

hib

AGARDograph 104



**High Frequency**

**Radio Communications**

**with Emphasis on Polar Problems**

*Editor*

**FINN LIED**



## AGARD

The Advisory group for Aerospace Research and Development (AGARD) initiated in 1951 is part of the North Atlantic Treaty Organization and consists of a number of permanent specialist scientific panels and committees from NATO countries. These panels are responsible for sponsoring technical meetings and symposia and the publication of technical papers.

Part of the AGARD publishing programme takes the form of good quality letterpress or litho, hard cover volumes of symposia proceedings or monographs on significant areas of astronomical and aeronautical research and development.

Technivision of Maidenhead, England have set up a publishing organization based on their wide technological background and utilizing their existing staff of Authors and Editors together with expanded art and production departments.

### AGARDograph 104

This handbook, amply illustrated throughout, is the result of the very valuable work done by staff members of the Norwegian Defence Research Establishment. A great deal of knowledge has been gained in recent years about the ionosphere in polar regions and the complex way in which the sun's energy interacts with this part of the earth's atmosphere. Lack of this information has restricted the full use of available communications in polar regions. It was to make this very necessary knowledge more widely available that caused this book to be published. Although this volume is designed to be the constant companion of Radio Communications Officers, it has been written to the standard of pre-university physics thus making the information understandable to amateurs and professionals alike and, with the increase of Solar activity due in the next few years, its information will be of immediate value.

## FUTURE PUBLICATIONS

### AGARD Conference Proceedings

#### Series No. 3

##### "PROPAGATION FACTORS IN SPACE COMMUNICATIONS"

This volume, edited by W. T. Blackband of the R.A.E. Farnborough, presents the text of papers read (and the ensuing discussions) at the 1965 annual symposium dealing with the propagation of electromagnetic waves in space. The Information covers a broad field and includes radar measurements, radio communications, propagation in a plasma sheath and some information on the ionosphere of Mars. The frequency spectrum covered extends from V.L.F. to laser beams.

### AGARDograph 106

##### "SOME EFFECTS OF RAISED INTRAPULMONARY PRESSURE IN MAN"

This monograph by Wing Cdr. J. Ernsting, PhD., B.Sc., M.B., B.S., R.A.F., is concerned with the effects on the body of raised intrapulmonary pressure and the reduction of these effects by applying compensating pressures to specific regions of the body. The information is of special value to those concerned with the protection of personnel engaged in high altitude or space flight and of considerable general interest to those engaged in the physiology of the respiration and circulation systems.

