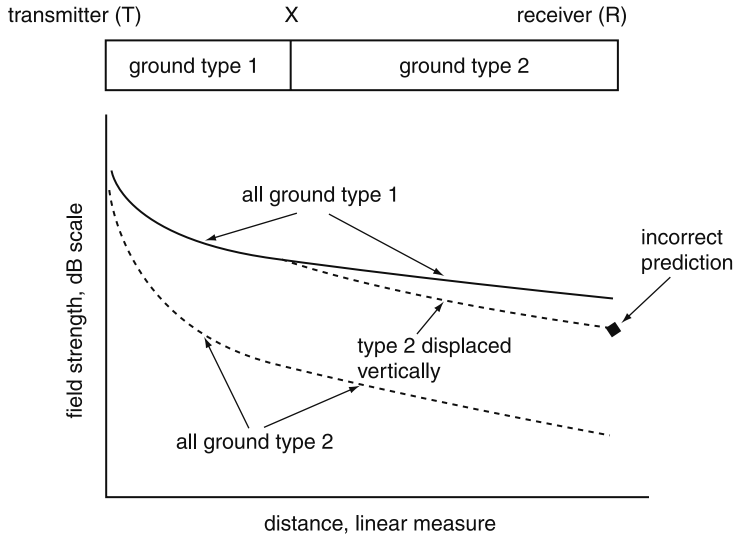


Figure 18.3 Smooth earth of mixed conductivity; Eckersley's prediction method

strength near the ground-type transition. An essential constraint on the form of the method was that it must satisfy the reciprocity requirement. A proposal was made, without proof or mathematical rigour, that the field strength might be estimated by a double application of Eckersley's method, followed by an averaging of the two results (expressed in dB). One application is made in the forward direction and then another in the reverse direction, as if the placement of T and R were reversed. Formation of an average forces the solution to satisfy reciprocity. Figure 18.4 shows how the Millington method should be applied.

When the method is applied to the situation of signals propagating from ground of high conductivity to ground of a lower conductivity, the disturbance at the boundary appears as a somewhat abrupt decrease in field strength. Eventually, the field variation takes on the character of the low-conductivity ground.

When the method is applied to signals propagating from ground of low conductivity to ground of high conductivity, a much more remarkable effect is predicted. The phenomenon is most marked at a land–sea boundary. The field strength undergoes an abrupt increase with range, immediately on crossing the coast. Millington sought to explain the unexpected recovery in terms of a redistribution of signal energy from elevated portions of the wavefront down to lower levels.

In 1950 Millington and Isted [16], in a carefully executed experiment, measured the recovery effect at 3 MHz and 75 MHz and demonstrated an excellent agreement with the new prediction method. Figure 18.5 shows the degree of prediction accuracy at 3 MHz.

The method can be applied to paths with more than two ground sections by exactly the same procedure. Eckersley's method is used for forward and reverse routes and then the average of the two results is computed.

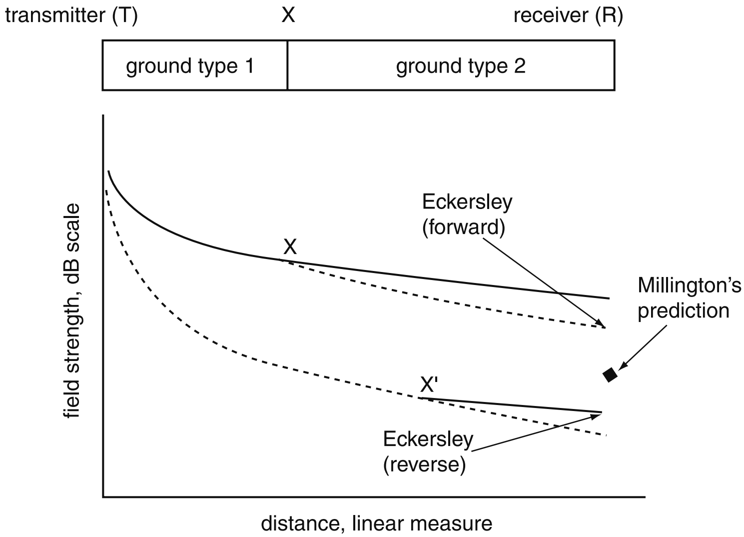


Figure 18.4 Millington's prediction method

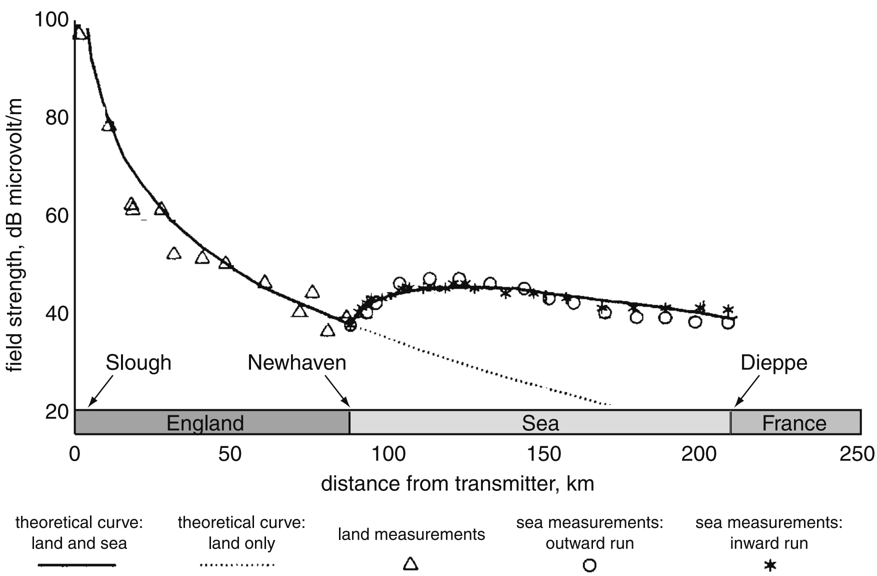


Figure 18.5 Calculated curve and experimental observation of the land–sea recovery at 3 MHz [16]

Since 1950, Millington's method has been very widely used and is still recommended by the ITU-R for surface-wave planning where ground conductivity changes occur along the propagation path.

18.3.7 The effects of buildings

When a surface wave encounters a built up area, its normal propagation characteristics are modified by the presence of electrically conducting structures. The kind of structures involved are steel-framed buildings, electrical wiring, electric lampposts, plumbing and also trees. Many of these structures can be considered to be earthed parasitic monopoles. Measurements of the horizontal magnetic field strength of MF surface waves in London have been made by Causebrook [17]. The field strength variations with range showed very marked minima and behaviour which is quite unlike anything which can be explained by published surface-wave curves, even in combination with Millington's method.

Causebrook demonstrated how his London measurements could be explained using Sommerfeld–Norton flat earth propagation theory. It transpires that the unusual behaviour is, in fact, present in the attenuation function, F (Eqn. 18.5). In rural areas the real component of the complex-valued w is negative and in this regime $|F|$ decreases monotonically with path length. In urban areas the man-made structures causes the real part of w to be positive. In this regime the attenuation function can have a distinct minimum at a path length of approximately 100λ which might account for the measured behaviour.

Causebrook showed that the effect of a bed-of-nails-type structure in urban areas will modify the surface impedance so that w enters the region near the minimum in $|F|$. It was not possible to deduce the surface impedance analytically. However, it was possible to derive an empirical model which is parametric in average building height and the fractional area covered by buildings.

Although there is no current ITU-R recommendation on how to model the effect of buildings in surface-wave system planning, the Causebrook approach is certainly worthy of consideration.

18.4 Sky-wave propagation below 2 MHz

18.4.1 What is the sky wave? – hops and modes

The sky wave is that part of the total received signal which relies on the presence of the ionosphere for its existence. The sky wave is a more easily understood concept than the surface wave.

In the frequency bands below 2 MHz there are essentially three methods of estimating the sky-wave field strength:

(i) a theoretical waveguide mode method by which propagation is analysed as the sum of waves corresponding to modes in the waveguide formed by the earth and lower ionosphere;

(ii) a theoretical method called wave hop in which the signals are modelled as one or more geometrical raypaths reflected from the lower ionosphere; this approach is similar, in principle, to that used in HF sky-wave prediction methods;
(iii) an empirical method; see ITU-R Recommendation P.1147 [5] for the planning of sound broadcasting services in the LF and MF bands.

To a degree, the three methods are complementary. The choice of method depends mainly on the combination of frequency and ground ranges of interest. Figure 18.6 is a rough guide to the regions of applicability of each method.

The three methods are outlined in Sections 18.4.2, 18.4.3 and 18.4.4, respectively.

18.4.2 Waveguide-mode field strength prediction theory

At frequencies below about 60 kHz the distance between the earth and ionosphere is less than a few wavelengths and the cavity tends to act as a waveguide. For example, at the main Omega frequency of 10.2 kHz, the separation is less than approximately three wavelengths. Surface wave and sky wave therefore cannot be considered independently, except at short ranges.

Propagation losses for ELF/VLF and lower LF signals in the earth–ionosphere waveguide are very modest. However, as discussed in Section 18.5, poor transmitting antenna efficiency can be a severe constraint. Communication services suffer from unavoidably narrow bandwidths and therefore low data rates. On the other hand, the excellent phase stability of continuous-wave signals makes these bands ideal for long-range hyperbolic navigation services.

Figure 18.7, derived from Morfitt *et al.*, shows a typical VLF fieldstrength/

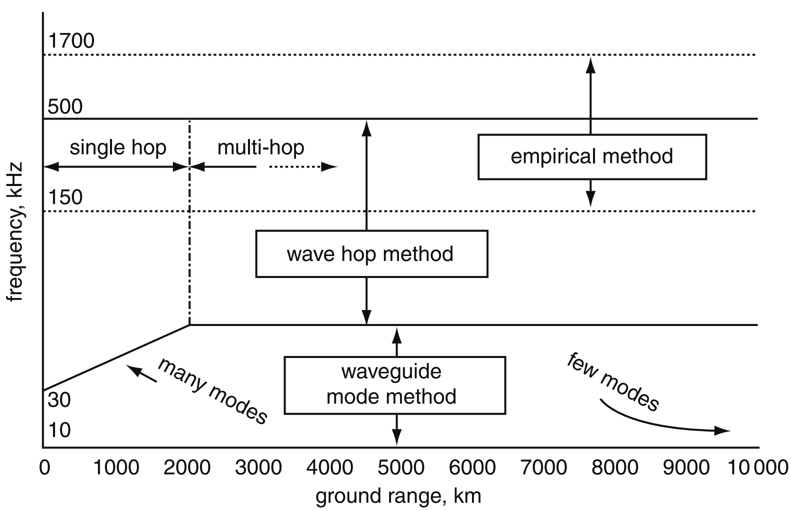


Figure 18.6 Sky-wave prediction methods; approximate domains of applicability

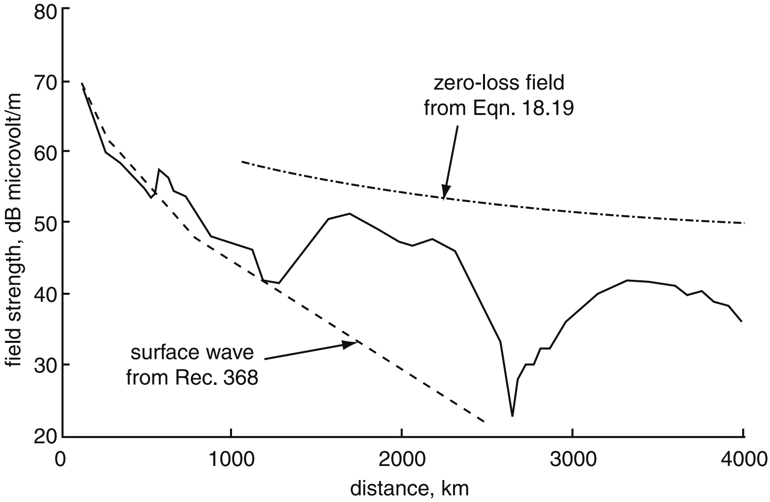


Figure 18.7 Measured field strength at night on 15.567 kHz

distance curve measured over a sea path at night at about 15.5 kHz. It can be seen that the measured curve agrees closely with the ITU-R surface-wave curve for distances up to 1200 km. Beyond this the influence of the ionosphere is clearly apparent.

It is interesting to compare the measured field strength with that predicted by a very simple prediction model. Imagine the earth–ionosphere waveguide to be loss free and that the radiated power P is distributed uniformly over the wavefront. If the earth were flat the area of the cylindrical wavefront would be $2\pi hd$, where d is the distance from the transmitter and h is the height of the ionosphere ($d >> h$). The power flux over the wavefront would therefore be $P/2\pi hd$, which is also equal to E^2/Z_0, where E is the field strength. The field strength, in mV m^{-1}, according to this simple model is thus given by:

$$E = 245\sqrt{\frac{P}{hd}} \tag{18.18}$$

where P is in kW and h and d are in km.

Eqn. 18.18 shows that the field strength would decrease as the square root of the distance; this is of course less than the inverse square law spreading in free space. The rate of attenuation is decreased still further by Earth curvature; if this is taken into account the expression for E becomes:

$$E = 245\sqrt{\frac{P}{ha\sin(d/a)}} \tag{18.19}$$

where a is the radius of the earth in km.

Figure 18.7 also shows the field strength calculated using Eqn. 18.19, for h = 90 km. Comparison with the measured field strengths shows the attenuation within the earth–ionosphere waveguide to be small at night. The attenuation measured during the day is somewhat greater.

Clearly, there are important features in the relation between measured field-strength variation and range which are not described by the very simple theories offered above. In the example shown in Figure 18.7, signal-strength minima occur at 1200 and 2700 km. Various theories have been evolved to explain low-frequency propagation and a survey can be found in Recommendation ITU-R P.684-2 [24]. One useful theory treats the total field as being the sum of the main waveguide modes which can propagate in the cavity. Destructive interference between the active modes gives rise to the observed signal-strength nulls. The higher the frequency and the shorter the range the greater is the number of significant modes. At 15.5 kHz, for example, the earth–ionosphere waveguide can support at least four TM (transverse magnetic) modes. At ELF, on the other hand, it is usually necessary to consider only one mode.

An expression due to Wait [20] for the sum of waveguide modes excited by a short monopole can be written:

$$E = 300\sqrt{\frac{P}{a\sin(d/a)}}\frac{\sqrt{\lambda}}{h}\exp\left\{-j(kd+\pi/4)\right\}\sum_n \Lambda_n \exp(-jkS_n d) \qquad (18.20)$$

where λ is the wavelength, in km, k is the free-space wavenumber $= 2\pi/\lambda$, Λ_n is the excitation factor of the nth mode and kS_n is the propagation coefficient of the nth mode.

The terms Λ_n and S_n are complex. The excitation factors give the relative amplitude and phase of the various modes excited in the waveguide by the source. The real part of the propagation coefficient kS_n contains the phase information of each mode and the imaginary part gives its attenuation rate. These factors depend on wavelength, ionospheric height, electrical properties of the ground and the reflection coefficients of the ionosphere.

ITU-R Recommendation P.684 describes in more detail how the reflection coefficients of the ionosphere may be computed and used in the above waveguide-mode prediction method. In addition, the Recommendation describes a more advanced form of waveguide-mode theory which provides a full-wave solution. Various workers have developed computer programs to evaluate waveguide-mode methods. An implementation of the algorithm is the program suite due to Ferguson [21].

18.4.3 Wave-hop field strength prediction theory

At frequencies above about 60 kHz (wavelengths shorter than 5 km), and at lower frequencies when the path length is less than approximately 2000 km, it is no longer appropriate to model the propagation mechanism as a waveguide because of the large number of significant modes. Instead it is more straightforward to use the ray theory to compute the sky-wave field strength and combine this with the surface-wave field strength derived separately using, say, the method introduced in Section 18.3.4.

Figure 18.8 shows a measured variation of field variation with range, together with the theoretical curve for the surface wave only. The oscillation in the measured field is due to interference between the surface wave and one-hop sky-wave modes. The range of oscillation is small at short ranges because the sky wave is small compared with the surface wave. At ranges beyond about 1500 km the sky wave is dominant and the oscillations decay. The interference nulls are generally known as the Hollingworth pattern. The field-strength nulls change location in response to the diurnal changes in sky-wave reflection height.

Sky-wave reflection takes place at heights near 70 km during the day and 90 km at night. In these lower reaches of the ionosphere the variation of electron concentration with height is significant within the wavelength of an LF signal. Under such circumstances the magnetoionic theory applied so successfully at HF is not valid. Waves are returned to the earth by partial, rather than total, reflection.

The reflection coefficients used in a wave-hop prediction procedure may be empirical or based on a theoretical model.

LF sky waves propagate strongly at night and may be almost as strong during the winter days in temperate latitudes. Solar-cycle variations appear to be small.

A complete account of a wave-hop field-strength prediction method is presented by the ITU-R in Recommendation P.684. The basic formula for predicting the effective sky-wave field strength E_s when reception is *via* a small in-plane loop antenna is:

$$E_s = 300\sqrt{P}\,\frac{2}{L}\cos\Psi_{\|}R_{\|}DF_tF_r \tag{18.21}$$

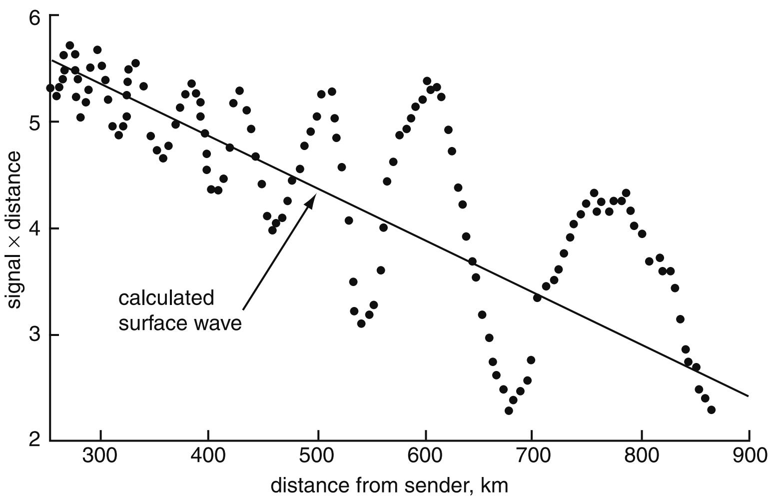


Figure 18.8 Hollingworth interference pattern measured at 85 kHz [22]

where

P = radiated power (kW)
L = sky-wave (slant) path length (km)
$_{\|}R_{\|}$ = ionospheric reflection coefficient which gives the ratio of electric field components parallel to the plane of incidence
D = an ionospheric focusing factor
F_t = a transmitting antenna factor
F_r = a receiving antenna factor
Ψ = angle of departure and arrival of the sky wave at the ground relative to the horizontal

For propagation beyond the one-hop ground distance of about 2000 km it is necessary to compute the signal strength after multiple ionospheric reflections and the intermediate ground reflection(s). For this purpose a more general form of Eqn. 18.21 is

$$E_s = 300\sqrt{\mathrm{P}}\,\frac{2}{L}\cos\Psi\,{}_{\|}R_{g\|}{}^{(n-1)}D^n D_g{}^{(n-1)}F_t F_r \pi^n{}_{i=1\|}R_{g\|} \qquad (18.22)$$

where

n = number of ionospheric hops
$_{\|}R_{g\|}$ = effective reflection coefficient of finitely conducting ground
D_g = divergence factor caused by the spherical Earth (approximately equal to $1/D$).

Recommendation 684 gives graphs of the above factors. An unusual feature of the method is its consideration of negative elevation angles in evaluating antenna factors. A negative elevation angle path corresponds to a geometry where the signal propagates, at each end of a zero-elevation sky-wave hop, by sections of diffraction around the bulge of the spherical earth.

18.4.4 An empirical field-strength prediction theory

In the planning of sound broadcasting systems in the LF and MF bands, the ITU-R recommends the use of an empirical sky-wave field-strength prediction method. A complete description of the method is given in Recommendation ITU-R P.1147 [18]. The salient points of the method are presented here. Sky-wave propagation in these bands differs from that at lower frequencies because the radio waves return to earth by ionospheric refraction rather than partial reflection. Therefore the magnetoionic theory may be applied.

One of the principal characteristics of the frequency band above 150 kHz is that sky waves propagate efficiently at night but are greatly attenuated during the day. This attenuation occurs mainly in the D region of the ionosphere. At sunset the D region ionisation decays rapidly, and waves reflected from the higher E or F regions become significant.

For LF and MF broadcasting the surface wave is the most important propagation mode because it provides a stable signal at all times. Nevertheless, the sky-wave mode is important because it can provide a greatly extended, although inferior quality, nighttime coverage. It can also give rise to troublesome nighttime interference between stations which function independently with the surface-wave mode during the day.

18.4.4.1 General features

As the sky-wave propagates from transmitter to receiver it is subject to various losses. These losses are considered in more detail in the following Sections. In practice, most transmitting and receiving antennas used in the LF and MF broadcasting bands use vertical polarisation, and this is assumed here.

The main ITU-R prediction formula for the annual median sky-wave field strength is:

$$E = P + G_v + G_h + G_s - L_p\, 20 \log p - k_r\,(p/1000)^{0.5} - L_t - L_r \qquad (18.23)$$

where

E = annual median of half-hourly field strengths in dB (V m^{-1})
P = radiated power in dB (kW)
G_v = a transmitting antenna gain factor due to vertical directivity (dB relative to the maximum gain of a small monopole)
G_h = a transmitting antenna gain factor due to horizontal directivity (dB); $G_h = 0$ for an azimuthally omnidirectional antenna
G_s = a 'sea gain' associated with sea near the transmitting and/or receiving antenna (dB)
L_P = a polarisation coupling loss associated with magnetoionic effects (dB)
A = an empirical factor which depends on geomagnetic latitude (dB)
p = slant propagation distance (km)
kr = an empirical loss factor in which is bundled ionospheric absorption, focusing, terminal losses and intermediate ground reflection losses
L_t = an empirical hourly loss factor (dB)
L_r = loss factor incorporating the effects of solar activity (dB)

The prediction equation is applicable between 150 and 1700 kHz and path lengths up to 12 000 km, but should be used with caution for geomagnetic latitudes outside ±60°.