### AGARD Conference Proceedings

#### Series No. 6

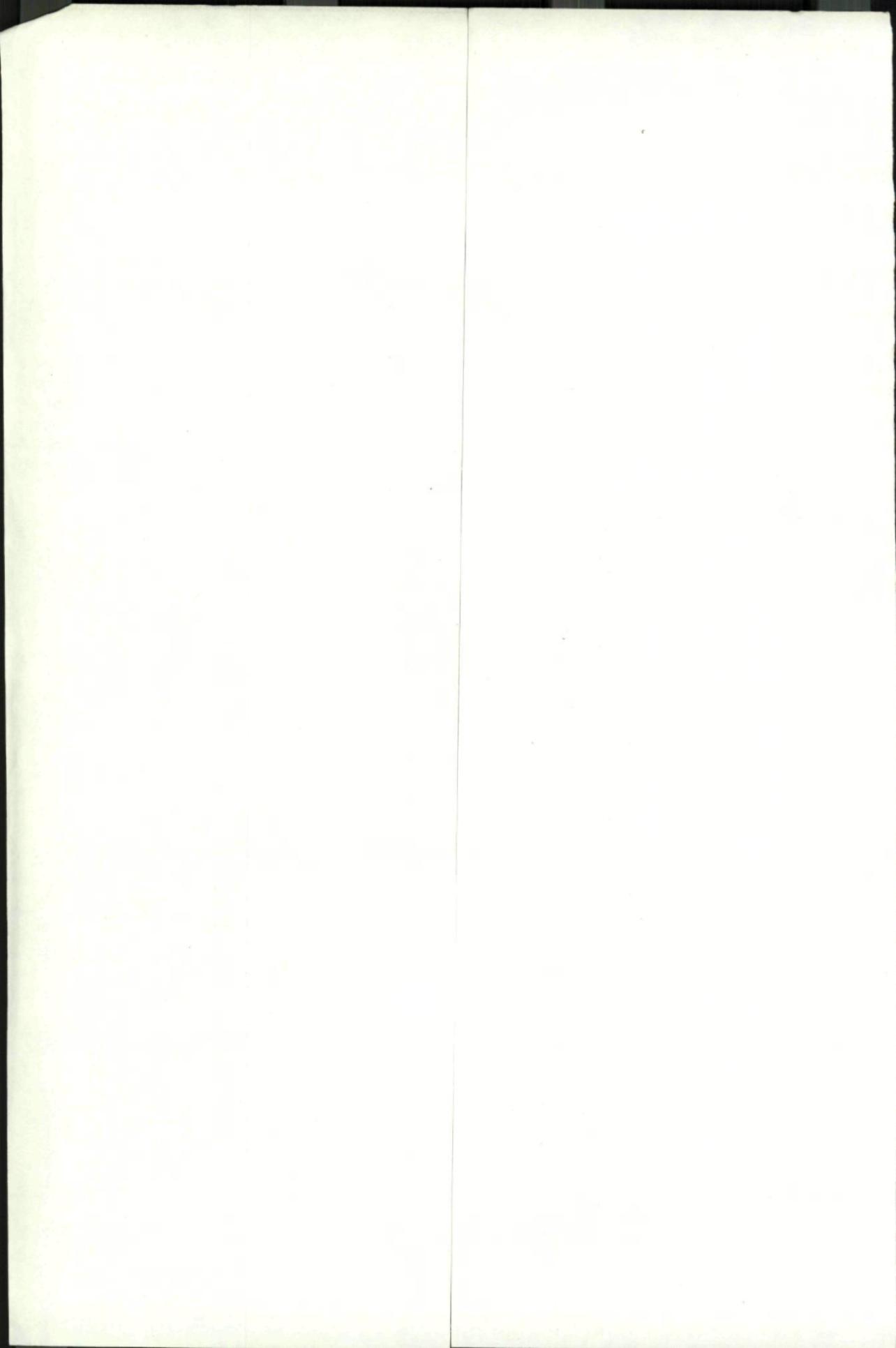
##### "AIRCRAFT FLIGHT INSTRUMENTATION INTEGRATED DATA SYSTEMS"

These proceedings, edited by D. Bosman, are the reports of the Tenth Symposium of AGARD Avionics Panel which discussed the aircraft flight recording and instrumentation systems. This book will interest those concerned in any branch of civil or military aeronautics and particularly researchers engaged in aircraft safety.

### AGARD Bibliography No. 5

##### "A BIBLIOGRAPHY OF REFRACTORY METALS"

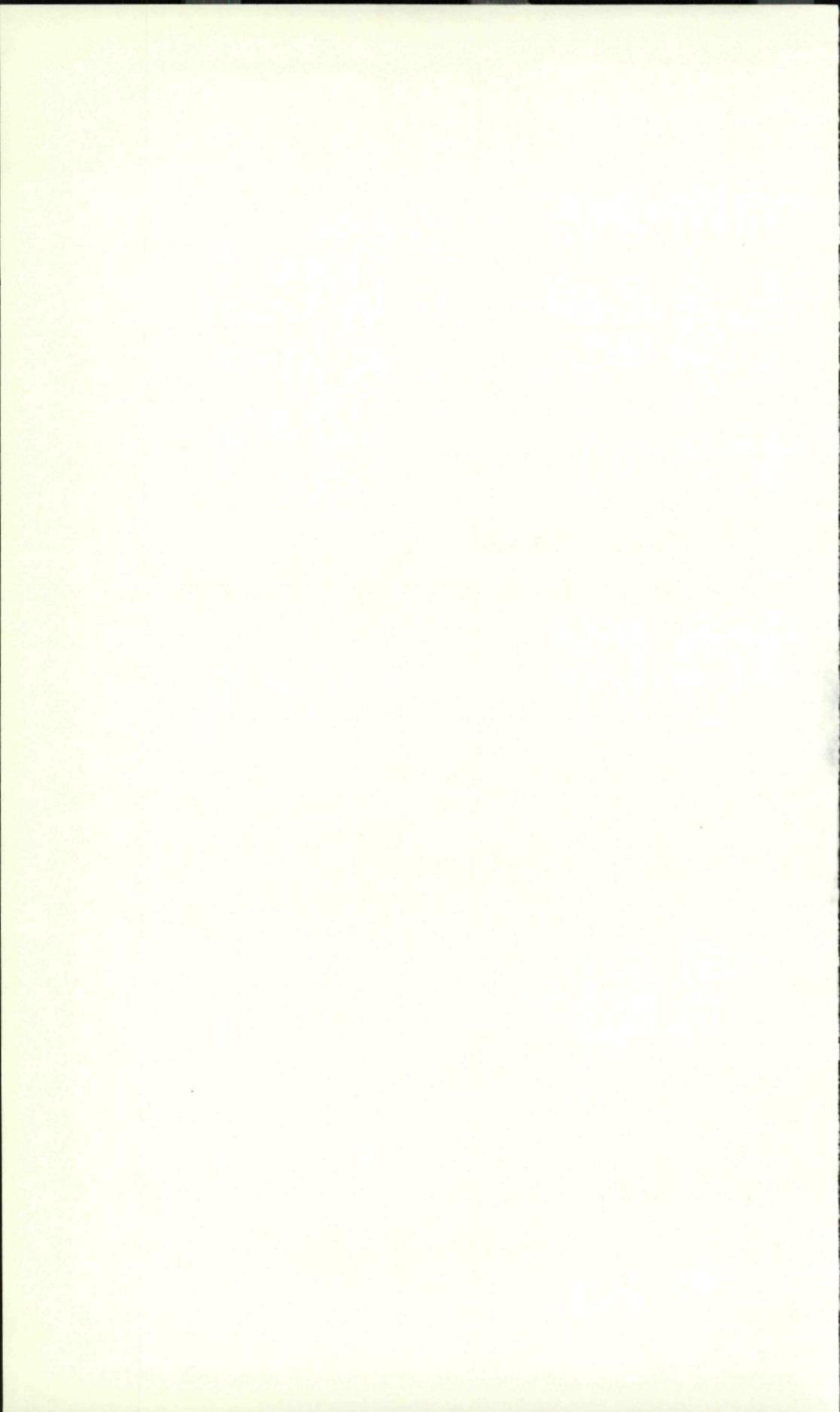
This reference work, which is believed to be the first published in its field was compiled by R. Syre, an authority in the field of metallurgy. The book is letterpress printed; the layout so designed to facilitate rapid consultation and incorporating an authors' index and list of patents.