18.4.4.2 Terminal losses and sea gain

The strength of the transmitted wave, and voltage induced in the receiving antenna, are both influenced by ground loss, which would be zero only if the ground were flat and perfectly conducting near the antennas. With flat but finitely conducting ground the interaction of direct and ground-reflected wave gives rise to a ground loss at each terminal of:

$$L_g = 6 - 20 \log |\, 1 + R(\Psi)| \quad 1\text{ dB} \qquad (18.24)$$

where R is the Fresnel plane-wave reflection coefficient for vertically polarised waves at elevation angle Ψ. As the elevation angle approaches grazing, R tends to −1 and the ground loss becomes infinite. When the earth's curvature is taken into account the losses remain large but finite.

Terminal loss factors due to imperfect ground do not appear explicitly in the Recommendation P.1147 prediction method. Instead they are implicitly included for average ground as part of the empirical term k_r.

Sea water has a much higher conductivity than land, with the result that the ground losses for terminals located within a few tens of kilometres of the sea can be much less than those for the average ground included in k_r.

For a terminal located on the coast, a correction factor, known as sea gain G_s, can be computed using reflection coefficients appropriate to first average ground (say a conductivity of 10 mS/m) and then sea. The difference in loss for the two ground types is the sea gain appropriate to one terminal. Figure 18.9 shows the result of such a calculation. Note that sea gain has maxima at ground ranges which are multiples of 2000 km owing to the presence of low-angle signals. At 2000 km the low-angle one-hop E region reflection is dominant, at 4000 km the low-angle two-hop mode dominates and so on. When a terminal is located inland, or the sea only occupies a narrow channel, then the sea gain for the terminal will be reduced. An algorithm to deal with such complications is presented in Recommendation P.1147.

When using the recommended prediction method to compute the coverage of an LF or MF broadcasting station it is relatively straightforward to allow for sea gain at the transmitter. Evaluating sea gain at the receiver is computationally intensive and is only really practical when a digital coastline database is available. In practice G_s at the receiver is often taken to be zero.

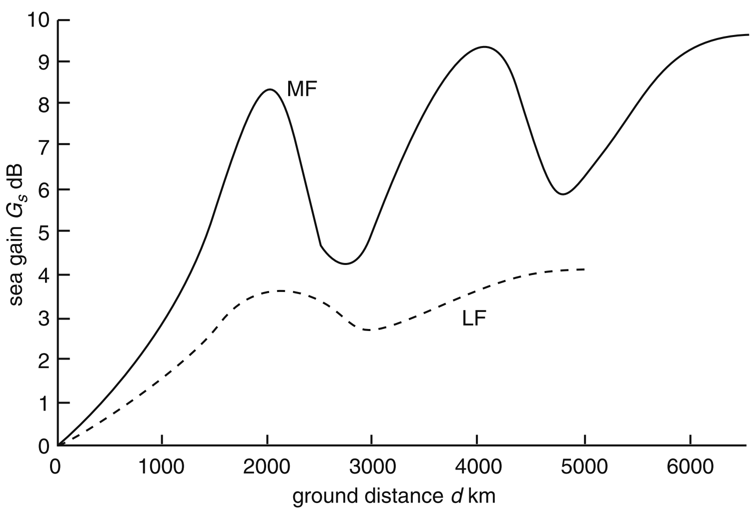


Figure 18.9 Sea gain G_s for a single terminal on the coast [18]

18.4.4.3 Polarisation coupling loss

The ionosphere, a plasma with an embedded magnetic field, is a birefringent medium. A linearly polarised radiowave incident on the bottom of the ionosphere splits into two waves which propagate independently. The division of power between the two waves depends on polarisation of the incident wave relative to the local geomagnetic field. One of these waves, the extraordinary, is more heavily attenuated than its partner, the ordinary wave. In general, on exit from the ionosphere the two waves have elliptical polarisation which may or may not be well orientated to excite a current in the receiving antenna.

The earth's gyromagnetic frequency varies between 800 kHz near the equator and 1600 kHz near the poles, and therefore lies in the MF broadcasting band. At the gyromagnetic frequency the extraordinary wave is almost completely attenuated so that only the ordinary wave contributes to the received signal. This effect also prevails over a band of frequencies either side of the gyrofrequency and the extinction of the extraordinary wave can be assumed in all MF planning. In the LF broadcasting band the effect can be ignored.

When the incident wave excites an extraordinary wave which is subsequently absorbed, this is a loss mechanism, the so-called polarisation coupling loss. A further loss occurs when the elliptically polarised ordinary wave which emerges from the ionosphere excites a current in a receiving antenna which responds only to the vertical component. The two loss mechanisms are essentially the same. The L_p in Eqn. 18.23 is the sum of polarisation coupling losses arising at the transmitting and receiving ends of the propagation path. On long paths it is necessary to compute losses for each terminal separately.

The major axis of the elliptically-polarised ordinary wave accepted by the ionosphere is parallel to the earth's magnetic field. On exit from the ionosphere the major axis is again parallel to the local field. Near the equator, where the earth's field is approximately horizontal, the normal vertically polarised transmitting antenna couples badly into the ordinary mode in directions to the east or west. Similarly, the normal vertically polarised receiving antenna is orthogonal to ordinary-wave energy arriving from the east or west. Thus the polarisation coupling loss per terminal can be very significant and is most extreme on east–west/west–east paths near the equator. Figure 18.10 shows the predicted coverage of an omnidirectional transmitting station on the equator. Instead of roughly circular coverage contours, the foreshortened coverage east and west is due to polarisation coupling losses.

18.4.4.4 Temporal variability

Sky-wave field strength in the LF and MF bands varies on timescales ranging from minutes to years.

Short-period variations, usually Rayleigh distributed and measured in minutes, arise due to continuous turbulence in the ionosphere. Occasionally, when only two sky-wave modes are present, the fading may be more severe.

Short-term median field strength measured on one day will generally differ

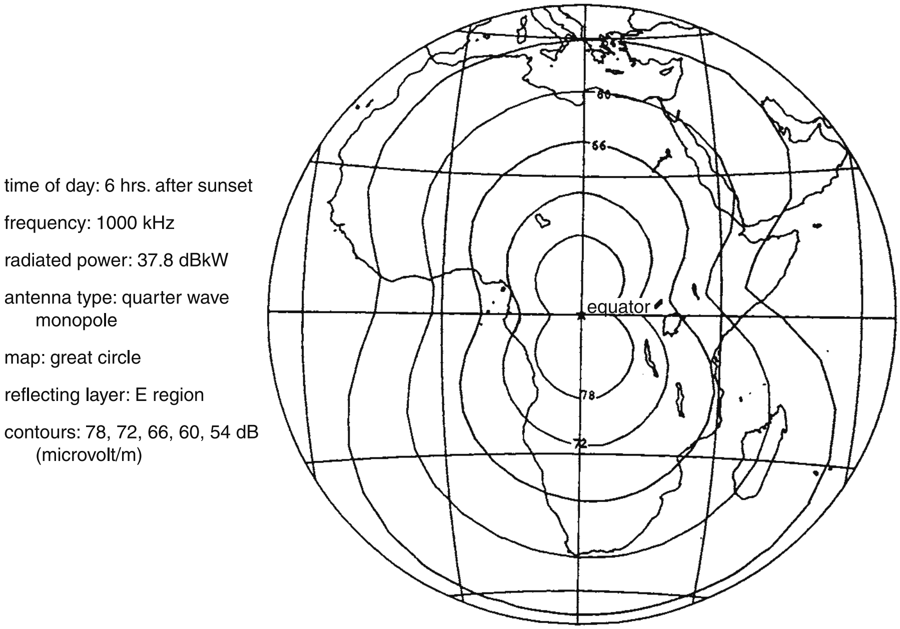


Figure 18.10 Coverage of an MF transmitter on the equator

from an equivalent measurement on the following day. A sequence of such day-to-day measurements can often be approximated by a log-normal distribution.

The diurnal variation in field strength is considerable. The sky-wave field strength is largest late at night and is weak or insignificant during the day. A large number of measurements under a variety of circumstances has permitted the estimation of an hourly average loss factor L_t, the form of which is shown in Figure 18.11. For multihop paths (>2000 km) the times of sunset and sunrise are taken to be those at a point 750 km from the terminal where the sun sets last or rises first, because the remainder of the path is then in darkness.

Seasonal variations in field strength also arise. In the MF band equinoxial months are associated with higher field strengths than others. The overall seasonal variation can be as much as 15 dB at the lower frequencies in the MF band, but is only about 3 dB at 1600 kHz. In the LF band different variations are observed, for example a pronounced summer maximum.

18.5 Antenna efficiency

Radio transmissions in the frequency band below 2 MHz are normally made from a vertical mast, the transmitter being connected between the base of the structure and a radial-wire ground screen. The emission from such an antenna is azimuthally omnidirectional and vertically polarised. Some MF broadcasting stations use two or more masts, located and phased to achieve azimuthal gain.

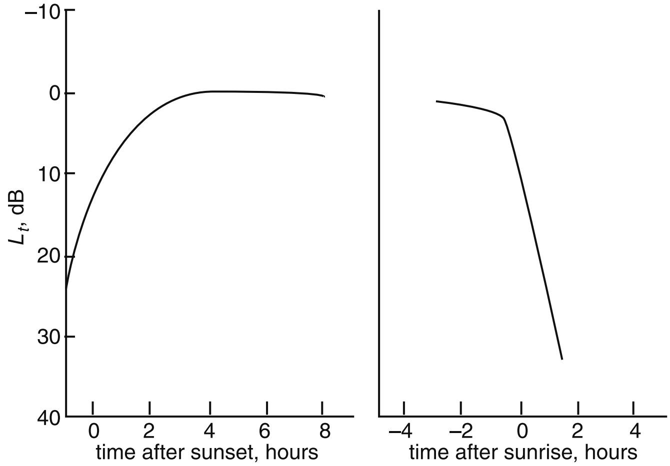


Figure 18.11 Hourly loss factor L_t [18]

In the MF band it is possible to build masts which are a quarter wavelength high. Such an antenna is resonant, with a purely resistive input impedance of 36Ω which can be readily matched to a transmitter.

At frequencies below about 250 kHz it becomes impractical to build masts $\lambda/4$ high. The radiation resistance R_r of a vertical monopole of height h is given by $40\pi^2(h/\lambda)^2$. In the ELF and VLF bands this ratio is small and the radiation resistance of realisable structures can be very modest.

Unfortunately, all of the power output by a transmitter cannot be radiated from the antenna. A portion of the transmitter power is dissipated in other resistive elements of the antenna system such as the series-dielectric resistance R_{sd}, copper-loss resistance R_c, load-coil resistance R_i and ground-system resistance R_g. The efficiency e of the antenna system is given by the ratio:

$$e = R_r/(R_r + R_{sd} + R_c + R_i + R_g) \tag{18.25}$$

A second consideration is the reactive nature of the antenna impedance. This gives rise to large voltages at the base of the antenna and voltage breakdown can set a limit on the radiated power.

These limitations can be eased using antenna systems with a capacitive top loading to maximise R_r, and an extensive ground screen and low-loss components to minimise other resistances. The extent to which such measures are taken depends on the operational requirement and the economics of running the radio station. When a heavy investment in antenna construction is justified, moderately high efficiences can be achieved, even in the VLF band. For example, the VLF antenna system at Cutler, Maine, USA, has an efficiency of 86 per cent at 20 kHz. The structure is enormous, see Watt [23]:

maximum mast height: 298 m
average physical height: 201 m

number of masts:	26
area of the top loading:	2.25 km^2
ground screen:	3300 km of copper wire

The radiation resistance of this antenna is a mere 0.15 Ω.

18.6 Surface-wave/sky-wave interactions

At short ranges from a transmitter the surface-wave mode dominates. At long ranges the sky-wave mode will normally deliver the strongest signal. At intermediate ranges the surface wave and sky wave may be comparable and the interference between the two will lead to a region of signal peaks and troughs. Unlike the sky wave, the surface wave has a stable phase. Diurnal and short-term variations in the height and structure of the lower ionosphere cause the signal peaks and troughs to shift position. Gross variations in the sky-wave signal strength causes the region of interplay between the two modes to move towards or away from the transmitter.

For the long wavelengths of the VLF/LF bands the surface-wave/sky-wave interaction gives rise to the relatively stable Hollingworth pattern referred to in Section 18.4.3. The systems required to operate in the interference region must be designed to function within the minima of the pattern.

In the MF band the region of interaction is called the night fading zone (NFZ). The location of the NFZ depends on any factor which affects the relative amplitude of the surface and sky waves, e.g. antenna vertical radiation pattern, time of night, ground conductivity and polarisation. Other factors such as radiated power have no effect. The fading is especially disruptive to broadcast reception and in planning such systems it is prudent to estimate the NFZ location. A fading zone which, night after night, resides over an important urban area will soon cause listeners to retune their receivers. The NFZ is usually taken to be the region in which the difference between surface- and sky-wave field strengths is 8 dB or less. When designing a broadcasting station so that the NFZ is not too troublesome, it must be borne in mind that nighttime interference from other cochannel transmitters may, in fact, have a more significant effect on the quality of reception.

18.7 Background noise

The performance of any radio system is determined, in part, by the level of background noise with which the wanted signal must compete. This can be generated within the receiver or arrive at the receiver input terminals *via* the antenna. The four potential sources of background noise originate in the receiver itself, in the galaxy, from atmospheric lightning discharges and from man-made electrical equipment (see Chapter 2). Galactic noise will normally be reflected by the top side of the ionosphere at the frequencies considered here and can be ignored. For most applications the level of receiver noise is also insignificant compared with the atmospheric and man-made contributions. An exception is in submarine communication where the atmospheric and above-surface man-made noise contributions may be both heavily attenuated by the sea so that the receiver noise contribution becomes dominant. Chapter 2 also discusses the characteristics of radio noise.

18.8 Acknowledgments

Parts of the text in this Chapter concerning sky-wave propagation are based on material originally written by Dr. Philip Knight [24] for a previous edition of this book. Figures 18.9 and 18.11 are used with the kind permission of the ITU as copyright holder.

18.9 References

1 SOMMERFELD, A.: 'The propagation of waves in wireless telegraphy', *Ann. Phys.*, 1909, **28**, p.665
2 NORTON, K.A.: 'The propagation of radio waves over the surface of the Earth and in the upper atmosphere. Part 1', *Proc. Inst. Radio. Eng.*, 1936, **24**, pp.1367–1387
3 NORTON, K.A.: 'The propagation of radio waves over the surface of the Earth and in the upper atmosphere. Part 2', *Proc. Inst. Radio. Eng.*, 1937, **25**, pp. 1203–1236
4 NORTON, K. A.: 'The calculation of ground-wave field intensity over a finitely conducting spherical Earth', *Proc. Inst. Radio Eng.*, 1941, **29**, pp.623–639
5 VAN DER POL, B., and BREMMER, H.: 'The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere', *Philos. Mag. Ser. 7*, 1937, **24**, pp.141–176
6 VAN DER POL, B., and BREMMER, H.: 'The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere', *Philos. Mag. Ser. 7*, 1937, **24**, pp.825–864
7 VAN DER POL, B., and BREMMER, H.: 'The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere', *Philos. Mag. Ser. 7*, 1938, **25**, pp.817–834
8 VAN DER POL, B., and BREMMER, H.: 'The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere', *Philos. Mag. Ser. 7*, 1939, **26**, pp.261–275
9 BREMMER, H.: 'Terrestrial radio waves' (Elsevier, 1949)
10 ROTHERAM, A.S.: 'Ground-wave propagation. Part 1 Theory for short distances', *IEE Proc. F*, 1981, **128**, pp.275–284
11 ROTHERAM, A.S.: 'Ground-wave propagation. Part 2 Theory for medium and long distances and reference propagation curves', *IEE Proc. F*, 1918, **128**, pp.285–295

12 ROTHERAM, A.S.: 'Ground-wave propagation', *Marconi Rev.*, 1982, **45**, (1), pp.18–48
13 ITU-R Recommendation P.368-7: 'Ground-wave propagation curves for frequencies between 10 kHz and 30 MHz', 1994
14 ITU-R Recommendation P.341-5: 'The concept of transmission loss for radio links', 1999
15 ITU-R Recommendation P.832-2: 'World atlas of ground conductivities', 1999
16 MILLINGTON, G., and ISTED, G.A.: 'Ground-wave propagation over an inhomogeneous Earth. Part 2 Experimental evidence and practical implications', *Proc. IEE*, 1950, **97**, pp.209–222
17 CAUSEBROOK, J.H.: 'Medium-wave propagation in built-up areas', *Proc. IEE*, 1978, **125**, pp.804–808
18 ITU-R Recommendation P.1147: 'Prediction of sky-wave field strength at frequencies between 150 and 1700 kHz', 1999
19 ITU-R Recommendation P.684-2: 'Prediction of field strength at frequencies below about 500 kHz', 2001
20 WAIT, J. R.: 'Electromagnetic waves in stratified media' (Pergamon Press, New York, 1962)
21 FERGUSON, J.A.: 'Status of the naval ocean systems long wave propagation capability "LWPC"'.Paper presented at *Third Workshop on ELF/VLF Radio Noise*, Stanford University, USA, 1988
22 BRACEWELL, R.N., *et al.*: 'The ionospheric propagation of low- and very-low-frequency radio waves over distances less than 1000 km', *Proc. IEE*, 1951, **98**, Part III, pp. 221–236
23 WATT, A.D.: 'VLF Radio Engineering' (Pergamon Press, New York, 1967)
24 KNIGHT, P., *in* HALL, M.P., and BARCLAY, L.W.: 'Radiowave propagation', IEE Electromagnetic Waves Series 30 (Peter Peregrinus, London, 1989)

Chapter 19

Terrestrial line-of-sight links: Recommendation ITU-R P.530

R.G. Howell

19.1 Introduction

Recommendation ITU-R P.530 [1] presents propagation data for planning terrestrial line-of-sight radio systems, typically operating in frequency bands between 1 GHz and 45 GHz, and gives step-by-step methods for predicting the performance of these systems for percentage times down to 0.001 per cent. The recommendation deals only with effects related to the wanted signal. The interference aspects of terrestrial line-of-sight link design are covered in Recommendation ITU-R P.452 [2].

Although the prediction methods presented in Recommendation ITU-R P.530 are based on the physics of radio propagation, measured data was used as the main foundation and the majority of the formulae are semi-empirical statistical relationships.

The Recommendation divides broadly into methods for determining the antenna heights (with and without antenna height diversity) and methods for calculating the statistics of signal fading, enhancement and depolarisation. In the case of clear air effects the methods are subdivided into a quick (initial planning) method and a more detailed method, which requires terrain data for the area of the link.

Other topics covered are: the correlation of simultaneous fading on multihop links, the frequency and polarisation scaling of rain attenuation statistics, the statistics of rain fade duration and the prediction of the error performance and availability of digital systems.

19.2 Overview of propagation effects

The main propagation effects that must be considered in the design of terrestrial line-of-sight links are:

- free-space loss
- attenuation due to atmospheric gases
- diffraction fading due to obstruction of the radio path by the terrain or obstacles on the terrain
- signal fading and signal enhancement due to atmospheric refractive effects
- signal fading due to multipath arising from surface reflection
- attenuation due to precipitation or solid particles in the atmosphere
- variation of the angle-of-arrival at the receive antenna and angle-of-launch at the transmit antenna due to atmospheric refraction
- reduction in the system cross-polar performance in multipath or precipitation conditions
- signal distortion due to frequency selective fading and delay during multipath propagation.

The free-space loss and the losses produced by dry air and water vapour have been described in Chapters 2 and 7. Further details on the losses produced by atmospheric gases can be found in Recommendation ITU-R P.676 [3]. The following sections will discuss the other propagation effects having an impact on the link design and methods for predicting the performance of a link taking account of these effects.

19.3 Link design

19.3.1 General

The performance of a link cannot be determined until all the relevant link parameters, e.g. frequency, path length, antenna heights, meteorological factors for the area etc., are known. Therefore, the design of a link is an iterative process in which an initial design is examined and then, if necessary, changed until, hopefully, a design which meets the required performance objectives is achieved.

In a practical situation the general location, direction and capacity of the link are the primary (fixed) requirements. The main variables in the link design are the frequency band available, the locations of the link end points (usually determined by the location of suitable high points with a clear view in the required directions and the availability of sites for which planning consents can be obtained), the height, number and gains of the antennas, the transmitter powers and the receiver sensitivities and noise figures.

The constraint on the general location of the link determines the meteorological factors e.g. the incidence of extreme clear air radio refractive effects and the incidence of high rainfall rates. The constraints on the location of the end points then determines, at least initially, the path length and the terrain type and profile over which the radio path must pass.

Having chosen the frequency band and the end points of the link, the main factors determining the link design then relate to the antennas (number, heights and gains), transmitter powers and receiver characteristics (sensitivity and noise figure). Given the constraints on transmitter power required to avoid interference to other systems and the practical limitations on the receiver characteristics, the link design then centres around:

- Achieving sufficient clearance of the radio path above the terrain such that any losses produced by the terrain (e.g. due to diffraction) are within acceptable limits, both under average clear air conditions and during conditions of abnormal atmospheric radio refractivity (e.g. under conditions of subrefraction).
- Determining the frequency separation and/or antenna spacing and/or angular separation if diversity is required to achieve the required performance (see Section 19.5).

Having carried out the initial link design and estimated its performance it may be that, even with fading mitigation methods (see Section 19.5), the design target is not achieved and new link end points, or a different frequency band, must be considered.

19.3.2 Path clearance

19.3.2.1 Nondiversity case

Two basic concepts are relevant to the determination of path clearance. The first is the bending of the radio beam produced by variations in the radio refractive index of the atmosphere with height (see Chapter 7) and the second is that of Fresnel zones enclosing the path (see Chapters 3 and 8).

It was shown in Chapter 7 that the decrease in radio refractive index with increasing height under average clear air conditions bends the radio path downwards with a curvature that is less than that of the earth's surface. Hence, a plot of the path over the earth's surface results in a curved radio path over a curved terrain profile. Transforming this plot so as to change the curved radio path into a straight line results in a flatter terrain profile (i.e. an increased earth radius) and this transformation is often used to simplify the task of determining the clearance of the radio path over the terrain. The factor by which the earth's radius must be multiplied to achieve this (the k factor) is typically about 4/3 (see Figure 19.1). When designing a terrestrial line-of-site link radio refractive index, changes in the lowest 100 m of the atmosphere are of the most significance in determining the k factor to be used.