ROYAL AIRCRAFT  
ESTABLISHMENT

13 MAR 1968

LIBRARY

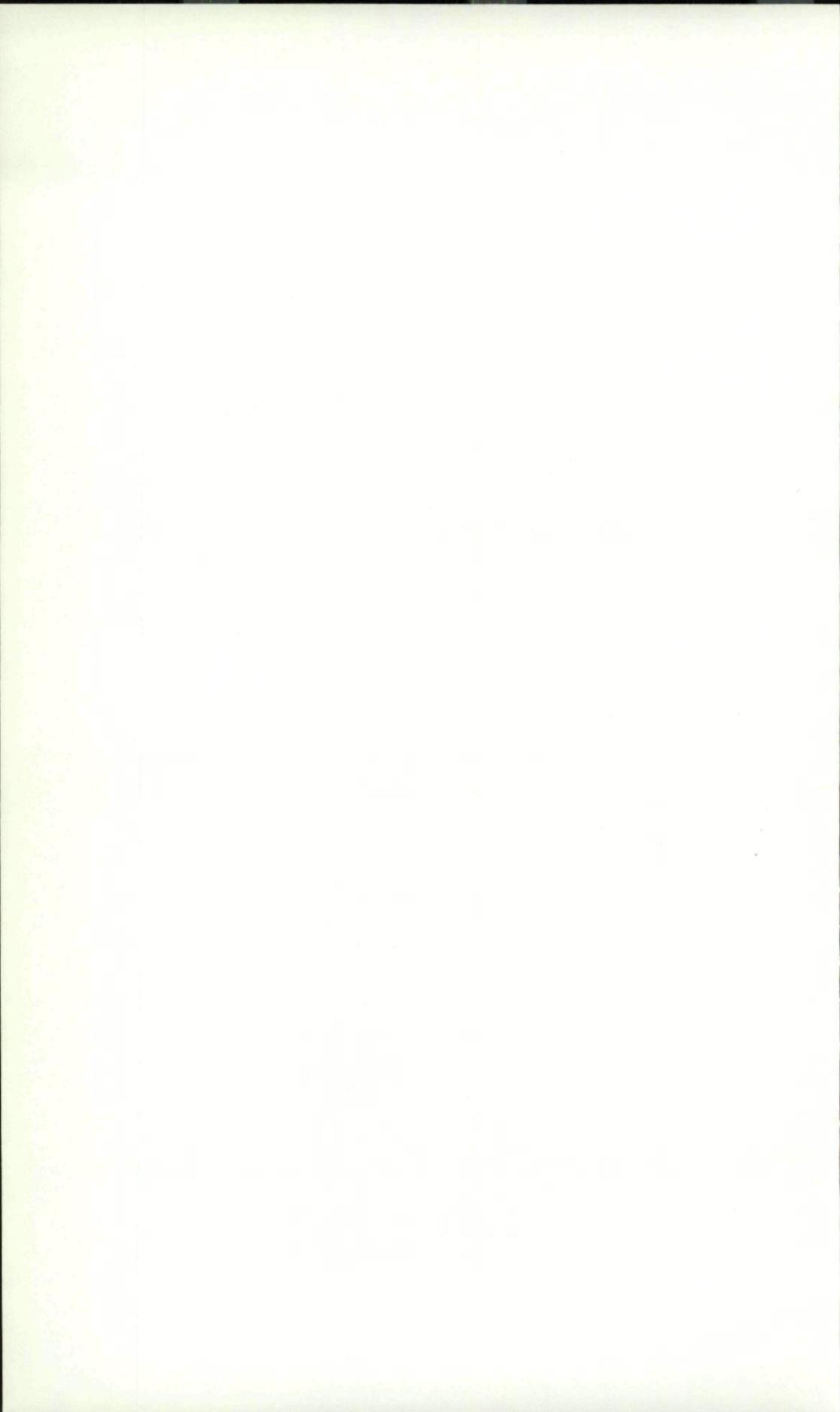


AGARDograph Number  
One Hundred and Four

**High Frequency  
Radio Communications  
With Emphasis on Polar Problems**



The Advisory Group for  
Aerospace Research and  
Development. NATO



Editor

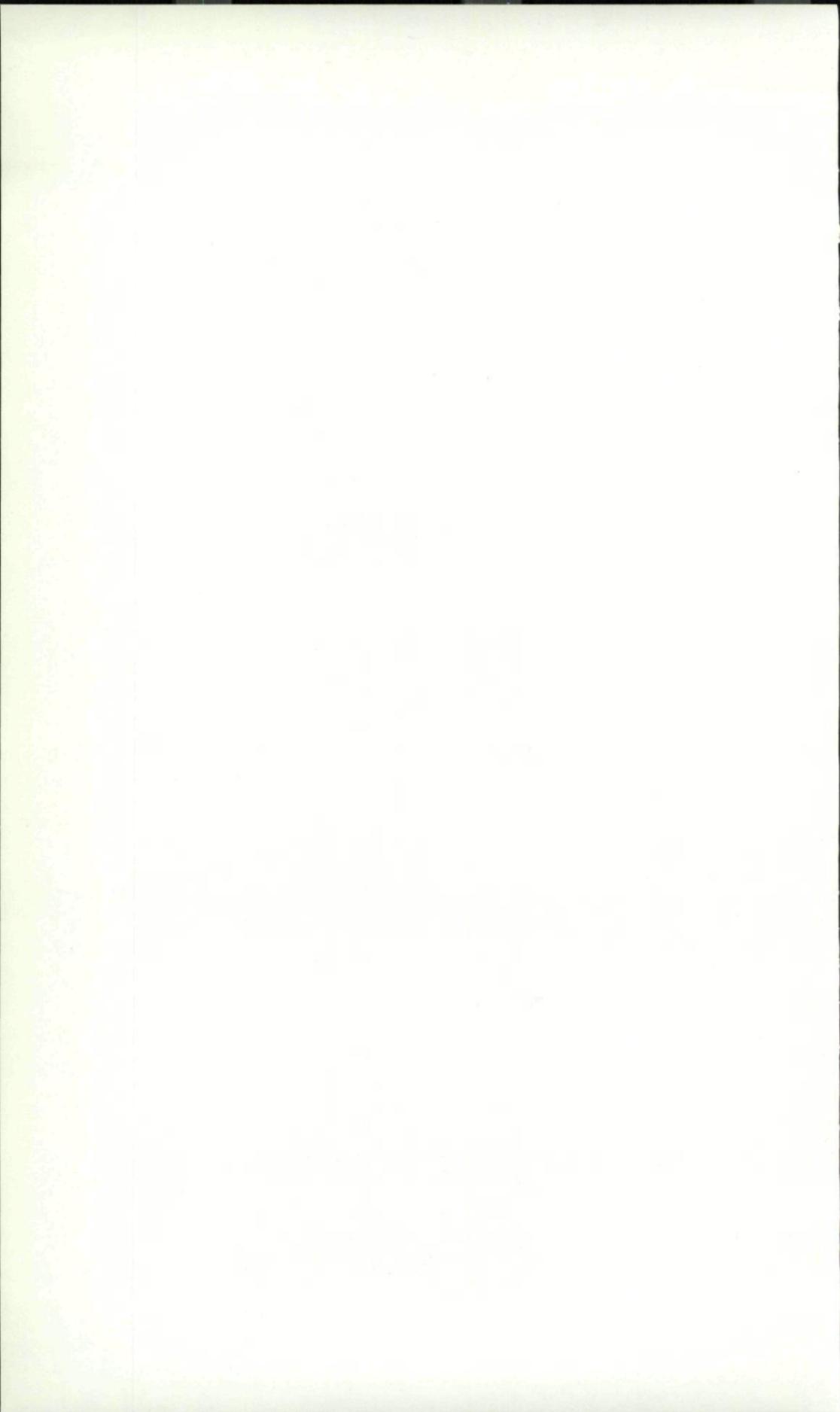
**FINN LIED**

Authors

**Knut W. Eriksen  
Bjørn Landmark  
Finn Lied  
Bernt Mæhlum  
Eivind V. Thrane**

Review

**B. Burgess**



Published by



Technivision  
Maidenhead, England

Printed by



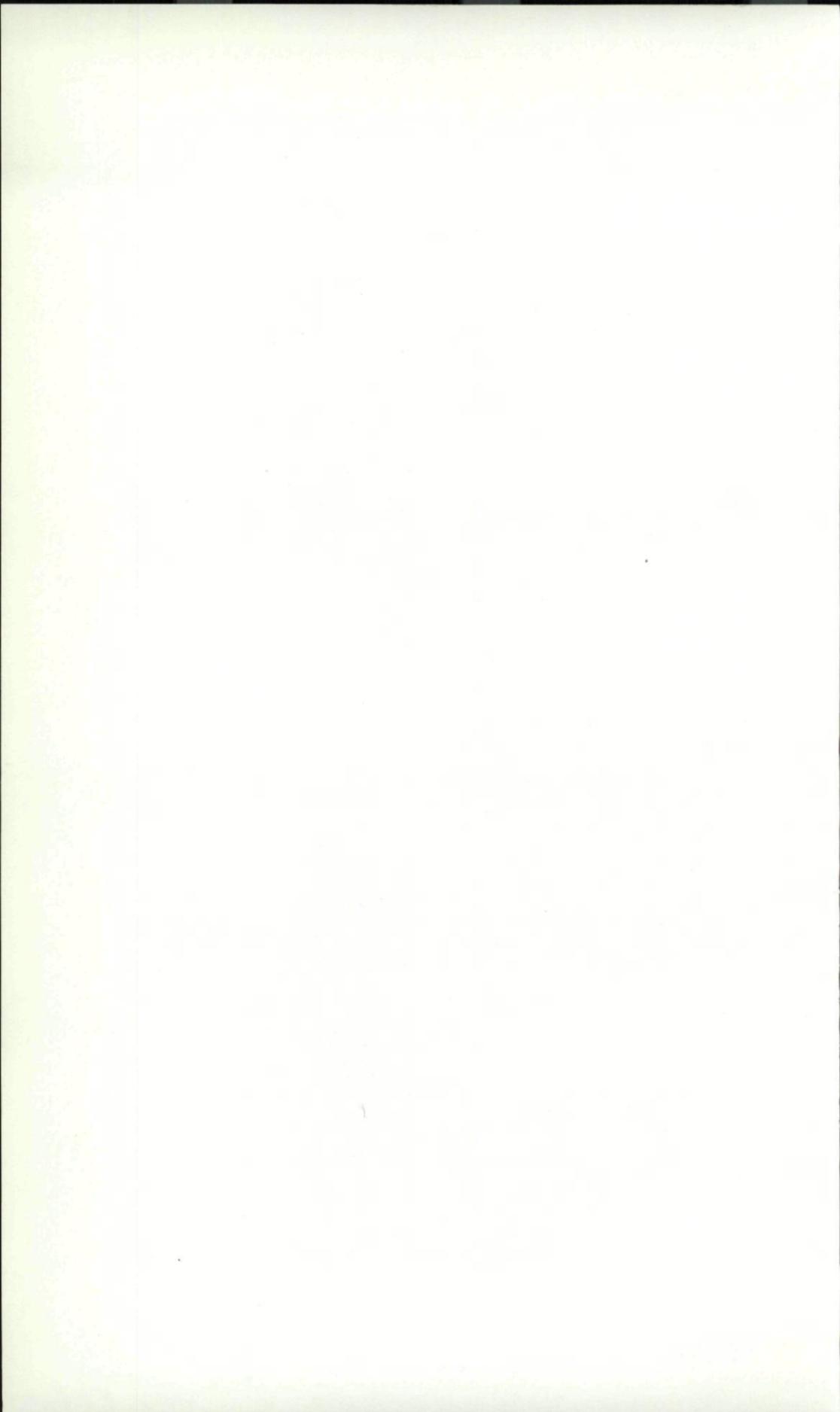
Unwin Brothers Limited  