If, instead of decreasing with height, the radio refractive index increases (as can happen under atmospheric conditions occurring for about 0.1 per cent of the worst month), the radio path curves upwards and to produce a path which is

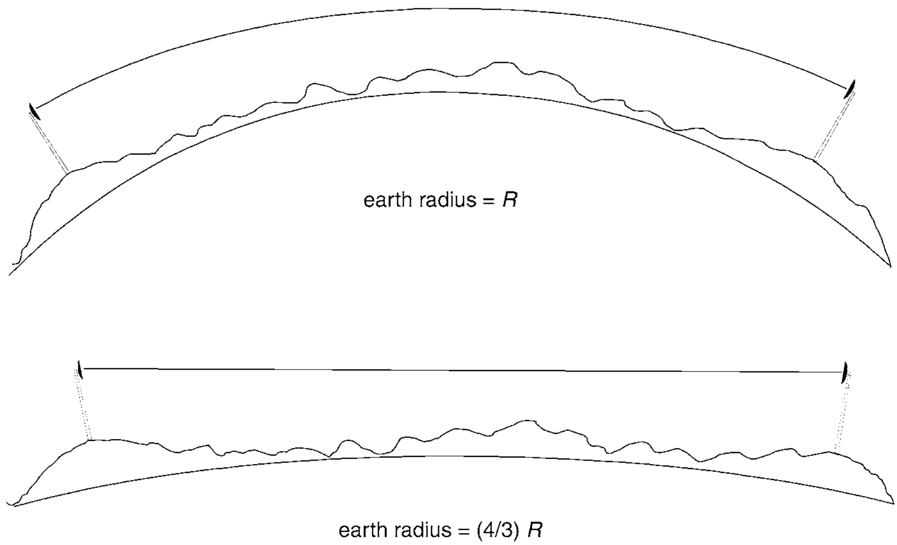


Figure 19.1 The use of a modified earth radius to simplify the determination of the required path clearance

a straight line the earth's radius must be reduced i.e. the terrain becomes more curved. Under these circumstances, i.e. subrefraction, the k factor might typically be about 2/3. The effect is as if the terrain is bulging up towards the radio path and hence there is an increased risk of partial or complete blockage of the signal.

Determination of the transmit and receive antenna heights centres on achieving sufficient clearance of the path above the terrain to avoid complete or partial blockage of the signal under both the average and the subrefractive conditions while taking account of the practical constraints on the antenna heights.

It was shown in Chapters 4 and 5, in the discussion of diffraction, that a radio opaque obstacle progressively moved towards a radio path does not produce an abrupt loss of signal but produces a ripple in the signal level. This results in signal levels both above and below the free-space level as the obstacle approaches the path, followed by a monotonic decrease in level below the free-space value before the edge of the obstacle reaches the path. This monotonic decrease in signal level then continues as the obstacle continues towards the path and after the obstacle protrudes into the path. The important points here are that the obstacle has an effect on the signal level even though the object is clear of the path and that the signal never completely disappears however much the object penetrates into the path, although for all practical purposes the signal level would be too low to be of use after the object has penetrated significantly into the path.

The key concept in discussing the required clearance is that of the Fresnel zones around the path (see Chapter 4). In the above discussion on the penetration of the obstacle into the path, significant could be considered as the point at which the obstacle penetrates more than about one Fresnel zone into the path.

It should be noted that the shape of the edge of the obstacle has a very significant effect on the diffraction loss. For example, with the path just grazing the edge of a sharp obstacle (e.g. a knife edge) the diffraction loss is about 6 dB compared with about 15 dB for an object having a large radius of curvature (e.g. a smooth earth surface). A clearance of about 60 per cent of the first Fresnel zone is required to achieve a free-space signal level.

In the method given in Recommendation ITU-R P.530 transmit and receive antenna heights are determined which, under average refractivity conditions ($k = 4/3$), give a clearance above the most significant obstacle for the whole of the first Fresnel zone. Then, the antenna heights required to give a clearance of a fraction (varying between 0 and 0.6, depending on the climate and the type of obstruction) of the first Fresnel zone for the conditions of subrefraction likely to occur for 0.1 per cent of the worst month are determined. The value of k to be used for the subrefractive condition (an effective k value k_e) should be determined from radio refractivity data averaged along the path and hence it is dependent on the path length. The effective k value, as a function of path length, is given in Recommendation ITU-R P.530 in the form of a graph. Finally, having determined the transmit and receive antenna heights for the above two conditions, the larger of the two pairs of values is chosen.

Larger path clearances than those determined by the above method for the worst month subrefractive condition may be necessary for frequencies above 10 GHz in order to reduce the risk of diffraction loss during conditions of subrefraction. Therefore, the diffraction loss should be calculated for this condition (e.g. using $k = 2/3$ and the methods given in Chapter 8) and, if necessary, the antenna heights increased.

Having determined the required antenna heights the link performance, taking account of the propagation conditions along the path, can now be estimated. If the performance target cannot be achieved using single antennas at each end of the link the improvement provided by various mitigation techniques can be examined (see Section 19.5).

19.3.2.2 Height diversity case

A method of reducing the link outage if the performance target cannot be achieved using single antennas is to use two vertically spaced antennas at one or both ends of the link, i.e. using height diversity. The principle behind the use of height diversity is that if the variation of signal strength with height is examined during multipath fading, peaks and troughs are observed. Hence, with suitable antenna spacing, a trough at one antenna will coincide with a peak at the other antenna and, with an appropriate signal combining mechanism, the system will suffer less overall fading. Furthermore, by pointing the two antennas at slightly different angles of elevation a further advantage can be obtained from the variations in the angle of arrival of the signal at the two antennas (angle diversity) i.e. when the angle of arrival at one antenna is not optimum (not on boresight) it may be closer to optimum at the other. Therefore, the design task is to determine a suitable spacing and angular offset between the two antennas.

In the method given in Recommendation ITU-R P.530 the initial height of the upper antenna is determined by the method described in Section 19.3.2.1. It may be necessary to change this height after determining the link outage probability with diversity. The height of the second (lower) antenna is that which gives a path clearance, under median k value conditions, of between 0.6 and 0.3 of the first Fresnel zone for path obstructions extending along a portion of the path and between 0.3 and 0 if there are only one or two isolated obstacles. In the absence of a measured median value for k a value of 4/3 should be used. Alternatively, the clearance of the lower antenna may be chosen to give about 6 dB of diffraction loss during normal refractivity conditions, or some other loss appropriate to the fade margin of the system.

If, after calculating the link outage with the antenna spacing derived above, the link performance is still not satisfactory some further improvement may be obtained by increasing the height of the upper antenna and/or adjusting the angular offset between the two antennas. Methods for calculating the improvement given by these methods are given in Recommendation ITU-R P.530.

19.4 Prediction of system performance

19.4.1 Propagation in clear air

The physical basis for the effects on radio propagation of variations in the atmospheric radio refractive index in clear air are discussed in detail in Chapter 7. In this Section an overview of the methods given in Recommendation ITU-R P.530 for determining the statistics of signal fading and enhancement produced by these effects is given. The reader is referred to the Recommendation for the details of the step-by-step algorithms and graphical methods.

19.4.1.1 Multipath fading on a single hop of a link

Two methods are given in Recommendation ITU-R P.530 for predicting the statistics of multipath fading on a single hop of a link:

(i) a quick method, suitable for planning applications, which does not require detailed information on the terrain;
(ii) a more detailed method which requires the standard deviation (s_a) of the terrain heights in the region of the path.

Both methods require the point refractivity gradient in the lowest 65 m of the atmosphere not exceeded for one per cent of an average year (dN_1 N km^{-1}). If measured point refractivity data for the region of the link is not available dN_1 can be obtained from Recommendation ITU-R P.453 [4].

These two methods are further sub-divided into:

- a method applicable to the small percentage time/deep-fading (> about 25 dB) region of the distribution;
- a method applicable to all percentage times and fade depths down to 0 dB.

First, the fading statistics for an average worst month are determined and then, if necessary, these can be used to derive statistics for an average year.

The key parameter used in all the methods is a geoclimatic factor K applicable to an average worst month. Formulae for calculating this from dN_1 and, in the case of the detailed method from s_a in addition to dN_1, are given. Alternatively, a method for deriving K from measured fading data is given in Appendix 1 to the Recommendation.

In the method applicable to the small percentage time/deep-fading region of the distribution K is used, in conjunction with the link path length (d km), the path inclination, the frequency and the altitude of the lower of the two antennas, to calculate directly the percentage time p_w that a fade of a given depth A dB is exceeded in an average worst month.

In the method applicable to all percentage times the deep-fading distribution derived above is combined with an interpolated distribution covering fade depths down to 0 dB. First, a multipath occurrence factor p_0 is calculated, which is then used to determine the fade depth at which the transition between the deep-fading distribution and the shallow-fading distribution occurs. For fade depths greater than this value the small percentage time/deep-fading distribution is used. For fade depths less than the transition value an interpolation procedure is given.

In both cases either a step-by-step algorithmic method or a graphical method can be used (the required graphs are given in the Recommendation). An approximate value for the lower frequency limit for which the methods are valid is given by $15/d$ GHz.

19.4.1.2 Simultaneous clear-air fading on multihop links

Measurements indicate that, in clear-air conditions, fading events exceeding 20 dB on adjacent hops of a multihop link are almost completely uncorrelated. This suggests that, for analogue systems with large fade margins, the outage time for a series of hops in tandem is approximately given by the sum of the outage times for the individual hops.

For fade depths not exceeding 10 dB, the probability of simultaneously exceeding a given fade depth on two adjacent hops can be estimated from:

$$P_{12} = (P_1\,P_2)^{0.8} \tag{19.1}$$

where P_1 and P_2 are the probabilities of exceeding this fade depth on each individual hop.

The correlation between fading on adjacent hops decreases with increasing fade depth between 10 dB and 20 dB, so that the probability of simultaneously exceeding a fade depth of greater than 20 dB can be approximately expressed by:

$$P_{12} = P_1 P_2 \tag{19.2}$$

19.4.1.3 Signal level enhancement

Large signal level enhancements can occur during the same general conditions that result in multipath fading. A procedure, similar to that described above covering fading for all percentage times, is used to determine the distribution of signal level enhancement.

First, a distribution valid for enhancements greater than 10 dB, and derived using the fade depth exceeded for 0.01 per cent of the average worst month (using the methods described above), is produced. Then, using an interpolation procedure, the distribution of enhancements between 10 dB and 0 dB is obtained and combined with the distribution for enhancements greater than 10 dB to obtain a distribution covering all enhancement levels.

19.4.1.4 Conversion from average worst month to average annual distributions

It should be noted that the methods discussed above for predicting clear air signal fading and enhancement give statistics for the average worst month. In some cases link performance criteria are specified in terms of both average worst month and average annual percentage times. Therefore, methods for converting from the average worst month distribution to the average annual distribution are of use to the link designer. Recommendation ITU-R P.530 provides methods for carrying out this conversion.

The annual distribution is obtained by combining two distributions, one derived from the deep-fading range of the average worst month distribution and the other derived from the shallow-fading range of the average worst month distribution. In the deep-fading range a logarithmic geoclimatic conversion factor ΔG, which is a function of the link latitude, path length and path inclination, is used to derive the average annual percentage time p from the average worst month percentage time p_w.

Methods are also given in the Recommendation for the conversion of the average worst month distributions of fading to shorter worst periods in the range from 1 h to 720 h and for the conversion of the average worst month enhancement distribution to an average annual enhancement distribution.

19.4.1.5 Reduction of system cross-polar performance due to clear air effects

Cross-polar discrimination (XPD) is a measure of the separation between orthogonally (typically vertical and horizontal) polarised signals that can be achieved by an antenna. A reduction in the XPD of an antenna (for example, produced by off-axis reception or transmission) can result in cochannel interference and, to a lesser extent, adjacent channel interference. The effect of multipath propagation on the arrival or transmission angle of signals combined with the antenna cross-polar patterns determines the reduction in the system XPD for small percentage times and clear-air propagation conditions.

A method for estimating the probability of outage due to this effect is given in Recommendation ITU-R P.530. The method is based on a multipath activity parameter which is closely related to the frequency of occurrence of multipath in an average worst month. This is then used to calculate the probability of outage, taking into account the number of antennas and their spacing (in the case of space diversity) and, if used, the effect of cross-polar cancellation (see Section 19.5.3).

19.4.2 Propagation in rain

19.4.2.1 Attenuation by rain

The physics of attenuation by rain has been discussed in detail in Chapter 12. This Section presents an overview of the method given in Recommendation ITU-R P.530 for predicting the long-term statistics of rain attenuation.

The primary input to the procedure is the rainfall rate $R_{0.01}$ (mm h^{-1}), averaged over a period of 1 min and exceeded for 0.01 per cent of the time for the region of the link. If this information cannot be obtained from long-term measurements carried out in the location of the link an estimate can be obtained from rainfall rate maps given in Recommendation ITU-R P.837 [5].

An effective path length, which is less than the actual path length and which reduces with increasing value of $R_{0.01}$, is calculated. This takes account of the nonuniformity of the rainfall rate along the path, particularly for high rainfall rates e.g. >20 mm h^{-1}.

An estimate of the path attenuation exceeded for 0.01 per cent of the time is then obtained by multiplying the effective path length by the rain-specific attenuation (dB km^{-1}) for a rainfall rate of $R_{0.01}$ mm h^{-1}. The rain-specific attenuation is given in Recommendation ITU-R P.838 [6] as a function of frequency, polarisation and rainfall rate.

Two formulae are given in the recommendation for deriving the path attenuation exceeded for percentage times in the range 1 to 0.001 per cent from the 0.01 per cent value. One applies to links between 30° north and 30° south and the other applies to links outside this region.

It should be noted that the propagation model described above gives the rain attenuation statistics for an average year. System planning often requires statistics for an average worst month. These can be derived from the average annual statistics using methods given in Recommendation ITU-R P.841 [7].

At frequencies where both rain attenuation and multipath fading must be taken into account the exceedance percentages for a given fade depth corresponding to each of these mechanisms can be added.

19.4.2.2 Frequency and polarisation scaling of rain attenuation statistics

When reliable long-term rain attenuation statistics are available at one frequency

(f_1) an estimate of the statistics for a link operating at another frequency (f_2), having the same hop length and in the same climate, can be made using the relationship:

$$A_2 = A_1 \, (\Phi_2/\Phi_1)^{1-H} \tag{19.3}$$

where

$$\Phi(f) = \frac{f^2}{1 + 10^{-4} f^2} \tag{19.4}$$

$$H = 1.12 \times 10^{-3} (\Phi_2/\Phi_1)^{0.5} (\Phi_1 A_1)^{0.55} \tag{19.5}$$

and A_1 and A_2 are the equiprobable values of the excess rain attenuation at frequencies f_1 and f_2 (GHz), respectively.

Similarly, when reliable long-term rain attenuation statistics are available for a link, either for vertical (V) or horizontal (H) polarisation, the attenuation statistics for the same frequency and the same link, but for the other polarisation, may be estimated using the relationships:

$$A_V = \frac{300 A_H}{335 + A_H} \tag{19.6}$$

or

$$A_H = \frac{335 A_V}{300 + A_V} \tag{19.7}$$

where A_V and A_H are equiprobable values in dB.

19.4.2.3 Statistics of the number and duration of rain fades

There is strong evidence that the duration of rain fades is much longer than that of multipath fades. If an event is defined to be the exceedance of attenuation A (dB) for a given period of time (τ) or longer then the relationship between the number of fades $N_\tau(A)$ and A is given by:

$$N_\tau(A) = a \, A^b \tag{19.8}$$

where a and b are coefficients that are expected to depend on frequency, path length and other variables such as climate. Measurements carried out in Scandinavia on a 15 km path at 18 GHz for a one-year period indicate that for $\tau = 10$ s, $a = 5.7 \times 10^5$ and $b = -3.4$.

If $T(A)$ is the total time for which an attenuation of A dB is exceeded and $D_m(A)$ is the mean duration of the fades contributing to this total time then:

$$D_m(A) = T(A)/N_\tau(A) \tag{19.9}$$

19.4.2.4 Rain fading on multihop links

If the occurrence of rainfall were statistically independent of location, then the overall probability of fading P_T for a linear series of hops in tandem would be given to a good approximation by summing the individual probabilities i.e.:

$$P_T = \sum_{i=1}^{n} P_i \tag{19.10}$$

where P_i is probability for the ith of the total of n hops.

Alternatively, if rain events are correlated over distances comparable to two or more hops in tandem then the combined fading probability for a multihop link is given by:

$$P_T = K \sum_{i=1}^{n} P_i \tag{19.11}$$

where K is a factor (<1) that includes the effect of the spatial correlation of the rainfall and is a function of the general orientation of the link relative to the prevailing direction of rainstorms in the area, the hop length, the total number of hops and the percentage exceedance time of interest. Graphs, giving a series of curves for K, are given in Recommendation ITU-R P.530.

19.4.2.5 Depolarisation by rain

The polarisation state of signals transmitted through rain may be changed because of the nonspherical shape of large raindrops in heavy rain (see Chapter 12). This reduces the cross-polar discrimination (XPD) of the system and may cause outage due to cochannel interference and, to a lesser extent, adjacent channel interference. The effect is least for vertically or horizontally polarised signals (the normal case for terrestrial line-of-sight links) and greatest for linearly polarised signals at 45° to the horizontal or for circularly polarised signals. However, some reduction of XPD still occurs even for vertically or horizontally polarised signals due to the tilting of the axis of symmetry of the raindrops away from the vertical produced by wind shear near the ground.

Recommendation ITU-R P.530 gives a method for estimating the statistics of XPD from those of the attenuation measured on the link. The basis of the method is the relationship observed between the equiprobable values of XPD and copolar attenuation (CPA) derived from concurrent long-term measurements. This takes the form:

$$XPD = U - V(f) \log CPA \tag{19.12}$$

where

$$U = U_0 + 30 \log f \tag{19.13}$$

and

$$V(f) = 12.8 f^{0.19} \qquad \text{for } 8 \le f \le 20 \text{ GHz} \tag{19.14}$$

$$V(f) = 22.6 \qquad \text{for } 20 < f \le 35 \text{ GHz} \tag{19.15}$$

A mean value for U_0 of about 15 dB (with a lower bound of 9 dB) is suggested for fades in excess of 15 dB.

19.5 Mitigation methods

The effects of slow, relatively nonfrequency selective clear-air fading (i.e. flat fading) due to beam spreading, and the faster frequency-selective clear-air fading due to multipath propagation can be reduced by both nondiversity and diversity techniques.

19.5.1 Techniques without diversity

Several methods can be employed to reduce the effects of multipath fading without the need for diversity. These are:

(a) *Reducing the level of ground reflections*: links should be sited where possible to reduce surface reflections, thus reducing the occurrence of multipath fading and distortion. Techniques for achieving this include the siting of the ends of over-water links at locations which place surface reflections on land rather than water and the siting of over-land and over-water links so as to avoid large flat highly reflecting surfaces on land. Another technique is to tilt the antennas upwards slightly, thus reducing the gain of the antennas in the direction of the reflection; however, in this case a trade-off must be made between the resultant loss in effective antenna gain under normal refractivity conditions and the reduction of multipath fading.

(b) *Increasing the path inclination*: links should be sited to take advantage of the terrain in ways that will increase the path inclination, since increasing path inclination reduces the effects of beam spreading, surface multipath fading and atmospheric multipath fading. In addition, the positions of the antennas on the radio link towers should be chosen to give the largest possible inclinations, particularly for the longest links.

(c) *Reducing the path clearance*: reducing the path clearance can reduce the incidence of multipath fading, however, a trade-off must be made between the reduction in the effects of multipath fading and distortion and the risk of increased fading in conditions of subrefraction.

19.5.2 Diversity techniques

Diversity techniques include space, angle and frequency diversity. Frequency diversity should be avoided whenever possible so as to conserve spectrum. Whenever space diversity is used, angle diversity should also be employed by tilting the antennas at different upward angles. Angle diversity can be used in situations in which adequate space diversity is not possible or to reduce tower heights.

The amount of improvement afforded by all of these techniques depends on the extent to which the signals in the diversity branches of the system are uncorrelated. For narrowband analogue systems, it is sufficient to determine the improvement in the statistics of fade depth at a single frequency. For wideband digital systems, the diversity improvement also depends on the statistics of in-band distortion.

The diversity improvement factor, I, for a fade depth A is defined by:

$$I = p(A)/p_d(A) \tag{19.16}$$

where $p_d(A)$ is the percentage of time in the combined diversity signal branch with a fade depth larger than A and $p(A)$ is the percentage time for the unprotected path. The diversity improvement factor for digital systems is defined by the ratio of the exceedance times for a given BER with and without diversity.

Recommendation ITU-R P.530 gives methods for predicting the diversity improvement factor and the outage when space diversity, frequency diversity, angle diversity and combinations of these are used.

19.5.3 Cross-polar cancellation

A method for reducing the interference that may result from the depolarised component of a signal appearing in the orthogonally polarised channel is to take a small proportion of that signal (equal in amplitude to the interference) and add it in antiphase to the signals (wanted and interference) in the orthogonally polarised channel. Using this technique some or all of the interference can be cancelled out.

19.6 References

1 Recommendation ITU-R P.530-8: 'Propagation data and prediction methods required for the design of terrestrial line-of-sight systems', 1999
2 Recommendation ITU-R P.452-9: 'Prediction procedure for the evaluation of microwave interference between stations on the surface of the earth at frequencies above about 0.7GHz', 1999
3 Recommendation ITU-R P.676-4: 'Attenuation by atmospheric gases', 1999
4 Recommendation ITU-R P.453-7: 'The radio refractive index: its formula and refractivity data', 1999
5 Recommendation ITU-R P.837-2: 'Characteristics of precipitation for propagation modelling', 1999
6 Recommendation ITU-R P. 838-1: 'Specific attenuation model for rain for use in prediction methods', 1999
7 Recommendation ITU-R P. 841-1: 'Conversion of annual statistics to worst-month statistics', 1999

Chapter 20

The principal elements of Recommendation ITU-R P.452

Tim Hewitt

20.1 Introduction

Nowadays, many new technologies and additional users are seeking access to an increasingly busy radio spectrum. An evolving propagation and interference modelling capability is clearly needed to support the necessary spectrum management changes and to resolve the spectrum sharing and system deployment issues that arise within a multioperator, multiservice situation.