Old Woking, Surrey, England

Copyright



© November 1967  
The Advisory Group for Aerospace  
Research and Development. Nato

Library of Congress Catalog Card No. 67-22235



# Contents

	Page
Review	I
Preface	III
Summary	V
Acknowledgements	VI
Chapter 1	1
Propagation of Radio Waves via The Ionosphere	1
Introduction	1
Principle of Radio Communication Via the Ionosphere	1
Variations of the Ionospheric Layers	16
Maximum Useable Frequency	29
Absorption, Radio Noise and Lowest Useful High Frequency	34
Chapter 2	53
Prediction of Ionospheric Conditions	53
Introduction	53
Methods of Predictions	53
Form and Presentation of Predictions	58
Reliability and Use of Predictions	74
Chapter 3	79
Communication Problems in the Polar Regions	79
Introduction	79
The Polar Ionosphere	79
Methods of Study	83
Polar Cap Absorption (PCA)	93
Auroral Absorption	95
High Latitude Sporadic E	100
Communication in Other Frequency Bands	105
Chapter 4	107
Prediction and Reduction of Polar Communication Difficulties	107
Introduction	107
Predictions of Disturbances	107
De-Routing of Traffic	112
Oblique Incidence and Backscatter Sounding	116
Appendix 1	125
The Earth's Magnetic Field and its Time Variations	125
Subject Index.	127