In the middle of the 1980s it became clear that the methods then available to model interference propagation at microwave frequencies would not be adequate for the tasks which lay ahead. Considerable effort was therefore invested in devising more flexible and reliable means of predicting microwave interference. These notes provide an introduction to one such development, a new set of propagation models developed by a major European collaborative project (COST 210 [1]), and subsequently adopted in 1992 by ITU-R Study Group 3 as Recommendation ITU-R P.452 [2]. Since then the ITU-R Study Group has progressively extended the capabilities and accuracy of the original COST 210 method to cope with new interference scenarios and to provide full global applicability.

20.2 The COST 210 approach to developing a new prediction procedure

Clear-air interference propagation embraces a mix of propagation mechanisms, most of which involve complex interactions between the atmosphere and the earth's surface. The distances over which interference levels can be significant (several hundred kilometres in the case of the anomalous propagation modes), with conditions that vary with time and in different locations. It was (and remains) unrealistic to consider modelling the propagation environment on a wholly theoretical basis, and semi-empirical modelling techniques had to be employed. However, it was possible to ensure that these models addressed the correct physical phenomena by supporting the modelling activities, with more fundamental propagation research backed up by propagation measurements. The research programme supporting the COST210 interference prediction procedure therefore comprised three complementary elements: radio meteorological analysis, propagation experiments and propagation modelling.

20.3 Radio meteorological studies

These studies, which primarily addressed the clear-air effects, were undertaken to identify the most important anomalous propagation mechanisms to be addressed by the propagation models, and to indicate how their relative importance varied over inland, sea and coastal areas. One particular study [3] looked at the atmospheric (rather than propagation) processes that give rise to short-term interference propagation conditions, and resulted in models of inland, coastal and anomalous propagation which were confirmed using the measured radio data. This work also showed how mixed (e.g. land plus sea) paths could be characterised to determine the dominant radio meteorological conditions.

Other studies yielded data which allowed the development of a location-dependent effective earth radius (or k factor) model to allow diffraction modelling across a range of time percentages. This model was derived from a combination of European experimental radio measurements and radiosonde data.

20.4 Propagation measurements

Experimental measurements were vital to both the development and the testing of the COST210 propagation models. A search was made for existing data, but with limited success as very little suitable material existed even for Europe, and virtually none existed for the remainder of the world.

Significant use was therefore made of the extensive database of measurements compiled during the COST210 project. Altogether measured results from some 64 paths (24 preCOST210, 40 during COST210) from inland, sea and coastal areas were used, covering distances of 60–1000 km and frequencies from 0.7–19 GHz. Figure 20.1 shows the largest of several COST210 networks of experiments to emphasise the extent of the supporting measurement campaigns.

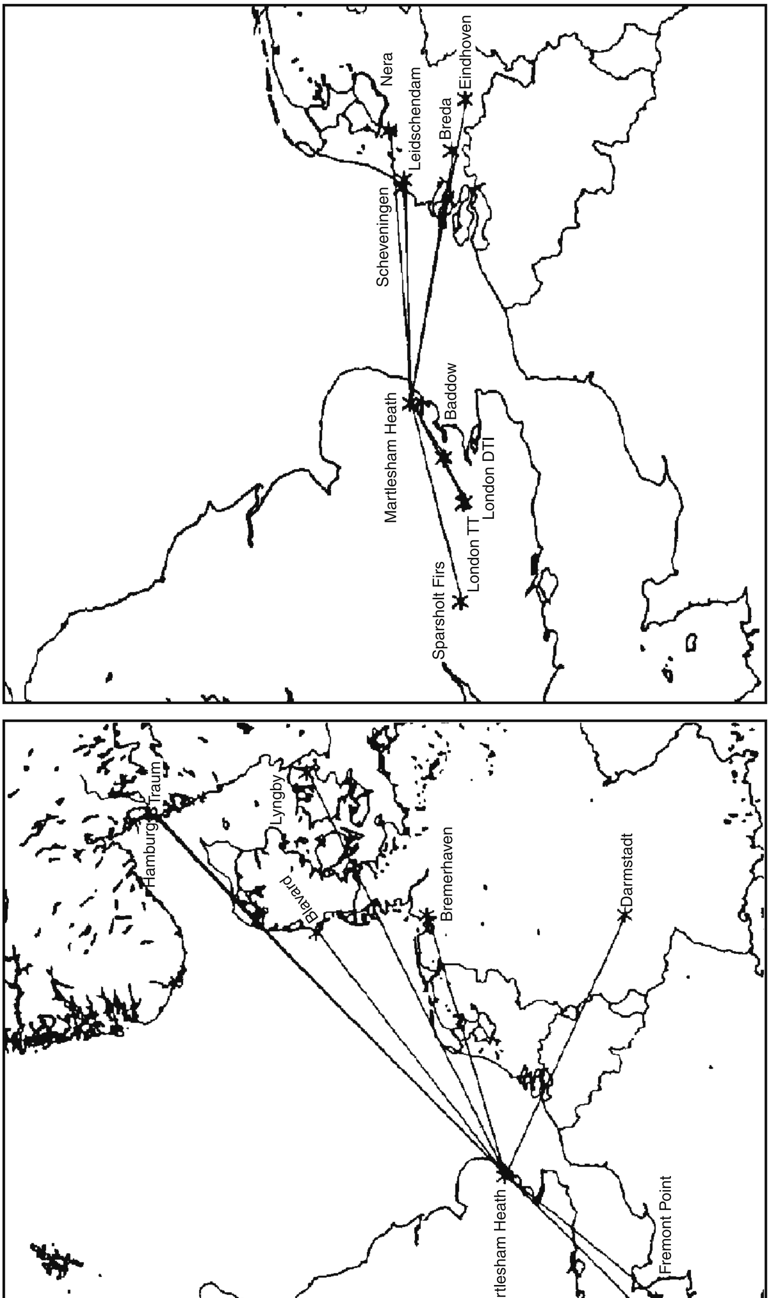


Figure 20.1 1.3 GHz propagation measurements forming one of several networks of experiments set up during the COST210 project

20.5 The propagation modelling problem

The preliminary research work confirmed that seven principal propagation mechanisms would need to be addressed during the development of the new microwave interference prediction procedure. Six of these are clear-air mechanisms, the seventh is hydrometeor scatter. The interference propagation mechanisms can be divided into two classes:

(i) The continuous, or long-term, propagation effects that are present, to a greater or lesser extent, all of the time. These are shown in Figure 20.2. Such propagation mechanisms set the background interference levels with which the radio systems must coexist. Radio planners must ensure that there is adequate separation between systems to keep the background interference level below the (generally very low) limits to be found, for instance, in the ITU-R Recommendations. The long-term propagation loss criterion is generally specified as a basic transmission loss that must be exceeded for all but 10 or 20 per cent of time.

(ii) The anomalous, or short-term, propagation effects that exist for small percentages of time (Figure 20.3). These can support very high levels of interference over significant distances. The propagation loss criterion for short-term interference is usually defined as the loss that will just bring the wanted system to the unavailable state. For digital transmission, this is generally when the error rate rises to 1 in 10^3. This situation is usually only permitted for a small percentage of time, e.g. 0.01 per cent of time or less.

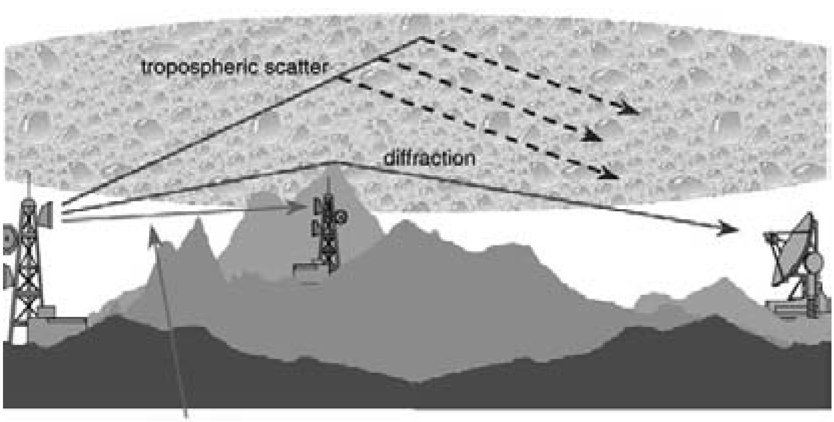


Figure 20.2 Long-term (continuous) interference propagation mechanisms

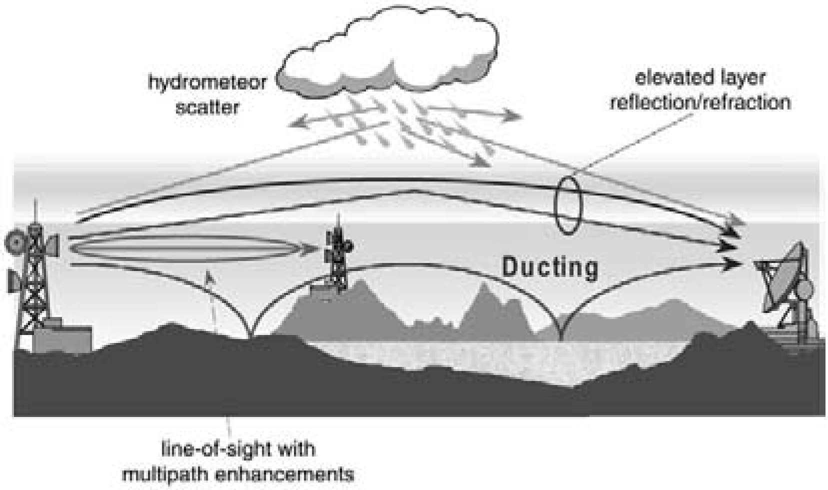


Figure 20.3 Short-term (anomalous) interference propagation mechanisms

20.6 Worst month considerations

When discussing time percentages in the interference context it is important to differentiate between percentages of an average year and those pertaining to the so-called worst month. Telecommunications system error performance objectives are usually defined in terms of error rate (for digital systems) or noise levels (for analogue circuits) that must be achieved for a specified time percentage '. . . of any month'. Engineers planning radio systems to carry high quality telecommunications services often require interference propagation data for the worst month, rather than the average year statistics usually provided by empirical models. Propagation models have not previously offered this level of sophistication, and it was a COST210 objective to overcome this deficiency.

Therefore, to allow the new method to provide a complete prediction package, data from the longer running experiments (over ten years in one case) was used to develop a model [4] which allowed the required worst month time percentage(s) to be translated into equivalent average year percentage(s) which could then be used as inputs to the new prediction models.

20.7 Path profiles

Central to all the propagation calculations is the path profile. The new method assumes that the user has access to a digital terrain database of terrain heights above sea level. Terrain databases are commonly available now, some free of charge. However, for one-off calculations, it is still possible to derive heights from suitable scale maps.

As Figure 20.4 implies, the effective earth's radius must be added to the profile. The value of the effective radius depends on geographical location and time percentage. To allow reasonable modelling of this variability Recommendation ITU-R P.452 includes a simple k factor model as indicated in Figure 20.5.

The COST210 model uses two radio meteorological parameters to determine the local behaviour of the effective earth's radius. These parameters are:

(i) ΔN, the mean radio refractive index lapse rate (in N units km^{-1}) through the first 1 km of the atmosphere at a given location. This parameter represents

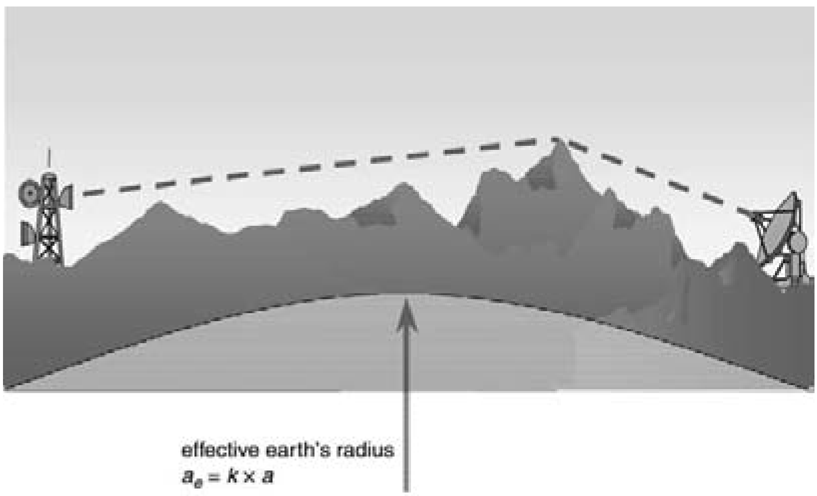


Figure 20.4 Path profile and earth's curvatures

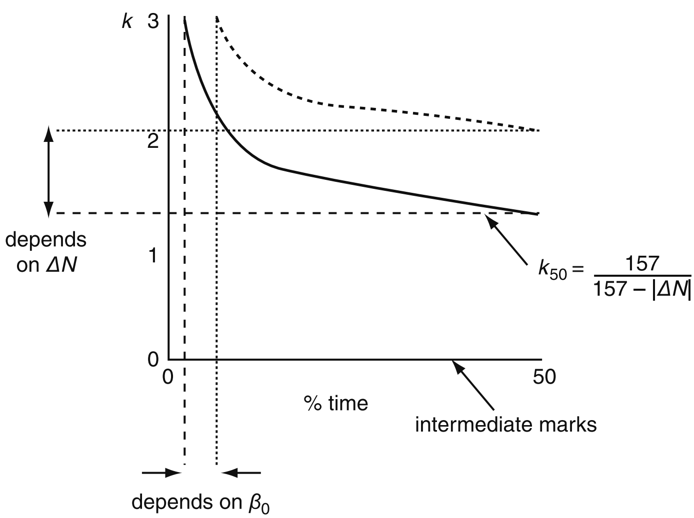


Figure 20.5 The k factor model

the bulk radio refractive index under well mixed atmospheric conditions. Within NW Europe ΔN is approximately $-39\ N\,\mathrm{km}^{-1}$ giving the well known k factor value of 4/3. However, globally, ΔN can vary from about 30 to 80 (average year) or 50 to 100 (worst month).

(ii) β_0, the percentage of time that the modulus of the mean lapse rate through the first 150 m of the atmosphere exceeds $100N$ units km^{-1} at a given location on the earth's surface. This parameter represents the percentage of time that fully developed clear-air anomalous propagation is deemed to occur at a given location. $\beta_0$varies around the world from 1 to 50 per cent (average year) and 2 to 90 per cent (worst month).

The earlier versions of ITU-R Recommendation P.452 provided world maps of the values of these two parameters. These maps were based on the original work of Bean *et al*. [5]. More recently the maps have been replaced by a latitude-dependent β_0 function which returns acceptable results while considerably simplifying the procedure.

The k factor model is only used for percentages of time down to β_0 per cent, as below this value the anomalous propagation mechanisms are deemed to dominate the prediction. For deriving basic path geometry information it is the profile which represents local median atmospheric conditions (k_{50} in Figure 20.5) that is required. However, to determine diffraction effects over a range of time percentages, the profile must be reanalysed for the different effective earth radius conditions applicable to each time percentage of interest.

20.8 Path characterisation

Once the path profile is prepared for the correct radio meteorological conditions, the path can be characterised into one of two types:

(i) line-of-sight (with or without subpath diffraction)
(ii) transhorizon

This characterisation then determines which propagation models should be applied to the path.

For transhorizon paths, Recommendation P.452 model provides a detailed method for analysing the path profile to determine all the necessary terrain-related parameters (horizon distances and elevation angles, terrain roughness etc.).

20.9 The propagation models

The initial COST210 studies showed that new propagation models would have to be developed to cover the principal mechanisms identified *via* the radio meteorological studies and measurements. Because of the complexity of the overall propagation problem (as illustrated by Figures 20.2 and 20.3), the approach taken was to provide individual models for each mechanism, with each model covering an appropriate time percentage range. The overall prediction comprises a composite of these individual predictions.

The models developed within COST210, together with the main areas of innovation, are described in the sections that follow.

20.9.1 Line of sight

This, the simplest of propagation modes, was treated with a four-element model comprising:

- free-space loss;
- gaseous absorption; as is general throughout the procedure, gaseous absorption is calculated using the standard method provided in Recommendation ITU-R P.676–4 [6];
- signal enhancements at the smaller time percentages due to multipath and focusing effects (these can be surprisingly significant – see Figure 20.6) and, where appropriate;
- subpath diffraction effects; this latter element can give up to 6 dB additional loss or up to ~1.5 dB reduced loss, so it was important that it was included; the subpath diffraction effects are derived using the methods of Recommendation ITU-R P.526–6 [7].

20.9.2 Terrain diffraction

Much effort within the COST210 project went into the development of a diffraction model which would cope with any path geometry but be straightforward to

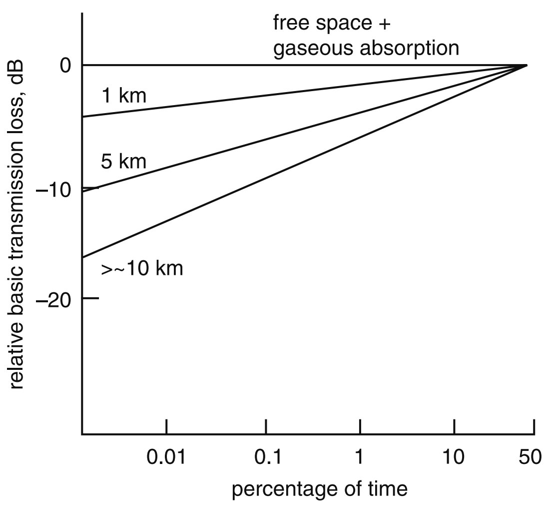


Figure 20.6 Reduced line-of-sight loss due to multipath effects

implement. The result (described in detail in an earlier paper [8]) is a simplification of a more rigorous model [9] which combines Epstein–Peterson [10] and ITU-R [11] formulation with a cascaded cylinders approximation of the terrain to give an extremely versatile model. Variability across the time percentage range was provided *via* the *k* factor model similar to that mentioned above.

However, many difficulties arose over the practical implementation of this diffraction model. The most difficult problem was to optimise the method of determining the radii of the cylinders that represented the main terrain obstacles. Visual inspection of a section of a path profile may readily suggest a correct choice of cylinder radius for a particular obstacle. However, the use of mathematical and software algorithms and a digital terrain database to do the same task is far more troublesome. Due to quantisation effects, terrain databases tend to yield far too many close-spaced obstacles, and hence very different results for different terrain database resolutions. In reality, even the visual approach is subjective and, hence, depends on the degree of detail in the profile.

Resolution-tolerant methods of smoothing the terrain to avoid such problems were developed, and when tested were found to give good results. However, these tended to make the model too complex for the end users and too slow to run on all but the more powerful computers.

From version 8 of Recommendation 452 (452-8) onwards the cascaded cylinders model was abandoned and was replaced by a modification of the Deygout [12] method (Figure 20.7). This model is addressed in Chapter 8 so the detail is not included here. It suffices to say that this approach is less sensitive to the fine detail of the digitally-derived profile, and gives more consistent results.

20.9.3 Tropospheric scatter

As it sets the long-term background interference level on longer paths once the diffraction field has become negligible, troposcatter needed to be included in the

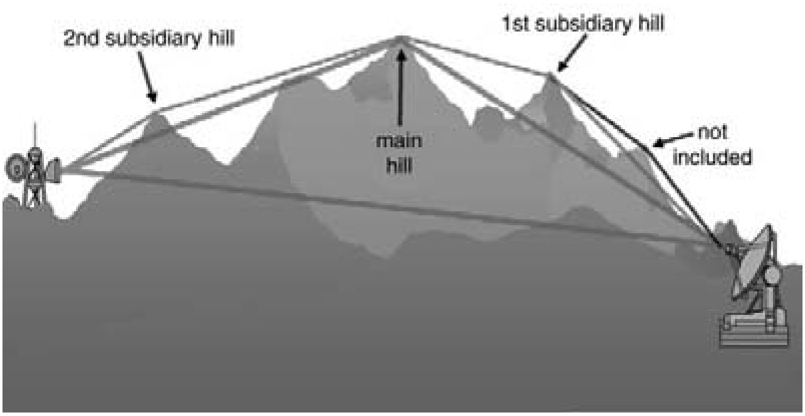


Figure 20.7 The new modified Deygout diffraction model

overall propagation prediction procedure. Propagation models used for the design of troposcatter communications systems are inappropriate as they concentrate on the low signal, rather than enhanced signal, region of the distribution. Hence, something more appropriate had to be developed.

Troposcatter propagation results from scattering from turbulent refractive index inhomogeneities within a common volume seen by the beams of two antennas (Figure 20.8). Key parameters are the size and height of the common volume, the scatter angle (or angular distance – see also Figure 20.14 in Section 20.9.4.3) and the frequency.

For operational troposcatter communications systems the common volume is formed by the intersection of the main beams of carefully aligned antennas, sited to maximise the common volume and minimise the scatter angle. In typical interference scenarios the common volume will arise *via* partial intersections of the sidelobes of the two antennas, or the main beam of one and the sidelobes of the second. Little theoretical work exists on tropospheric scatter coupling under such complex conditions. Those models which do attempt to deal with this issue (e.g.[13]) become extremely complex.

Given all the uncertainties, a very simple model was adopted for interference prediction applications. The chosen model was a derivative of the earlier Yeh [14] model and testing showed this to be effective for this application. The changes to the original Yeh model included an improved frequency dependence (with a transition from a 30log(*f*) to a 20log(*f*) above 2 GHz), and the addition of an empirically-derived cumulative distribution to provide results for time percentages below 50 per cent.

The troposcatter model makes use of a third radio meteorological parameter N_0, which is the sea-level atmospheric refractivity at a given location. Again, a map is provided in the ITU-R text and the path centre value is used for the calculation.

Note that a low value of water vapour concentration (3 g m^{-3}) is used to

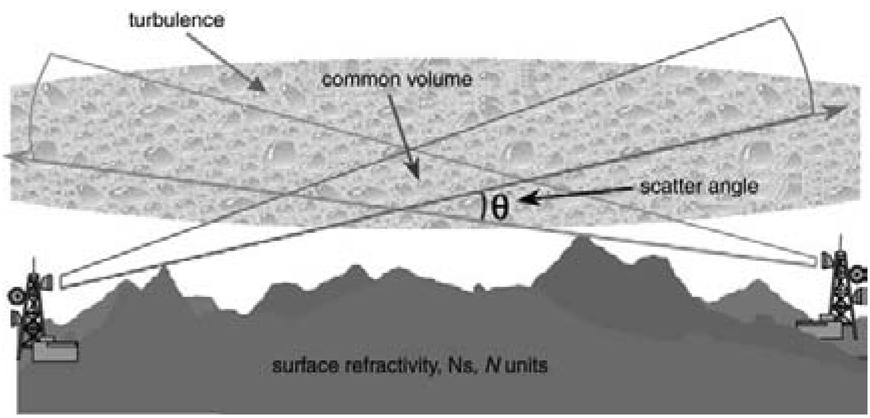


Figure 20.8 Tropospheric scatter propagation

calculate the gaseous absorption in the troposcatter case. This is because much of the troposcatter path exists well above the earth's surface.

The main elements of the troposcatter model are shown in Figure 20.9.

20.9.4 Ducting and elevated layer reflection/refraction

These two anomalous propagation modes, which in most situations dominate the interference prediction, provided perhaps the biggest challenge. Because the two modes exhibit similar statistical characteristics for distances below about 250 km, a single solution was found which covered both mechanisms to good effect. As will be seen, a correction is applied for the difference in distance dependence that occurs beyond this distance.

The short-term anomalous propagation model for layer reflection/refraction and ducting is based on the determination of values for two parameters:

- a time percentage, β per cent, derived from β_0 per cent
- a basic transmission loss, L_{br} dB that will occur at this time percentage.

These two values position a reference point on a cumulative distribution of basic transmission loss against time percentage (see Figure 20.10). Once the point L_{br},β is established, the remainder of the cumulative distribution is developed for time percentages above and below this reference point. This approach provides the flexibility to cater for a wide range of path input conditions.

The operation of this model can best be understood by reviewing the factors influencing the translation of the reference point against each axis of the prediction graph.

20.9.4.1 Factors affecting the reference time percentage, β per cent

The reference time percentage for the point, β per cent, is based on the value of the parameter β_0 described in Section 20.7. However, β_0 is the point value for the general location of the path, and it must be corrected for four factors (Figure 20.11) which determine the actual percentage incidence of anomalous propagation over the particular path. These factors are

frequency | distance | scatter angle | surface refractivity (location)

$$L_{bs} = 190 + k(f) + 20 \log d + 0.573\theta - 0.15\, N_0$$

$$+ L_c + A_g - 10.1\,(-\log(p/50))^{0.7} \text{ dB}$$

basic transmission loss due to troposcatter | antenna gain correction | gaseous absorption (3 g m^{-3}) | empirical cumulative distribution

$k(f) = 30 \log f$ below 2 GHz; $3 + 20 \log f$ at higher frequencies

Figure 20.9 Principal elements of the tropospheric scatter model

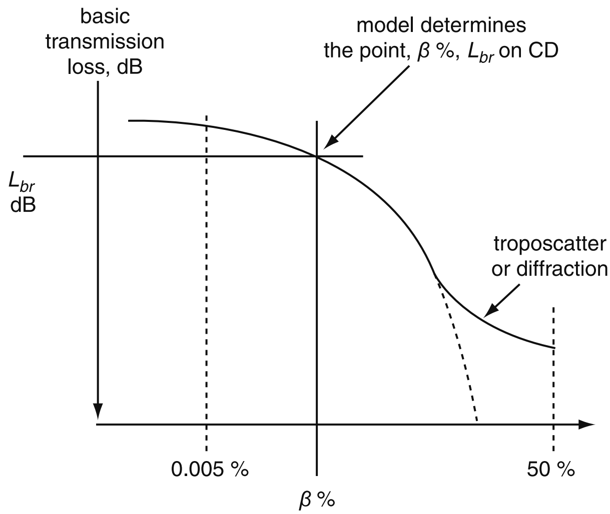


Figure 20.10 Basic approach used for the anomalous propagation

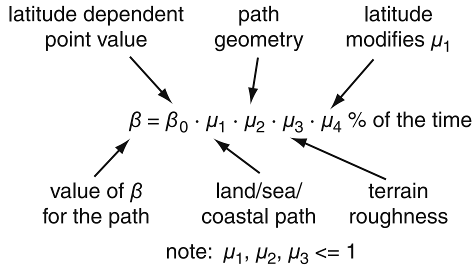


Figure 20.11 Factors determining the final value of β

- the mix of inland, coastal land and sea zones in the path; this determines the value of the parameter μ_1;
- the path geometry as determined from the analysis of the path profile; this controls the value of μ_2;
- the terrain roughness along the path; also determined from the path profile analysis, this provides the value for μ_3;
- a latitude-dependent correction to the relative occurrence of anomalous propagation on land, sea and coastal paths, the parameter, μ_4, adjusts the value of μ_1.

20.9.4.2 Land, sea and coastal effects – determining the value of μ_1

The approach chosen to account for the land/coastal/sea proportions of the path resulted from the findings of the COST210 radio meteorological studies.

The studies confirmed that although the layer propagation conditions affected all three zones (inland areas, flat coastal regions and sea areas) to a similar extent, ducting conditions mainly affected sea paths. However, it was shown that ducting was in evidence in the low-lying coastal areas, but to a lesser extent than over the sea. It was also found that, if the relative incidence of over-sea anomalous propagation (dominated by the occurrence of ducting) is taken as unity, then that of inland paths (resulting purely from layer reflection/refraction) was about 0.15–0.2 per cent. The flat coastal areas were subjected to some ducting in addition to the layer effects, and the relative incidence of anomalous propagation was about 0.35 per cent.

A most important finding was that a relatively short land section of the path (about 40 km) would tend to isolate an antenna from the over-sea ducts. Such a path would then only exhibit the characteristics of an inland path (i.e. the characteristics of layer reflection/refraction propagation) irrespective of the length of the remaining sea section(s) (Figure 20.12 illustrates these effects). The value of β is therefore reduced from its initial value of β_0 per cent as the amount of inland and/or coastal land along the path increases.

It should be noted that the values for the relative incidence of anomalous propagation over inland, sea and coastal areas relate to the latitudes of NW Europe. For other latitudes it is necessary to correct the ratio by means of the factor μ_4. This is taken care of in the model by including μ_4 and the latitude of the path centre in the calculation of μ_1.

To assist with the evaluation of μ_1, a coastal zone database was developed based on the inland, coastal land and sea definitions of Recommendation 452 and digital terrain height data from a global 30 arc second terrain database. This data was initially created as a raster database (see Figure 20.13). However, the data has recently been vectorised by the ITU Radiocommunications Bureau, and the vector coastal zone boundaries are now included in the ITU-R International Digital World Map (IDWM) which is available for purchase from the ITU.

20.9.4.3 Path geometry – setting the value of μ_2

The value of β is further modified by the path geometry. The longer the path length, and the lower the antennas, the greater the path angular distance will become. Figure 20.14 explains the modified angular distance concept used in this context.

As the angular distance increases the stratified atmosphere has an increasingly difficult task to do in order to achieve a given low basic transmission loss. The greater the demands on the refractive index structure, the lower will be the probability that the necessary conditions will actually occur. Thus the value of β reduces with increasing path angular distance.

It is important at this point to note that a significant difference exists between the distance dependence of the ducting and layer reflection/refraction mechanisms. Ducts (especially over the sea) extend for considerable distances (1000 km or more). To a first approximation, the loss within the duct can be considered as a linear specific attenuation resulting from leakage from the duct and scattering by the earth's surface (particularly over land areas). Once the energy is trapped

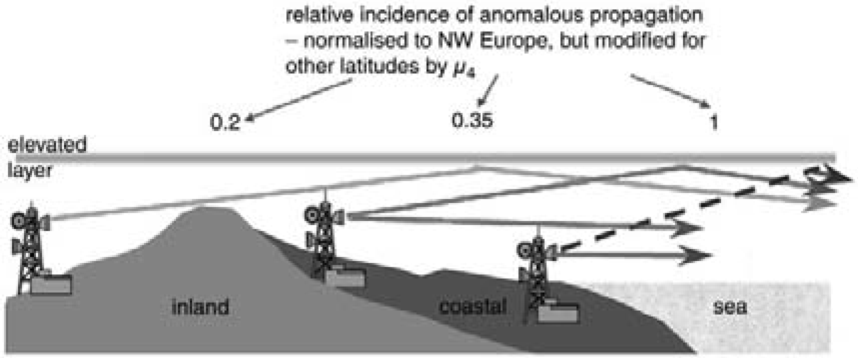


Figure 20.12 *Relative incidence of anomalous propagation over inland, sea and coastal areas (values shown are for NW Europe)*

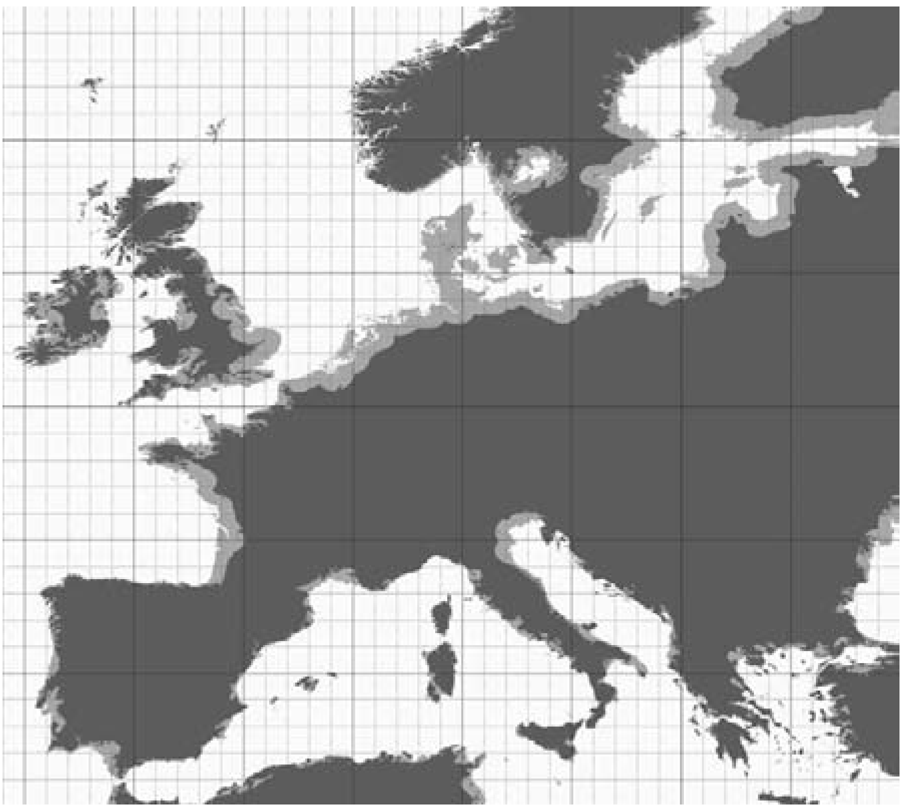


Figure 20.13 *Coastal zone data for use with Recommendation 452 (note that the ITU-R provides a vector form of this data with ITU Digital World Map)*

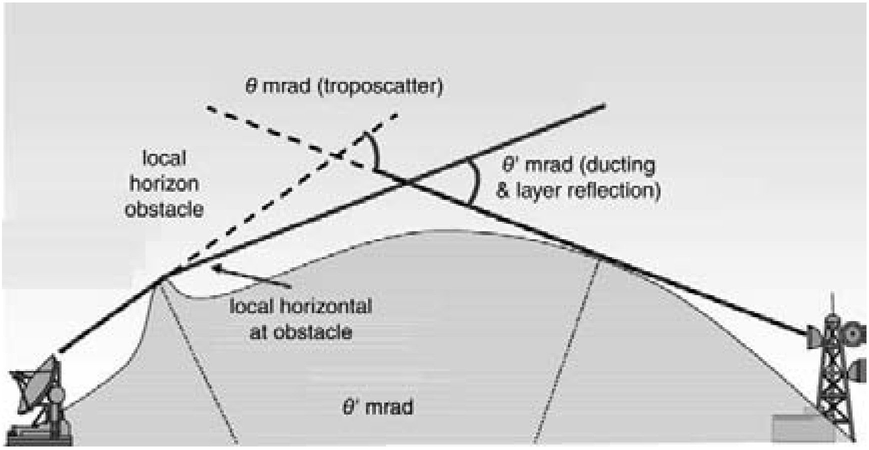


Figure 20.14 The modified angular distance θ′ mrad, used for the clear air anomalous propagation calculation (note that θ′ can have a different value than θ used for the troposcatter calculations when one or other horizon subtends a positive angle at the relevant antenna)

within the duct structure it will follow the curvature of the earth. Additional distance makes little difference, apart from the effect of the specific attenuation, until the physical limit of the duct is reached.

It was noted above that over short to medium distances (up to about 250 km) the evidence from measured data shows duct and elevated layer propagation to exhibit very similar losses, although their relative occurrence will be different.

However, beyond about 250 km, different factors come into play. In the layer reflection/refraction case, the energy is not constrained to follow the earth's curvature. As the distance increases around the curvature of the earth, the angle of incidence of a signal arriving at the elevated super-refractive layer increases. At a given distance (depending primarily on overall path geometry) the angle will become such that the energy will tend to pass through the layer, unless a more intense (and hence less probable) layer is present. Furthermore, the height at which the more intense layer will need to exist will increase as the path length increases.

During the COST210 radio meteorological studies, an analysis of long-term radiosonde data confirmed that the probability of finding a given layer strength reduces with height, and thus the probability of a given basic transmission loss existing will be still further reduced. The combination of these two effects leads to a relatively rapid reduction in the probability of low-loss paths existing over inland areas as the distance increases. At a given small time percentage this translates into a greater increase in loss with increasing distance in the inland case when compared with the more linear distance dependence of the ducting scenario (Figure 20.15). The net result is that on inland

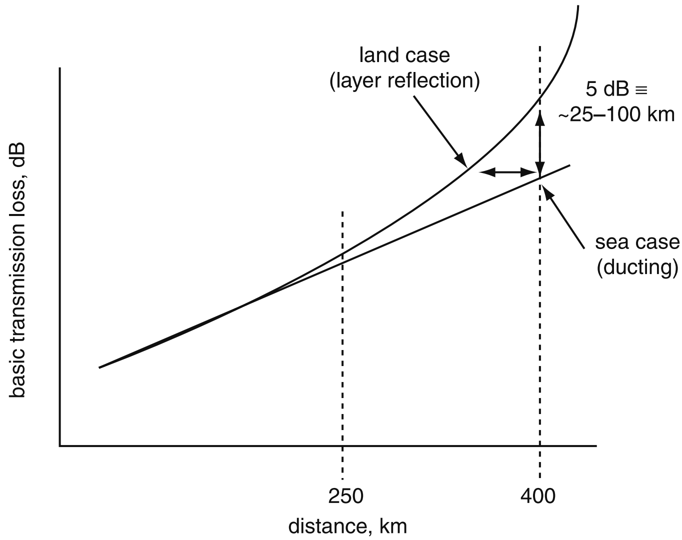


Figure 20.15 Difference in distance dependence between sea/coastal and inland cases for distances beyond ~250 km

paths significant levels of interference become increasingly rare at distances beyond about 400 km.

20.9.4.4 Terrain roughnes – setting the value of μ_3

The final factor influencing β is terrain roughness. The way the model deals with this is shown in Figure 20.16.

Figure 20.16*a* shows that, with relatively flat land, several options exist whereby strong interfering signals can become established. Layer reflection/refraction, surface ducts and low elevated ducts all have an opportunity to create low-loss paths. The probability of such paths occurring is therefore high.

Figure 20.16*b* represents the case of moderate terrain roughness. Here, the surface duct mode is effectively suppressed as the surface will block the signals and/or inhibit ground reflections. However, the elevated layers and elevated ducts can still become established. With only some modes possible, the low-loss paths can still occur, but the probability of this happening is significantly reduced.

Figure 20.16*c* shows the situation with high terrain roughness, where only the elevated layer reflection/refraction can be supported. The probability of a low-loss path is now very considerably reduced.

Any further increase in terrain heights would suppress even this layer reflection/refraction mode, leaving troposcatter as the dominant mechanism. Increasing terrain roughness therefore progressively reduces the probability of a low-loss path.

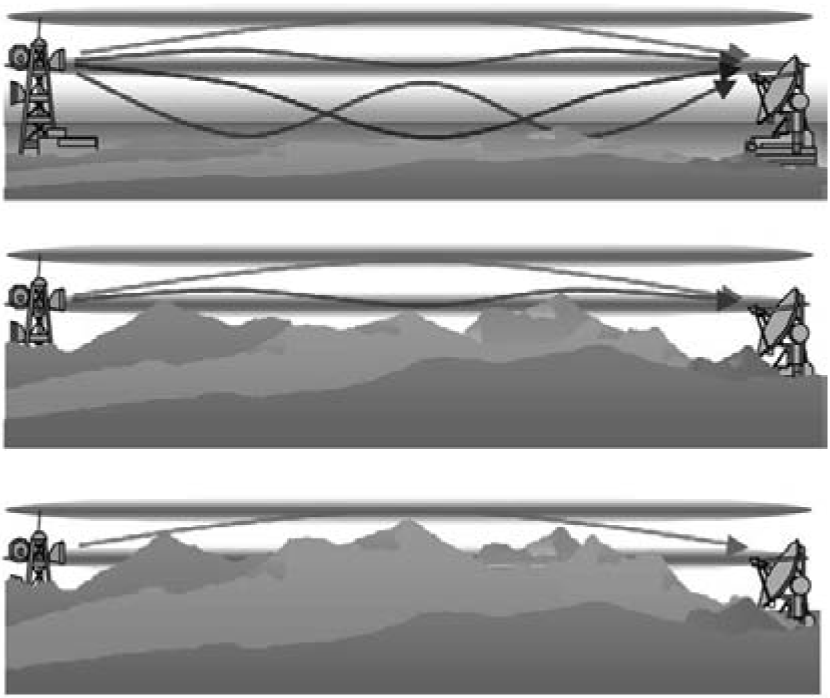

(a)
(b)
(c)

Figure 20.16 The impact of increasing terrain roughness
(a) flat terrain – high probability of interference
(b) rough terrain – less probability of interference
(c) rugged terrain – low probability of interference

20.9.5 Factors affecting the basic transmission loss reference value, L_{br} dB

Just as with parameter β, the basic transmission loss reference value, L_{br} dB, has a number of elements which contribute to its eventual value. Figure 20.17 shows the main elements of the calculation of L_{br}.

This apparent simple equation encapsulates a wide range of concepts needed to model the clear-air anomalous propagation path. The equation is based on the following assumptions:

- that there is normal propagation between the antennas and their respective horizons
- coupling into and out of the anomalous propagation structure (layer or duct) occurs at the two radio horizons
- that, on average, there is 6 dB coupling loss into the ducts or layers at each horizon
- that the losses within the duct/layer structure can be modelled by a simple specific attenuation

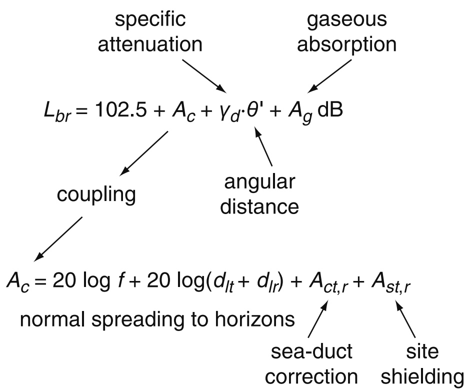


Figure 20.17 Principal elements of the calculation of L_{br}

- that positive horizon angles (created by terrain obstacles) can cause diffraction losses even under anomalous propagation conditions
- that antennas at low heights close to the sea (i.e. those already within the over-sea duct structure) will not exhibit the 6 dB coupling loss.

These assumptions are elaborated in the sections that now follow.

20.9.5.1 Free-space loss between the antennas and their respective horizons

This is accounted for by an equation of the form $92.5 + 20\log f + 20\log(d_{lt} + d_{lr})$ dB, where d_{lt} and d_{lr} are the two horizon distances. Originally the 2 × 6 dB coupling losses were added to the 92.5 dB constant to give 104.5 dB. However, the final tuning of the Recommendation P.452 model to give a best fit against measured data resulted in this constant being adjusted to the 102.5 dB now to be found in Recommendation P.452.

20.9.5.2 Specific attenuation within the stratified layer or duct

Based on earlier work within (what was then) CCIR Study Group 5, the specific attenuation, γ_d dB/km, within the anomalous propagation structure was defined as $0.05f^{1/3}$ dB/km. The distance here is the angular distance term θ' converted into its kilometre equivalent. In the calculation of this angular distance, positive horizon angles are set to (near) zero to correctly reflect the coupling assumptions and to avoid double accounting with the site shielding losses.

20.9.5.3 Coupling at the radio horizons

Figure 20.18 explains the thinking behind the horizon coupling argument. Coupling will occur most effectively when the angle between the radio ray and the stratified atmosphere is at a minimum.

For negative horizon angles (top picture of Figure 20.18) the angle is a min-

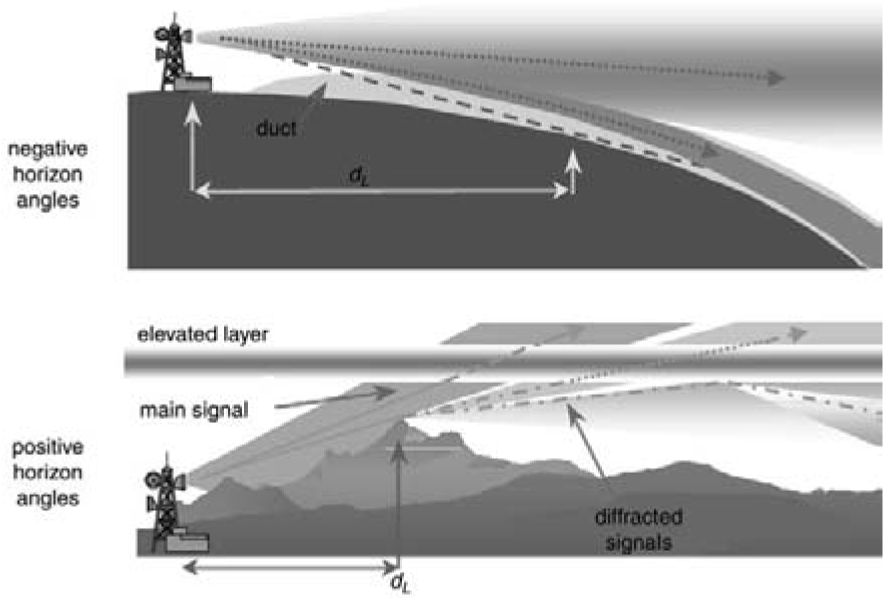


Figure 20.18 Coupling into the stratified atmosphere at or near the antenna horizons

imum where the ray is tangential to the earth's surface (note that a negative angle is below the local horizontal at the antenna). Some of the energy from the antenna will not be coupled in, hence the assumed 6 dB coupling loss. Note that it is always the antenna gain in the direction of the horizon (which is not usually the boresight gain) that is used for the prediction calculation.

When the terrain creates a positive horizon angle (lower picture of Figure 20.18), the assumption is that the high-angle energy will pass through the layer or duct. It will then only be the diffracted rays leaving the obstacle at angles close to the local horizontal that are coupled into the stratified atmosphere at just beyond the horizon point. Again, there will be a coupling loss. Obviously, there will also be the diffraction loss associated with propagation over the obstacle, but this is accounted for separately (see site shielding in Section 20.9.5.5).

In reality, the coupling will not happen at a point but will occur over a range of distances around the horizon point. In the case of a negative horizon angle, this range will extend either side of the actual radio horizon (especially as the duct has a finite thickness). In the positive horizon angle case, the distance range over which coupling actually occurs will generally be beyond the actual obstacle, depending on the height of the layer with respect to that of the obstacle. This is perhaps the least well developed part of the model, and improvements could be added during a future revision.

20.9.5.4 Close coupling into over-sea ducts

Where an antenna is at a low height and is close to the sea it will probably already be within the sea-duct structure. In these circumstances the assumed 6 dB coupling loss may no longer be correct, as more energy could be trapped within the duct.

To protect against this eventuality, Recommendation P.452 provides a sea-duct coupling correction of up to 6 dB reduction in the coupling loss (Figure 20.19). The value of the correction depends on the antenna height above sea level, and on the distance of the antenna from the sea. It has little effect beyond about 5 km from the coast.

20.9.5.5 Terrain site shielding

The losses over the terrain obstacles that give rise to positive horizon angles are known as site shielding losses. These losses are calculated using a simple diffraction model which uses the horizon angle and horizon distance as input parameters (Figure 20.20).

It should to be noted here that Recommendation P.452 employs a correction to the simple diffraction model to give a safety margin for distant horizon cases. This was necessary because distant horizons could allow higher than expected signal levels to arrive at a slightly elevated angle *via* layer reflections or *via* duct structures that follow smooth rolling terrain (see Figure 20.21). The correction applies the site shielding loss only to positive horizon angles greater than $0.1d_L$ mrad, rather than to all angles greater than zero. This then provides an increasing margin of protection as the horizon distance increases.

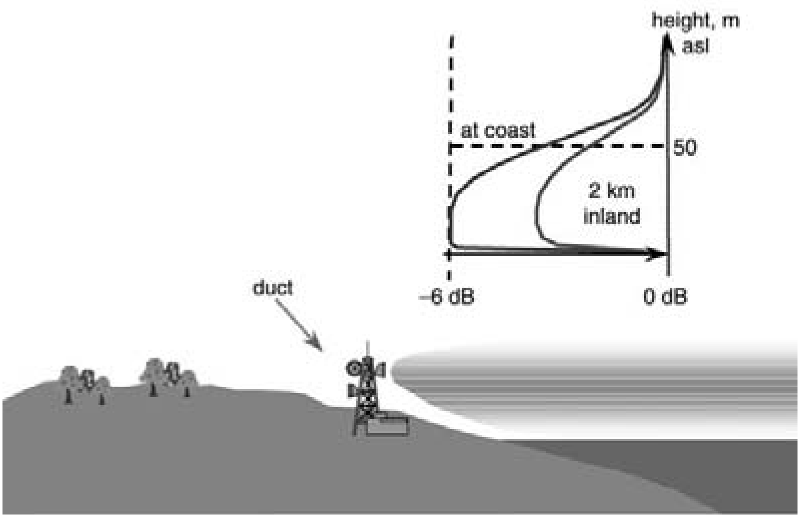


Figure 20.19 Close coupling into over-sea ducts

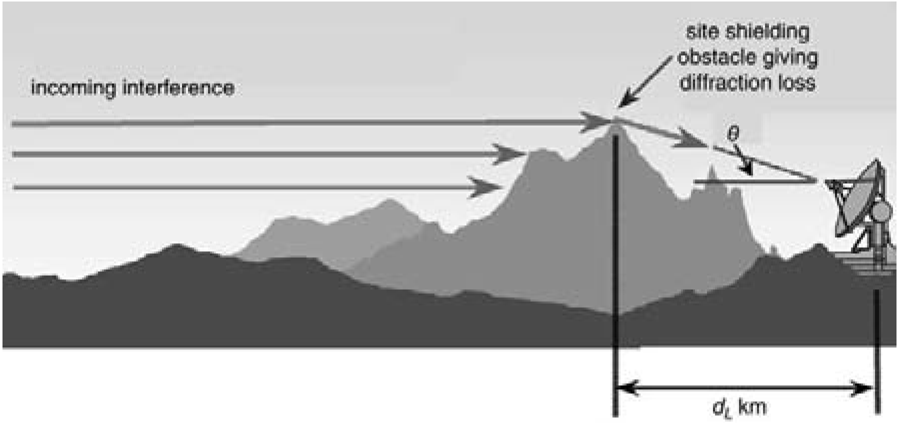


Figure 20.20 Site shielding diffraction losses

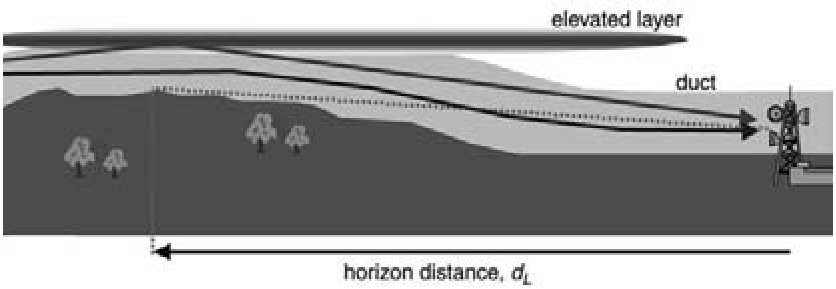


Figure 20.21 Interfering signals bypassing low horizon obstacles situated at significant distances from the antenna

20.9.6 Local clutter losses

With new telecommunications applications being found for radio all the time, the Recommendation P.452 model needs to be continually developed to match new interference scenarios. In the past, microwave links have been synonymous with tall towers and large antennas, and with signals being kept deliberately well clear of any natural or man-made clutter. However, the trend now is to use radio within the telecommunications access network, i.e. to form the final link between the network and the customer. This broad range of applications has gained the generic name of fixed wireless access (FWA).

FWA systems tend to be of the point-to-multipoint variety, and often have to have antennas at heights which are comparable to, or below, the surrounding clutter heights. This makes service provision difficult, but can represent a bonus in terms of protection from interference. The question then arises as to how much additional protection can be assumed?

This question is actually difficult to answer because, once down among the clutter, the dominant interference path rarely remains that which lies along the great circle between interferer and victim. In reality clutter is a three-dimension propagation problem (Figure 20.22) with signals coming over and through buildings and being reflected and scattered off other buildings and trees etc. The direction from which the strongest interference will come is normally totally unknown. This is a particularly important consideration as it means that the antenna discrimination in the direction of the great-circle path cannot necessarily be relied upon.

Within the ITU-R a cautious approach is being taken to the introduction of clutter losses into this Recommendation. A clutter loss model was proposed (Figure 20.23) which provided a conservative approach while allowing at least some advantage to be taken from clutter protection in order to improve frequency reuse. The clutter loss model offered the following features:

- a loss related to the antenna height as a fraction of the local clutter height, but which reduced with increasing distance from the clutter;
- a region from 80 to 100 per cent of nominal clutter height where little additional loss was assumed due to uncertainties over actual clutter heights;
- a frequency-dependent maximum additional loss (20–40 dB for 0.7–40 GHz); this was significantly less than the normal diffraction loss that would exist were it to be assumed that the interference arrived by a single path over the top of the clutter, and allowed for the problems represented in Figure 20.22 to be accommodated.

The ITU-R found even this to be insufficiently conservative, and in the event ITU-R Study Group 3 adopted (as an interim measure) the 0.7 GHz curve from Figure 20.23 to apply at all frequencies. However, as this allows for up to

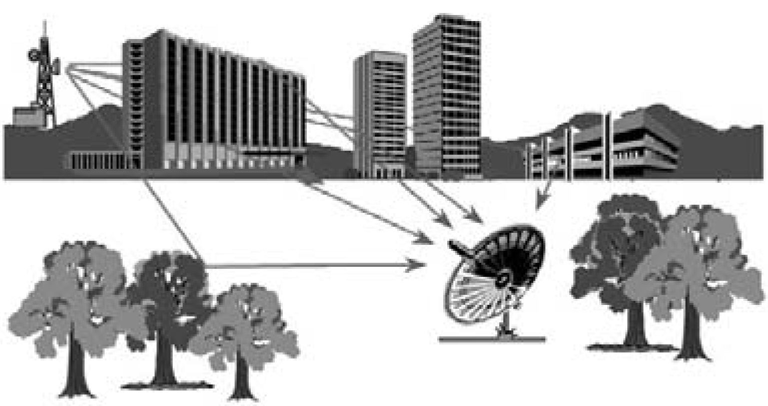


Figure 20.22 Clutter effects: a three-dimensional problem

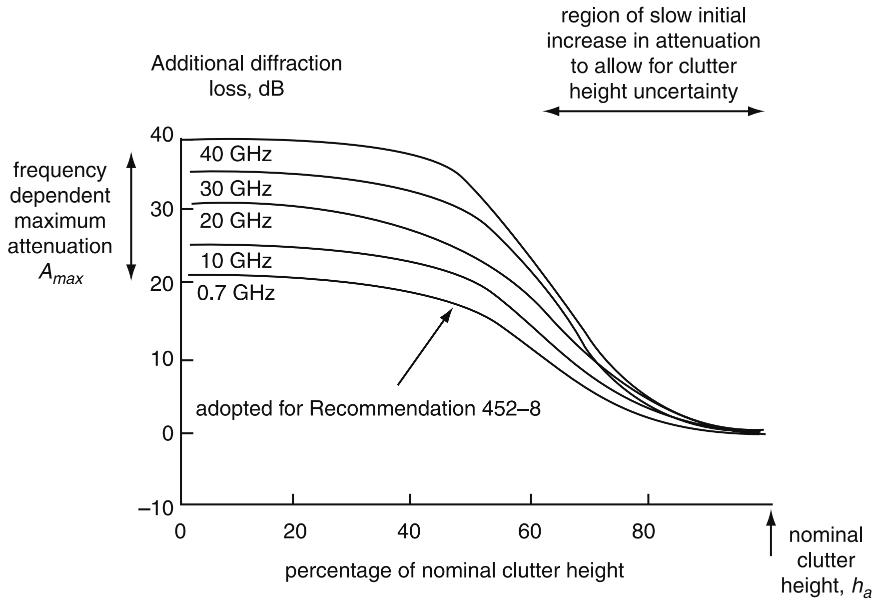


Figure 20.23 The clutter loss model

20 dB additional loss to be included in frequency-sharing calculations, it still represented an important step forward.

The application of this model remains optional within Recommendation P.452. When it is required it can be included by a simple amendment to the basic Recommendation P.452 calculation, but the decision must be made at the start of the prediction calculation.

The basic Recommendation P.452 procedure is used to calculate the basic transmission loss to the top of the assumed clutter height at an assumed clutter distance from the far end. In cases with high clutter heights this will significantly change the path geometry compared with the no clutter calculation. Once the basic transmission loss to the top of the clutter is found, the additional loss due to the actual antenna height can be added. The method of application is shown in Figure 20.24.

20.10 Hydrometeor scatter

20.10.1 The COST210 and Recommendation 452 models

Hydrometeor scatter is a very different problem to the clear-air issues discussed so far. It arises in situations where the main beams of two antennas intersect (wholly or partially), and the common volume thus formed fills with rain.

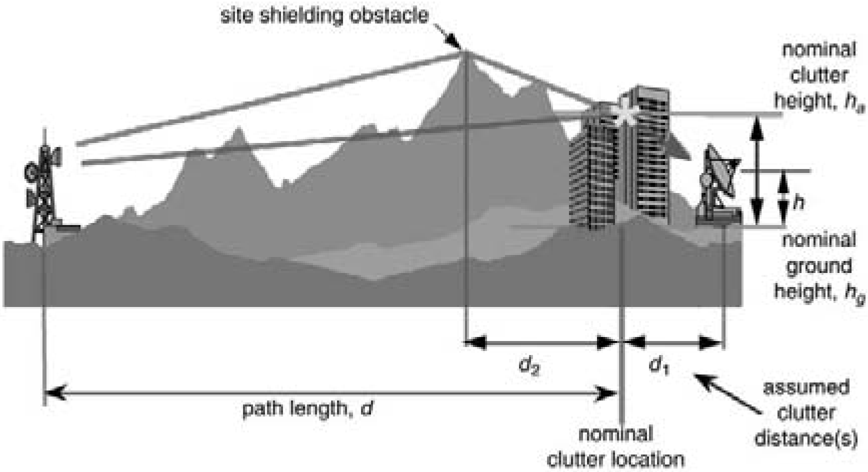


Figure 20.24 Applying the clutter loss correction

When one of the antennas is associated with a sensitive earth station receiver, it is also possible to get significant levels of interference *via* the sidelobes of such an antenna. This can be accommodated by the COST210 hydrometeor scatter model.

The theoretical aspects of rainscatter are covered in another Chapter and thus there is no need to include the detail here. However, the hydrometeor scatter propagation model included within Recommendation P.452 is a powerful and flexible tool that was also developed within the COST210 project. The most comprehensive version of the model is well documented in the COST210 report.

Within the ITU-R context, the full COST210 model exists as software (written in FORTRAN) which may be obtained from the ITU. The full model can accommodate a wide range of input conditions and path geometry. The paper version as documented in Recommendation P.452 itself is more limited in its scope because it would have been difficult to capture all aspects of the COST210 model in an ITU-R recommendation. Recommendation P.452 contains sufficient material to provide an awareness of the versatility of the model. However, those who have a need to make hydrometeor scatter calculations should make themselves aware of both the Recommendation P.452 and the original COST210 versions.

As with the clear-air models, the development of this model was supported by many new long-term hydrometeor scatter measurements.

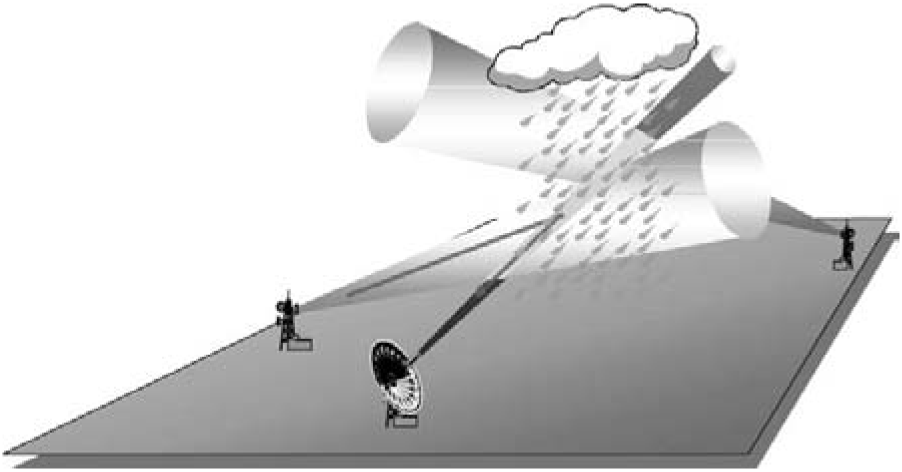


Figure 20.25 The hydrometeor scatter problem

20.10.2 Basic scenario of the Recommendation P.452 version

The hydrometeor scatter scenario modelled by Recommendation P.452 is shown in Figure 20.26. The two intersecting beams form a common volume that can become rain filled. In the Recommendation P.452 model, one beam is deemed to be narrow compared with the other. Such a scenario arises in the case of an earth station (the narrow beam) working in the same band as a terrestrial microwave link (the broader beam) – a situation commonly encountered in practice.

The wider of the two beams is considered to contribute to interference down to the −18 dB points of its radiation pattern. The common volume is formed in the region between h_{min} and h_{max} (see Figure 20.26), i.e. that part of the narrow beam/wide beam intersection that falls within a cylindrical rain cell of diameter d_s km.

The rain has a maximum height depending on its location in the world (see Recommendation ITU-R P.839–2 [15]), and a statistical rain rate distribution according to the location, as given in Recommendation ITU-R P.837–2 [16].

Depending on the geometry, it is possible for part of the common volume to be above the maximum rain height, and the model allows for this in the calculations.

The common volume is assumed to be rain filled below the maximum rain height. The level of interference is determined by an integration along the narrow beam from h_{min} to h_{max}, calculating the following principal factors (see Figure 20.27):

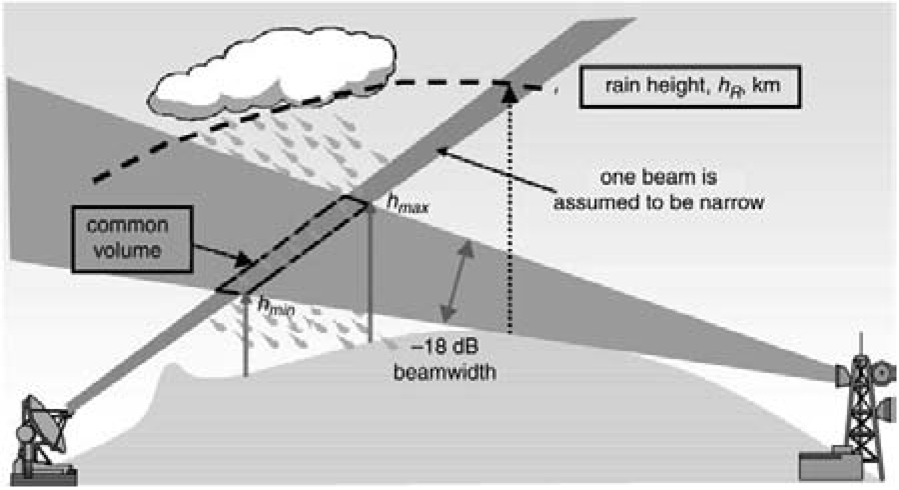


Figure 20.26 Hydrometeor scatter scenario in Recommendation 452

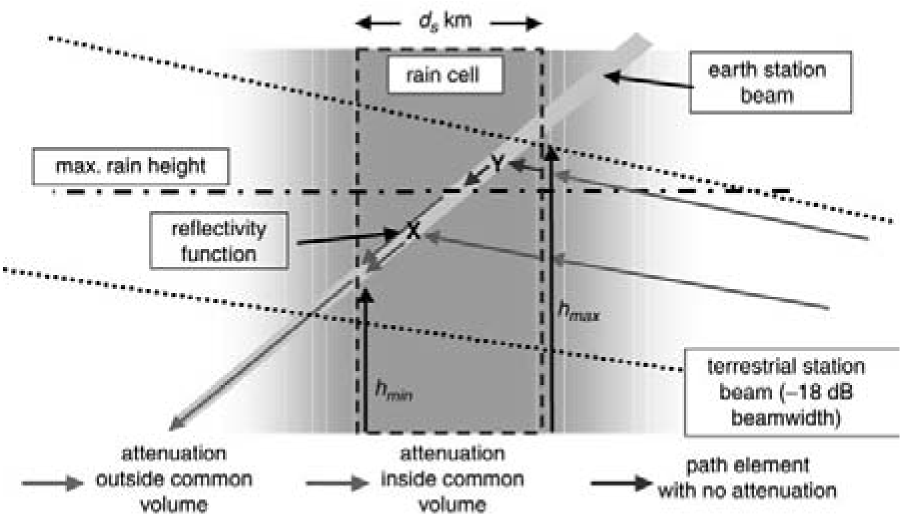


Figure 20.27 The principal elements of the hydrometeor scatter integration

(i) The attenuation between the two antennas and the common volume, assuming an exponential rain rate decay with increasing distance from the common volume. If any path elements are above the rain height, no attenuation is calculated for such elements.
(ii) Attenuation within the rain-filled common volume, assuming a uniform

rain density in the common volume itself. Again, no attenuation is assumed for path elements above the rain height.

(iii) The scattering caused by the rain in the common volume. This scattering continues above the rain height to account for the effects of ice, but the reflectivity reduces at 6.5 dB km^{-1} above the top of the rain.

By its nature, the calculation using this integral provides the transmission loss between the antennas, rather than the basic transmission loss as determined by the clear-air calculations. If the predicted levels are to be compared, then the simplest way is to adjust the clear-air calculation to also give transmission loss. This is not a major problem as the two antenna gains must be known in order to undertake the calculation.

20.11 Deriving the overall prediction

Once the appropriate propagation calculations have been completed an overall prediction for the path is obtained by taking the lowest loss among the individual propagation model results for each time percentage of interest. If the whole time percentage range is calculated, then the overall result represents a lowest-loss envelope across all the individual cumulative distribution elements. Figure 20.28 illustrates the method used.

As shown in the Figure, different propagation mechanisms can dominate the

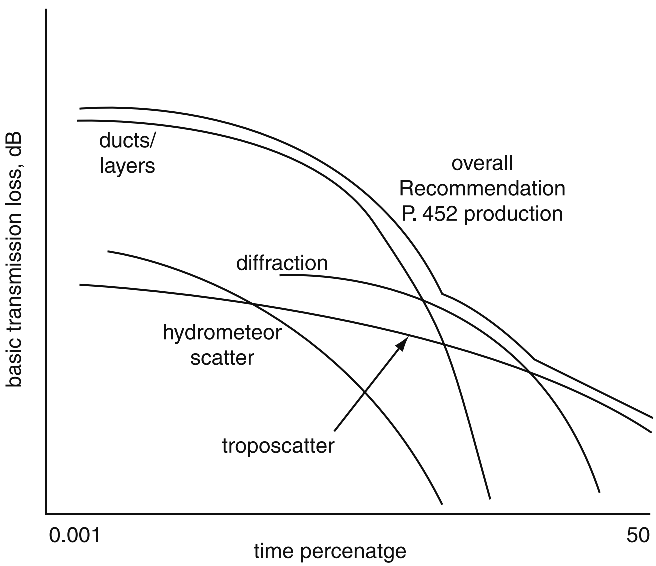


Figure 20.28 Deriving the overall prediction

prediction over different parts of the time percentage range, depending on the specific input conditions for the calculations. This can, and does, happen on practical paths, once again illustrating the complexity of the problem that the recommendation sets out to solve.

Recommendation ITU-R P.452 has now been used for many millions of path calculations associated with the planning of new systems and the protection of what exists already. Although there is always a need to refine and improve such tools as the requirements grow more complex, the version which currently exists (Recommendation P.452–9) provides a reliable and flexible general purpose tool for interference analysis.

20.12 References

1 COST210 Management Committee: 'Influence of the atmosphere on interference between radio communication systems at frequencies above 1 GHz'. Final report of COST Project 210, Commission of the European Communities, Luxembourg, April 1991

2 Recommendation ITU-R P.452-9: 'Prediction procedure for the evaluation of microwave interference between stations on the surface of the earth at frequencies above about 0.7 GHz', 1999

3 BYE, G.D.: 'Radio-meteorological aspects of anomalous propagation in NW Europe'', *BT Technical Journal*; 1988, **6**, **3**

4 HEWITT, M.T. *et al.*: 'Worst month prediction model for transhorizon microwave interference', *Electron. Lett.*, 1989, **25**, (24) pp.167

5 BEAN, B. *et al.*: 'A world atlas of atmospheric radio refractivity', ESSA monograph 1, US Dept. of Commerce, Institute for Telecommunications Sciences, Boulder, Colorado, 1966

6 Recommendation ITU-R P.676-4, 'Attenuation by atmospheric gases', 1999

7 Recommendation ITU-R P.526-6 'Propagation by diffraction', 1999

8 HEWITT, M.T. and MEHLER, M.J.: 'A versatile diffraction modelling techniques for microwave interference propagation prediction', International Wroclaw Symposium on *Electromagnetic compatibility*, Wroclaw, Poland, 1992

9 SHARPLES, P.A. and MEHLER, M.J.: 'A cascaded cylinder model for predicting terrain diffraction loss at microwave frequencies', *Proc. IEE* 1989, **136**, pp.331–337

10 EPSTEIN, J. and PETERSON, D.W.: 'An experimental study of wave propagation at 850 Mc', *Proc IRE*, 1953, **41**, pp.595–611

11 Recommendation ITU-R P.526-3: 'Propagation by diffraction', 1992

12 DEYGOUT, J.: 'Multiple knife-edge diffraction of microwaves', *IEEE Trans. Antennas Prop.*, **14**, pp.480–489

13 RICE, P.L. *et. al.*: 'Transmission loss predictions for tropospheric communication circuits'. NBS technical note 101, revised January 1967, US National Bureau of Standards, vol. II

14 YEH, L.P.: 'Simple methods of designating troposcatter circuits', *IRE Trans. Comms. Sys.*, **CS-8** (III), p.119

15 Recommendation ITU-R P.839-2: 'Rain height model for prediction methods', 1999
16 Recommendation ITU-R P.837-2, 'Characteristics of precipitation for propagation modelling', 1999

Chapter 21

Earth–space propagation: Recommendation ITU-R P.618

R.G. Howell

21.1 Introduction

Recommendation ITU-R P.618 [1] presents propagation data and step-by-step methods for predicting the performance of earth–space radio systems operating in frequency bands between about 1 GHz and 55 GHz and applicable to average annual percentage times in the range 0.001 per cent to 50 per cent. This recommendation deals only with effects relating to the wanted signal. Interference aspects are covered in Recommendations ITU-R P.452 [2], P.619 [3] and P.1412 [4].

There is insufficient space within this Chapter to present all the details of the prediction methods given in Recommendation ITU-R P.618 and, therefore, a broad description will be given together with references to related ITU documents.

21.2 Overview

21.2.1 General

The propagation effects relevant to the design of earth–space radio systems are somewhat different from those applicable to terrestrial line-of-sight radio links. In particular, multipath propagation produced by atmospheric layers can be neglected except for paths at low elevation angles (<4°). However, the effects of the ionosphere (at frequencies below about 4 GHz), high altitude ice particles and short-term variations in atmospheric radio refractive index, producing rapid amplitude and phase variations in the signal (mainly at elevation angles below 5°), must be taken into account. Also, during fine weather conditions, the radio noise emission from objects within the main beam of the earth-station antenna is usually significantly less than that for terrestrial line-of-sight antennas. This allows low-noise receiving systems to be used, however, the effects of radio noise from the cosmos, the atmosphere and the environment of the antenna must then be taken into account.

The effects of rain are similar to those on terrestrial line-of-sight links with the exception that, in the case of earth–space links, the effective path length through the rain is determined by the maximum height of the rain and its vertical and horizontal distribution, rather than the length of the link and the horizontal distribution of the rain along the path (the situation with terrestrial line-of-sight links).

21.2.2 Propagation effects

The main propagation effects that must be considered in the design of earth–space radio systems are:

- free-space loss
- ionospheric effects: Faraday rotation of the plane of polarisation, dispersion, excess time delay and ionospheric scintillation
- absorption by atmospheric gases
- absorption, scattering and depolarisation by hydrometeors (water droplets and ice particles of various forms in precipitation, clouds etc.)
- radio noise emissions from the cosmos, sun, moon and absorbing media within the atmosphere
- loss of signal due to divergence of the earth-station antenna beam resulting from the normal radio refractive index structure of the atmosphere
- a decrease in the effective antenna gain, due to phase decorrelation across the antenna aperture, caused by irregularities in the radio refractive index structure of the atmosphere
- slow fading due to beam bending caused by large-scale changes in the radio refractive index of the atmosphere
- more rapid fading (scintillation) and variations in angle of arrival, due to small-scale variations in the radio refractive index of the atmosphere
- possible limitations in bandwidth due to multiple scattering or multipath effects, especially in high-capacity digital systems
- attenuation by the local environment of the earth-station antenna (buildings, trees etc.)
- short-term variations in the ratio of the attenuations at the uplink and downlink frequencies, which may affect the accuracy of adaptive fade countermeasures e.g. transmitter power control
- for nongeostationary satellite (nonGSO) systems, the effect of the varying satellite elevation angle.

At elevation angles greater than 10° only gaseous attenuation, rain and cloud attenuation and possibly (to a lesser extent) scintillation will contribute significantly to the total atmospheric loss.

21.3 Ionospheric effects

The effect of the ionosphere on signals transmitted over earth–space paths may be important, particularly at frequencies below 4 GHz, but rapidly become less important as the frequency approaches and exceeds 10 GHz, see also Chapter 17. The effects include:

- Faraday rotation: a linearly polarised wave propagating through the ionosphere undergoes a progressive rotation of the plane of polarisation;
- dispersion, which results in a differential time delay across the bandwidth of the transmitted signal;
- excess time delay relative to the delay a signal would have if it traversed the same path with no ionosphere;
- ionospheric scintillation: inhomogeneities of electron density in the ionosphere cause refractive focusing or defocusing of radiowaves and lead to amplitude fluctuations termed scintillation; ionospheric scintillation is maximum near the geomagnetic equator and smallest in the mid-latitude regions, the auroral zones are also regions of large ionospheric scintillation; strong scintillation is Rayleigh distributed in amplitude, weaker scintillation is nearly log-normally distributed; these fluctuations decrease with increasing frequency and depend upon path geometry, location, season, solar activity and local time;
- phase fluctuations having a spectral density proportional to $1/f^3$, where f is the Fourier frequency of the fluctuation, accompany the amplitude fluctuations.

Table 21.1 shows the frequency dependence and order of magnitude of some of the ionospheric effects for frequencies between 1 GHz and 10 GHz [1]. In terms of telecommunications systems the effect of the ionosphere can be ignored at frequencies above about 10 GHz.

Further details on the effects of the ionosphere can be found in Recommendation ITU-R P.531 [5].

21.4 Clear-air effects

21.4.1 General

Other than atmospheric absorption, clear-air effects in the absence of precipitation are unlikely to produce serious fading in space telecommunication systems operating at frequencies below about 10 GHz and at elevation angles above 10°. However, at low elevation angles (≤10°) and at frequencies above about 10 GHz tropospheric scintillation can, on occasion, cause serious degradations in performance. At very low elevation angles (≤4° on inland paths, and ≤5° on over-water or coastal paths), fading due to multipath propagation effects can be particularly severe.

Table 21.1 The frequency dependence and magnitude of some ionospheric effects

Effect	*Frequency dependence*	*1 GHz*	*3 GHz*	*10 GHz*
Faraday rotation	$1/f^2$	108°	12°	1.1°
Propagation delay	$1/f^2$	0.25 μs	0.028 μs	0.0025 μs
Refraction	$1/f^2$	<0.6″	<4.2″	<0.36″
Variation in the direction of arrival (RMS)	$1/f^2$	12″	1.32″	0.12″
Absorption (auroral and/or polar cap)	$\sim 1/f^2$	0.05 dB	6×10^{-3} dB	5×10^{-4} dB
Absorption (mid-latitude)	$1/f^2$	<0.01 dB	<0.001 dB	$<10^{-4}$ dB
Dispersion (ps Hz^{-1})	$1/f^3$	0.0004	1.5×10^{-5}	4×10^{-7}
Scintillation (peak-to-peak)	see Rec. ITU-R P.531	>20 dB	~10 dB	~4 dB

21.4.2 Free-space loss and attenuation by atmospheric gases

The free-space loss and the losses produced by dry air and water vapour have been described in Chapters 2 and 6, respectively. The essential difference between the way the attenuations produced by atmospheric gases are treated when planning terrestrial line-of-sight links and when planning earth–space links is that in the former it can be assumed that the density of each gas is constant along the path, whereas in the latter the variation of density with height must be taken into account.

In Recommendation ITU-R P.618 an exponential decrease in density with height is assumed. This is then used to calculate equivalent heights for dry air and water vapour. These are the heights that each constituent of the atmosphere would have if its density was constant (equal to the ground level value) and the total zenith attenuation produced by the gas was equal to the actual zenith attenuation. For example, the water vapour equivalent height, away from the water vapour absorption band, is 1.6 km in clear weather and 2.1 km in rain. This height is dependent on the temperature of the atmosphere and the ground level water vapour density. The values given are for a water vapour density of 7.5 g m^{-3} and a temperature of 15 °C. The equivalent height for dry air at frequencies below 50 GHz is 6 km.

The total attenuation produced by atmospheric gases is calculated using the specific attenuation (dB km^{-1}) for each gas, the equivalent heights and the path elevation in formulae which take account of the path length up to the equivalent height. For path elevation angles above 10° a formula which neglects atmospheric refraction and Earth curvature can be used, giving an attenuation proportional to the cosecant of the path elevation. For path elevation angles equal to or less than 10° a more complex formula, which takes account of atmospheric refraction and earth curvature, is used. Further details on the losses produced by dry air and water vapour can be found in Recommendation ITU-R P.676 [6].

21.4.3 Phase decorrelation across the antenna aperture

Incoherence of the wave front of a wave incident on a receiving antenna is caused by small-scale irregularities in the refractive index structure of the atmosphere. This results in an antenna-to-medium coupling loss that can be described as a decrease of the antenna gain. The loss increases both with increasing frequency and decreasing elevation angle, and is a function of antenna diameter. Although not explicitly accounted for in the atmospheric refraction models discussed below, this effect is negligible in comparison.

21.4.4 Beam spreading loss

The amount of downward ray bending produced by the regular decrease of the radio refractive index of the atmosphere with height under average refractivity conditions is a function of the elevation angle of the ray, and increases as the elevation angle of the ray is reduced (particularly for elevations below about 3°). Hence, rays at the top and bottom of the antenna main beam travel along paths having slightly different elevation angles and the resulting differential ray bending (top to bottom) produces an additional divergence of the beam in the vertical plane. Therefore, the cross-sectional area of the beam is increased and the gain of the antenna is reduced. It should be noted that there is no increase in the divergence of the beam in the horizontal plane. The loss in the antenna gain resulting from this beam spreading is independent of frequency, over the range of 1–100 GHz (see Recommendation ITU-R P.834 [7] for further details).

The loss A_{bs} due to beam spreading in regular refractive conditions can be ignored at elevation angles above about 3° at latitudes less than 53° and above about 6° at higher latitudes.

At all latitudes, the beam spreading loss in the average year at elevation angles less than 5° can be estimated from:

$$A_{bs} = 2.27 - 1.16 \log(1 + \theta_0) \text{ dB} \quad \text{for } A_{bs} > 0 \tag{21.1}$$

where θ_0 is the apparent elevation angle (mrad), taking into account the effects of refraction.

Recommendation ITU-R P.618 also gives formulae (which are dependent on the latitude of the earth station) for calculating the loss produced by beam spreading for the average worst month.

21.4.5 Scintillation and multipath fading

The magnitude of tropospheric scintillation depends on the magnitude and structure of the refractive index variations, increasing with frequency and with the path length through the medium, and decreasing as the antenna beamwidth decreases because of aperture averaging. Monthly-averaged RMS fluctuations are well correlated with the wet term of the radio refractivity, N_{wet}, which depends on the water vapour content of the atmosphere. N_{wet} may be estimated for periods of a month or longer from meteorological data obtained at the surface.

At very small percentages of time, or conversely large fade depths (greater than about 10 dB), the fading at very low elevation angles (≤4° on inland paths, and ≤5° on over-water or coastal paths) is observed to be more severe than that predicted due to scintillation. The fading is also observed to have a character similar to multipath fading on terrestrial links. Like the distribution on terrestrial links, the distribution for very-low-angle satellite links also appears correlated with refractivity gradient statistics. The overall fading distribution shows a gradual transition from a scintillation distribution at large exceedance percentages to a multipath fading distribution (with a slope of 10 dB decade^{-1}) at small percentages.

Recommendation ITU-R P.618 gives a general method for predicting the cumulative distribution of tropospheric scintillation at elevation angles of greater than 4°. The method is based on monthly or longer averages of surface ambient temperature (t) and relative humidity (H) for the site. As the averages of t and H vary with season, distributions of scintillation fade depth exhibit seasonal variations, which may also be predicted by using seasonal averages of t and H in the method. The procedure has been tested at frequencies between 7 and 14 GHz, but is recommended for applications up to at least 20 GHz.

Recommendation ITU-R P.618 also gives methods for calculating both the deep fading (> 25 dB) and the shallow fading part of the scintillation/multipath fading distribution for paths having elevation angles of less than 5°. The methods estimate the average worst month and average year fading distributions, taking account of the combined effects of beam spreading, scintillation and multipath fading.

The net fade distribution due to tropospheric refractive effects is the combination of the beam spreading, scintillation and multipath fading effects described above. Tropospheric and ionospheric scintillation distributions may be combined by summing the respective time percentages for which specified fade levels are exceeded.

21.4.6 Angle of arrival

Because of the downward bending of the wave produced by the normal decrease in radio refractive index with increasing height an earth-station antenna must be pointed at an elevation angle slightly greater than that which would apply in the absence of refraction. For example, the total angular refraction (the increase in apparent elevation) is about 0.65°, 0.35° and 0.25°, for path elevation angles of 1°, 3° and 5°, respectively, for a tropical maritime atmosphere. For a polar continental climate the corresponding values are 0.44°, 0.25° and 0.17°. Other climates will have values between these two extremes. The day-to-day variation in apparent elevation is of the order of 0.1° (RMS) at 1° elevation, but the variation decreases rapidly with increasing elevation angle.

Short-term variations, due to changes in the refractivity-height variation, may be of the order of 0.02° (RMS) at 1° elevation, again decreasing rapidly with increasing elevation angle. In practice, it is difficult to distinguish between the effect of the short-term changes in the refractivity-height distribution and the effect of random irregularities superimposed on that distribution. A statistical analysis of the short-term angle-of-arrival fluctuation at 19.5 GHz and at an elevation angle of 48° suggests that both in elevation and azimuth directions, standard deviations of angle-of-arrival fluctuations are about 0.002° at the cumulative time percentage of one per cent. The seasonal variation of angle-of-arrival fluctuations suggests that the fluctuations increase in summer and decrease in winter. The diurnal variation suggests that they increase in the daytime and decrease both in the early morning and the evening.

Further information on elevation-angle errors due to refraction may be found in Recommendation ITU-R P.834 [7].

21.4.7 Propagation delays

Recommendation ITU-R P.834 [7] provides data on radio meteorologically-based methods for estimating the mean value and the variability in the propagation delay (or range error) produced by the troposphere for earth–space paths. At frequencies above 10 GHz, the delay produced by the ionosphere is generally smaller than that for the troposphere, but may have to be considered in special cases. Further information on the delay produced by the ionosphere can be found in Recommendation ITU-R P.531 [5].

Range determination to centimetre accuracy requires careful consideration of the various contributions to excess range delay. A dry atmosphere adds a delay equivalent to an additional 2.3 m to a zenith path. However, the largest source of uncertainty in the path length is produced by the water vapour component of the atmosphere, due to the variability of the water vapour concentration. For example, this produces a delay equivalent to an additional path length of 10 cm for a zenith path and a reference atmosphere with a surface water vapour concentration of 7.5 g m^{-3} and 2 km scale height (see Recommendation ITU-R P.676 [6]).

For current satellite telecommunication applications, additional propagation delays contributed by precipitation are sufficiently small to be ignored.

21.5 Hydrometeor effects

21.5.1 General

The physical basis for the absorption, scattering and depolarisation by hydrometeors has been discussed in Chapter 12. It was shown in this Chapter that the specific attenuation (dB km^{-1}) for rain increases very rapidly for frequencies above 10 GHz. The majority of the loss of the wanted signal when passing through rain is produced by absorption, although a small amount, particularly at frequencies above 30 GHz, is due to scattering. Although not significant in terms of its contribution to the loss of the wanted signal, the scattered component can be very important in terms of its potential to cause interference to other cochannel systems (see Chapter 20 and Recommendation ITU-R P.452 [2]).

21.5.2 Prediction of attenuation statistics for an average year

If statistics of rain attenuation, derived from long-term measurements carried out in the vicinity of the site and on a path at the same (or a similar) angle of elevation are available, it is recommended that the required attenuation statistics are obtained by scaling these data to the required frequency (see Section 21.5.2.1). If suitable measured data are not available, Recommendation ITU-R P.618 provides a method for predicting the statistics of rain attenuation for an average year from the statistics of point rainfall rate (see Section 21.5.2.2).

21.5.2.1 Frequency scaling of rain attenuation statistics

The following formulae can be used to scale rain attenuation statistics collected at one frequency (f_1 GHz) to those at another frequency (f_2 GHz). If A_1 and A_2 are the equiprobable values of the rain attenuation at frequencies f_1 and f_2, respectively, then:

$$A_2 = A_1(\varphi_2/\varphi_1)^{1-H(\varphi_1,\varphi_2,A_1)} \tag{21.2}$$

where

$$\varphi(f) = f^2/(1 + 10^{-4}f^2) \tag{21.3}$$

and

$$H(\varphi_1, \varphi_2, A_1) = 1.12 \times 10^{-3}(\varphi_2/\varphi_1)^{0.5}(\varphi_1\ A_1)^{0.55} \tag{21.4}$$

21.5.2.2 Calculation of rain attenuation statistics from rainfall rate statistics

Recommendation ITU-R P.618 provides a method for predicting the statistics of rain attenuation from the statistics of rainfall rate. Figure 21.1 shows the path geometry used in the propagation model.

The rain is assumed to extend from the ground to a rain ceiling. The length of the slant-path below the rain ceiling (L_s) is calculated:

(a) for a path elevation angle less than 5° taking account of atmospheric refraction and earth curvature
(b) for a path elevation angle greater than or equal to 5° using a simpler flat earth nonrefracting atmosphere model.

In each of the above cases the length of the projection along the ground of the

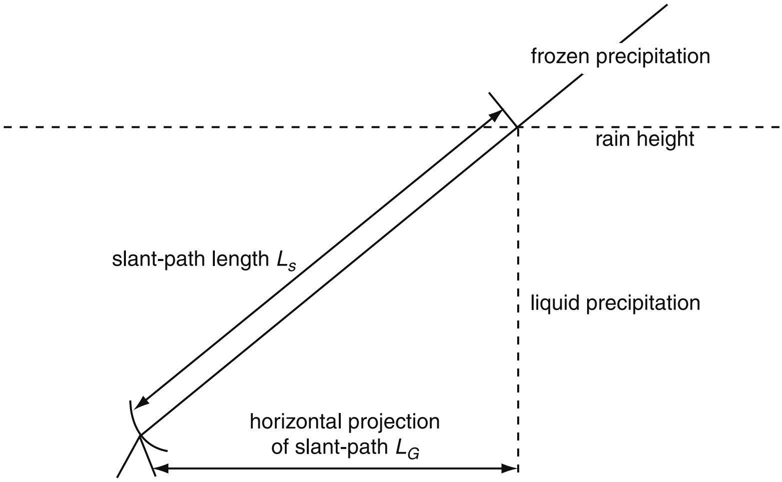


Figure 21.1 Path geometry used in the Recommendation ITU-R P.618 rain attenuation prediction model

slant path (L_G) is calculated. This length is then used to determine horizontal and vertical reduction factors, which take account of the spatial nonuniformity of the rain both horizontally and vertically. The rain-specific attenuation, corresponding to a rainfall rate equal to that exceeded for 0.01 per cent of an average year ($R_{0.01}$), is then used with the effective path length to calculate the attenuation exceeded for 0.01 per cent of an average year. The attenuation exceeded for other percentage times in the range 0.001 per cent to 5 per cent is then derived from the 0.01 per cent attenuation value.

Clearly, a key meteorological parameter used in this model is the rainfall rate exceeded for 0.01 per cent of an average year ($R_{0.01}$). Rainfall rate statistics are strongly dependent on the integration time used in determining the rainfall rates. An integration time of one minute should be used to determine the value of $R_{0.01}$ used in this model. If a value for $R_{0.01}$ is not available from long-term locally measured data then an estimate can be obtained from maps of rainfall rate given in Recommendation ITU-R P.837 [8].

21.5.3 Seasonal variations – worst month

It should be noted that the propagation model described in Section 21.5.2.2 gives the rain attenuation statistics for an average year. System planning often requires statistics for an average worst month. These can be derived from the average annual statistics using methods given in Recommendation ITU-R P.841 [9].

First, the average annual time percentage, p per cent, corresponding to the desired average worst month time percentage p_w per cent is determined using a relationship between the two that takes into account the climatic region. Then the attenuation exceeded for p_w per cent of the average worst month is equal to the attenuation exceeded for p per cent of an average year, as determined using either the propagation model described in Section 21.5.2.2, or by scaling from long-term statistics as described in Section 21.5.2.1.

Curves, giving the variation of worst month values from their mean, are provided in Recommendation ITU-R P.678 [10].

21.5.4 Year-to-year variability of rain attenuation statistics

Distributions of rain attenuation measured on the same path at the same frequency and on the polarisation may show marked year-to-year variations. In the range 0.001 per cent to 0.1 per cent of the year, the attenuation values at a fixed probability level are observed to vary by more than 20 per cent RMS.

21.5.5 Fade duration and rate of change

The durations of rain fades exceeding a specified attenuation level are approximately log-normally distributed. Median durations are of the order of several minutes. No significant dependence of these distributions on fade depth is evident in most measurements for fades of less than 20 dB, implying that the larger total time percentage of fades observed at lower fade levels or at higher frequencies is composed of a larger number of individual fades having more or less the same distribution of durations. Significant departures from a log-normal distribution seem to occur for fade durations of less than about half a minute. Fade durations at a specified fade level tend to increase with decreasing elevation angle.

Data are needed on the contribution of attenuation events shorter than ten seconds to the total fading time for the planning of integrated services digital network (ISDN) connections *via* satellite. This information is especially relevant for the attenuation level corresponding to the outage threshold, where events lasting longer than ten seconds contribute to system unavailable time, and shorter events affect system performance during available time.

Existing data indicates that, in the majority of cases, the exceedance time during available time is 2 per cent to 10 per cent of the net exceedance time. However, at low elevation angles, where the short period signal fluctuations due to tropospheric scintillation become statistically significant, there are some cases for which the exceedance time during available time is far larger than in the case at higher elevation earth–space paths.

The rate of change of attenuation (fading rate) can be an important factor when fade mitigation techniques, such as adaptive transmit power control (Section 21.11.2), are considered. There is broad agreement that the distributions of positive and negative fade rates are log-normally distributed and very similar to each other, however, the dependence of fade rate on fade depth has not been established.

21.5.6 Short-term variations in the frequency scaling ratio of attenuations

Data on the instantaneous ratio of the rain attenuation at different frequencies is relevant to the application of a variety of adaptive fade techniques. The frequency-scaling ratio has been found to be log-normally distributed, and is influenced by rain type and rain temperature. Data reveal that the short-term variations in the attenuation ratio can be significant, and are expected to increase with decreasing path elevation angle.

21.6 Estimation of total attenuation produced by multiple sources

For systems operating at frequencies above about 18 GHz, and especially those operating at low elevation angles and/or with small fade margins, the effect of multiple sources of simultaneously occurring atmospheric attenuation must be considered.

The total attenuation ($A_T(p)$ dB) is the result of the combined effect of rain, gases, clouds and scintillation.

Defining:

$A_R(p)$: attenuation (dB) due to rain exceeded for a probability p per cent, estimated using Recommendation ITU-R P.618 [1]

$A_C(p)$: attenuation (dB) due to clouds exceeded for a probability p per cent, estimated using Recommendation ITU-R P.840 [11]

A_G: mean gaseous attenuation (dB) due to water vapour and dry air, estimated using Recommendation ITU-R P.676 [6]

$A_S(p)$: attenuation (dB) due to tropospheric scintillation exceeded for a probability p per cent, estimated using Recommendation ITU-R P.618 [1]

where p is the probability of the attenuation being exceeded in the range 50 per cent to 0.001 per cent. A general method for calculating total attenuation ($A_T(p)$) for a given probability is given by:

$$A_T(p) = A_G + \sqrt{(A_R(p) + A_C(p))^2 + A_S^2(p)} \qquad (21.5)$$

However, for $p < 1$ per cent, $A_C(p) = 0$ because cloud attenuation is already included in the rain attenuation prediction method used in Recommendation ITU-R P.618.

Some simplifications can be made to Eqn. 21.5 under certain conditions as listed in Table 21.2.

21.7 System noise

The atmospheric contribution to the noise temperature of an earth-station antenna (and hence to the system noise) increases as the loss produced by absorption in atmospheric gases and rain along the path increases. For earth

Table 21.2 Simplification of Eqn. 21.5 under certain conditions

Condition	*Simplifying assumption*
Rain dominated	in most parts of the world, except dry climates, rain will dominate the total attenuation when 1% > p > 0.001%; in this case the other effects may be neglected and the total attenuation given in Eqn. 21.5 reduces to: $$A_T(p) = A_R(p)$$
Multiple effects	in most parts of the world, except dry climates, all effects may contribute a measurable level of propagation impairment when 5% > p > 1%; however, for elevation angles above about 10° the effect of scintillation may become negligible; in this case, Eqn. 21.5 reduces to: $$A_T(p) = A_G + A_R(p) + A_C(p)$$
Nonraining	in most parts of the world, except very wet climates, the non-rainy condition is present when p > 5%; during clear air conditions, rain attenuation is, by definition, 0 dB. In this case, Eqn. 21.5 reduces to: $$A_T(p) = A_G + \sqrt{A_C^2(p) + A_S^2(p)}$$ when operating at elevation angles greater than about 10°, scintillation may become negligible during clear air conditions and Eqn. 21.5 will further reduce to: $$A_T(p) = A_G + A_C(p)$$

stations with low-noise front ends, this increase in noise temperature may have a greater impact on the resulting signal-to-noise ratio than the attenuation itself.

The atmospheric contribution to the noise temperature of an earth station antenna may be estimated using the equation:

$$T_s = T_m\,(1 - 10^{-A/10}) \tag{21.6}$$

where:

T_s = sky-noise temperature (K) as seen by the antenna
A = path attenuation (dB)
T_m = effective temperature (K) of the medium

The effective temperature of the medium (T_m) is dependent on the contribution of scattering to the attenuation, the physical extent of clouds and rain cells, the vertical variation of the physical temperature of the scatterers and, to a lesser extent, on the antenna beamwidth. By comparing radiometric observations and simultaneous beacon attenuation measurements, the effective temperature of the medium at frequencies between 10 and 30 GHz has been determined to lie in the range 260–280 K for rain and clouds.

When the attenuation is known, the following effective temperatures of the atmospheric media may be used to obtain an upper limit to the sky-noise temperature at frequencies below 60 GHz:

$$T_m = 280 \text{ K for clouds} \tag{21.7}$$

$$T_m = 260 \text{ K for rain} \tag{21.8}$$

The sun and, to a lesser extent, the moon can occasionally contribute significantly to the noise temperature of an earth-station antenna at all frequencies. In addition, radio noise from other discrete sources such as Cygnus A and X, Cassiopeia A, Taurus and The Crab nebula could also be important. Radio noise from the cosmic background is only of importance at frequencies below about 2 GHz. For further information on sources of radio noise and methods for determining the system noise temperature of an earth station see Recommendation ITU-R P.372 [12].

21.8 Cross-polarisation effects

21.8.1 General

The capacity of space telecommunication systems can be increased by reusing frequencies on orthogonal polarisations. However, this technique can be restricted by depolarisation of the signal as it passes through the atmosphere. Of particular importance is the depolarisation produced by hydrometeors (i.e. rain or ice particles in various forms) in the troposphere and, at frequencies below 4 GHz, depolarisation produced by the ionosphere.

Depolarisation by hydrometeors can occur due to:

(a) The nonspherical (oblate) shape of large raindrops. This produces differential attenuation and phase shift between components of the incident wave polarised parallel and perpendicular to the principal axes of the raindrop.

(b) Ice particles, in the form of needles and plates. In this case the depolarisation results from a differential phase shift between components of the incident wave polarised parallel and perpendicular to the principal axes of the ice crystals.

The physical basis for both these effects is discussed in some detail in Chapter 12.

Depolarisation by the ionosphere results from Faraday rotation of the plane of polarisation of a linearly polarised wave by the ionosphere. This is discussed in Chapter 16 and also in Recommendation ITU-R P.531 [5]. As much as 1° of rotation may be encountered at 10 GHz, and greater rotations at lower frequencies. As seen from an earth station, the planes of polarisation rotate in the same direction on the uplinks and downlinks. Therefore, if the same antenna is used both for transmitting and receiving it is not possible to compensate for Faraday rotation by rotating the feed system of the antenna.

21.8.2 Calculation of long-term statistics of hydrometeor-induced cross polarisation

Recommendation ITU-R P.618 gives a method for predicting the long-term statistics of hydrometeor-induced cross polarisation from the long-term statistics of rain attenuation.

The following terms are used in the method:

(i) A frequency-dependent term (C_f): this takes account of the frequency dependence of the depolarisation produced by rain.
(ii) A term dependent on the statistics of rain attenuation (C_A): this is based on the observed correlation between the depolarisation and the attenuation produced by rain.
(iii) An elevation angle dependent term (C_θ): large raindrops can be modelled as oblate (flattened) spheres having their minimum dimension (i.e. the dimension along their axis of symmetry) aligned with a distribution of angles (canting angles) having a mean which is approximately vertical. Hence, the raindrop asymmetry, when viewed along the path, will decrease as the path elevation increases i.e. the raindrop cross section normal to the path will change from being approximately elliptical for a low elevation path to being approximately circular for a vertical path (an extreme case). Since, the depolarisation of a wave passing through a collection of raindrops is directly related to the asymmetry in the cross section normal to the path, waves traversing low-angle paths will suffer more depolarisation than those traversing high-angle paths. The elevation angle dependent term (C_θ) takes account of this variation.
(iv) A term giving the polarisation improvement factor (C_τ): this takes account of the polarisation type (linear or circular) and, if applicable, the orientation of the linearly polarised wave relative to the axis of symmetry of the raindrops. In the case of a linearly polarised wave, if τ is the tilt angle of the electric field vector with respect to the horizontal, then the improvement factor $C_\tau = 0$ for $\tau = 45°$ and reaches a maximum value of 15dB for $\tau = 0°$ or $90°$. For circular polarisation a value of 45° should be used for τ. Thus, a linearly polarised wave suffers the greatest depolarisation when it is aligned at 45° to the horizontal (equal to that for a circularly polarised wave) and the least depolarisation when it is aligned either vertically or horizontally.
(v) A raindrop canting angle dependent term (C_σ): This takes account of the spread in the raindrop canting angles around the distribution mean value. The effective standard deviation of the raindrop canting angle distribution expressed in degrees (σ) has a value of about 0°, 5°, 10° and 15° for 1 per cent, 0.1 per cent, 0.01 per cent and 0.001 per cent of the time, respectively.

First, the XPD (not exceeded for p per cent of the time) associated with the depolarisation produced by rain (XPD_{rain}) is calculated using:

$$XPD_{rain} = C_f - C_A + C_\tau + C_\theta + C_\sigma \tag{21.9}$$

Then, based on the observed correlation between rain and ice depolarisation, the XPD (not exceeded for p per cent of the time) associated with the depolarisation produced by ice crystals (C_{ice}) is calculated from XPD_{rain} using:

$$C_{ice} = XPD_{rain} \times \left[\frac{0.3 + 0.1 \log p}{2}\right] \tag{21.10}$$

Finally, the total XPD not exceeded for p per cent of the time (XPD_p), taking account of the effects of both ice and rain, is calculated using:

$$XPD_p = XPD_{rain} - C_{ice} \tag{21.11}$$

21.8.3 Joint statistics of XPD and attenuation

The joint statistics of XPD and attenuation may be required in some applications i.e. the probability that the XPD is less than a given value while at the same time a given attenuation is exceeded. The conditional probability distribution of XPD for a given value of attenuation, A_p, can be modelled by assuming that the cross polar to co-polar voltage ratio, $r = 10^{-XPD/20}$, is normally distributed. The parameters of the distribution are the mean value, r_m, which is very close to $10^{-XPD_{rain}/20}$, with XPD_{rain} given by Eqn. 21.9. The standard deviation of the distribution, σ_r, has the almost constant value of 0.038 for 3 dB $\le A_p \le$ 8 dB.

21.8.4 Long-term frequency and polarisation scaling of statistics of hydrometeor-induced cross polarisation

If long-term XPD statistics obtained at one frequency and polarisation tilt angle are available they can be scaled to another frequency and polarisation tilt angle using the following semi-empirical formula:

$$XPD_2 = XPD_1 - 20 \log\left[\frac{f_2\sqrt{1 - 0.484(1 + \cos 4\tau_2)}}{f_1\sqrt{1 - 0.484(1 + \cos 4\tau_1)}}\right] \tag{21.12}$$

where

$$4 \le f_1, f_2 \le 30 \text{ GHz}$$

and XPD_1, XPD_2 are the XPD values not exceeded for the same percentage of time at frequencies f_1 and f_2 and polarisation tilt angles, τ_1 and τ_2, respectively.

Eqn. 21.12 can be used to scale XPD data that include the effects of both rain and ice depolarisation, since it has been observed that both phenomena have approximately the same frequency dependence at frequencies less than about 30 GHz.

21.9 Bandwidth limitations

Anomalous dispersion in the vicinity of the absorption lines of atmospheric gases produces small changes in the refractive index. However, these refractive index changes are small in the bands allocated to earth–space communications, and will not restrict the bandwidth of systems.

Multiple scattering in rain can limit the bandwidth of incoherent transmission systems due to variation in time delays of the multiple-scattered signals, however, under these conditions the attenuation will present a far more serious problem. A study of the problem of bandwidth limitations imposed by the frequency dependence of attenuation and phase shift due to rain on coherent transmission systems showed that such bandwidth limitations are in excess of 3.5 GHz for all situations likely to be encountered. These are greater than any bandwidth allocated for earth–space communications below 40 GHz, and the rain attenuation will therefore be far more important than its frequency dependence.

21.10 Calculation of long-term statistics for nonGSO paths

The prediction methods presented in Recommendation ITU-R P.618 were originally developed for applications where the path elevation angle remains constant. However, prediction methods are also required for nongeostationary satellite (nonGSO) systems, where the elevation angle is continuously varying. A method, applicable to single satellite nonGSO systems, is given in Recommendation ITU-R P.618.

The approach used is to calculate the minimum and maximum elevation angles at which the system will be expected to operate and to divide this operational range into small angular increments (e.g. 5° wide). The percentage of time (p_i) for which the satellite is visible within each small angular increment range is then calculated. For a given propagation impairment level, the percentage of time (p_j) for which the level is exceeded for a fixed path having an elevation angle equal to the mid value of each small angular increment range is determined. The time percentages for which the impairment level is exceeded at each elevation angle are calculated from p_i and p_j and these are then summed over all the path elevations within the operational range to obtain the total time percentage for which the impairment level is exceeded.

In the case of multivisibility satellite constellations employing satellite path diversity (i.e. switching to the least impaired path), an approximate calculation can be made assuming that the spacecraft with the highest elevation angle is being used.

21.11 Mitigation techniques

21.11.1 Site diversity

The intense rain cells that cause large attenuation on earth–space links often have horizontal dimensions of no more than a few kilometres. Diversity systems able to reroute traffic to alternate earth stations, or with access to a satellite with extra onboard resources available for temporary allocation, can improve the system reliability considerably.

Two parameters are currently used to characterise diversity performance:

(i) The diversity improvement factor (I): defined as the ratio of the single-site time percentage (p_1) to the diversity time percentage (p_2), at the same attenuation level.
(ii) The diversity gain (G): defined as the difference (dB) between the single-site and diversity attenuation values for the same time percentage.

Both the diversity improvement factor and the diversity gain are mainly determined by the distance between the two sites, increasing with increasing site separation, and are only slightly dependent on the path elevation angle and the frequency.

Recommendation ITU-R P.618 gives procedures for calculating the diversity improvement factor and the diversity gain for systems operating in the frequency range 10 GHz to 30 GHz. These procedures are only recommended for time percentages of less than 0.1 per cent. For time percentages above 0.1 per cent, the rainfall rate is generally small and the corresponding site diversity improvement is not significant.

21.11.2 Transmit power control

A method of reducing the effects of rain fading is to increase the earth-station transmitter power to compensate for the loss of up-path signal using the fade on the down path to determine the amount by which the transmit power must be increased. To do this data are required on the instantaneous ratio between the attenuation produced by rain at the up-path frequency to that at the down-path frequency. This scaling ratio has been found to be log-normally distributed, and is influenced by rain type and rain temperature. Data reveals that the short-term variations in the attenuation ratio can be significant, and are expected to increase with decreasing path elevation angle.

21.11.3 Cross-polarisation cancellation

System capacity can be increased, without increasing the required bandwidth, by frequency reuse, i.e. using the same band of frequencies on orthogonal polarisations. However, for this to be effective adequate separation between signals on the two polarisations must be maintained. A method for reducing the interference that may result from the depolarised component of a signal appearing in the orthogonally polarised channel is to take a small proportion of that signal (equal in amplitude to the interference) and add it in antiphase to the interference in the orthogonally polarised channel. Using this technique some or all of the interference can be cancelled out.

Since the amplitude and phase of the signal required to carry out this cancellation is continually changing, due to variations in the propagation medium, the equipment required to derive the correct cancellation signal must continually track the amplitude and phase changes. Hence, key parameters in the design of the cancellation system are the magnitude and rates of change of amplitude and phase of the required cancellation signal.

Measurements at 6 and 4 GHz have shown that 99 per cent of the XPD variations are slower than ± 4 dB s^{-1}, or equivalently, less than $\pm 1.5°$ s^{-1} in the mean path differential phase shift. Therefore, the time constant of a depolarisation compensation system at these frequencies need only be about one second.

Experiments have also shown that a strong correlation exists between rain depolarisation at 6 and 4 GHz on earth–space paths, both on the long term and on an event basis, and uplink depolarisation compensation utilising concurrent downlink depolarisation measurements appears feasible. Only differential phase effects were apparent, even for severe rain events, and single-parameter compensation (i.e. for differential phase) appears sufficient at 6 and 4 GHz.

21.12 References

1 Recommendation ITU-R P.618-6: 'Propagation data and prediction methods required for the design of earth–space telecommunications systems', 1999
2 Recommendation ITU-R P.452-9: 'Prediction procedure for the evaluation of microwave interference between stations on the surface of the earth at frequencies above about 0.7 GHz', 1999
3 Recommendation ITU-R P.619-1: 'Propagation data required for the evaluation of microwave interference between stations in space and those on the surface of the earth', 1992
4 Recommendation ITU-R P.1412: 'Propagation data for the evaluation of coordination between earth stations working in the bidirectionally allocated frequency bands', 1999
5 Recommendation ITU-R P.531-5: 'Ionospheric effects influencing radio systems involving spacecraft', 1999
6 Recommendation ITU-R P.676-4: 'Attenuation by atmospheric gases', 1999
7 Recommendation ITU-R P.834-3: 'Effects of tropospheric refraction on radiowave propagation', 1999
8 Recommendation ITU-R P.837-2: 'Characteristics of precipitation for propagation modelling', 1999
9 Recommendation ITU-R P.841-1, 'Conversion of annual statistics to worst-month statistics', 1999

10 Recommendation ITU-R P.678-1: ‘Characterization of the natural variability of propagation phenomena’, 1992
11 Recommendation ITU-R P.840-3: ‘Attenuation due to clouds and fog’, 1999
12 Recommendation ITU-R P.372-6: ‘Radio noise’, 1999

Index

Propagation of Radiowaves

2nd Edition

The field of radio communications continues to change rapidly, and the second edition of this outstanding book, based on a popular IEE Vacation School, has been fully updated to reflect the latest developments. The introduction of new services and the proliferation of mobile communications have produced a growing need for wider bandwidths and the consequent need for frequency reuse. This book introduces the basic concepts and mechanisms of radiowave propogation engineering in both the troposphere and ionosphere, and includes greater emphasis on the needs of digital technologies and new kinds of radio systems. The content reflects the wide experience of an exceptional group of authors and the emphasis throughout is on modelling and prediction methods and the relevant ITU Radiocommunication Sector recommendations. *Propagation of Radiowaves, 2nd Edition* is essential reading for professionals involved in planning, designing and operating radio systems, as well as postgraduate students in the field.

Les Barclay OBE is a consultant in radiowave propagation and spectrum management. He is also visiting professor at the universities of Lancaster, Surrey and Bradford, where he lectures in radio regulation, spectrum management and propogation. He has been involved in the organisation and lecturing for all IEE residential courses in radiowave propagation, and has latterly chaired the organizing committee for these courses Prior to this, he was the deputy director at the Radiocommunications Agency, responsible for research. He also acted as international chairman of ITU-Study Group 3 on radiowave propagation and chairman on the IEE professional groups on antennas and propagation, as well as on radiocommunications. He is a Fellow of both the Royal Academy of Engineering and the IEE.

ISBN 978-0-85296-102-5

9 780852 961025 >

The Institution of Engineering and Technology
www.theiet.org
0-85296-102-2
978-0-85296-102-5