## Review

The Ionospheric Research Committee (since renamed The Electromagnetic Wave Propagation Committee) of the Avionics Panel, AGARD, recommended in August 1964, the production of an AGARDograph dealing with the problems associated with high frequency radio communications with particular emphasis being placed on communications in polar regions. A great deal of knowledge has been gained in recent years about the ionosphere in polar regions and the complex way in which the sun's energy interacts with this part of the earth's atmosphere. This information, with the consequent bearing on high frequency radio communication in these regions, was not felt to be widely appreciated by wireless operators and therefore, the full use of the available communication circuits was not possible. The dissemination of this information to service users of radio communications is in line with the policy of AGARD and hence the compilation of this handbook by Staff Members of the Norwegian Defence Research Establishment, who have made a great contribution to our knowledge of the Polar Ionosphere and its vagaries.

The authors state the main purpose of this book to be a provision of simple up to date information on different problems which may arise in radio wave communication, to give simple explanations for these and, when possible, point out how to circumnavigate these problems. The book, which, by its nature, devotes a lot of space to radio ionospheric communications in general, is intended to emphasise the problems that are particularly relevant to communication at high latitudes. It is aimed at the communications officer, who has a practical job to do. In achieving this aim, the authors require, for an understanding of this book, no more knowledge, as a start, than pre-university physics.

The handbook is divided into four chapters. The first chapter deals with radio-wave communications under normal ionospheric conditions. Here the basic principles of radiowave propagation via the ionosphere are developed in an easy to follow manner. The formation of the ionosphere, methods of studying it and its normal variations are described. This then leads to the concept of frequency limits imposed on high frequency communication. The upper limit, called the Maximum Usable Frequency (MUF), is governed by the properties of the ionosphere; the lower limit, the Lowest Usable High Frequency (LUHF), is determined by ionospheric absorption and through this, on the transmitter power, aerial efficiency and noise levels at or on the receiver.

Having discussed the basic principles of ionospheric radio communications, the problem of predicting the future behaviour of the ionosphere, both long term (months and years) and short term is tackled. The important propagation parameters are discussed and the methods involved and assumptions made in predictions presented. The reliability of the predictions, the reason for their failure and what to do when they do fail are important points which are discussed. This chapter is amply illustrated with practical examples, as indeed is the whole book; a necessary feature for this type of publication.

The authors, having built up the basic principles which govern the choice of the optimum communication circuit, move onto the topic which motivated the compilation of the handbook, viz., communication problems in polar regions. The final two chapters of the book are devoted to the problems of prediction and reduction of communication difficulties. Here the results of recent researches on the polar ionosphere are put into the context of communication problems. Familiarity with the pattern of events which accompany the numerous disturbances of the polar ionosphere will enable the operator to derive more benefit from warnings of disturbances. On this basis a great deal of trouble has been taken to describe the origin of the disturbances, their life cycle and how the various propagation parameters of importance to communications are likely to

vary. In particular the properties of polar cap absorption events, auroral absorption and high latitude sporadic E are discussed in detail. As an example of how they effect communications the following is a useful illustration. During geomagnetically disturbed conditions high frequency radio waves are heavily absorbed due to the enhanced ionization in the D region. However in polar regions an intense sporadic E layer is also formed under these conditions. This sporadic E layer will support a higher operating frequency, which will suffer less absorption by the D region, hence a communication circuit could be re-established.

While the information presented in this book is based on a wealth of published material which can be found in the technical journals, it is the first time that these important aspects of polar ionospheric disturbances have been gathered together and applied to the radio communication problem in a fashion that will enable the radio communication officer to use it to great advantage. This book will be a definite asset to any communication station and with increased solar activity due in the next few years, its pages will be well thumbed. As a byproduct, many amateur radio operators will find it a handy companion. The book, written in excellent English by five Norwegians, is easy to read and has an attractive style of presentation, which will go a long way towards helping the reader to get to grips with the complex phenomena of the regular and irregular interaction of solar energy with the earth's ionosphere and the resultant important consequences to radio communications.

# Preface

All signalling by means of radio occurs as the result of waves that travel from the transmitter to a receiver. These waves, which are electromagnetic, are generated by rapidly alternating currents in the antenna of the transmitter. Oscillating vacuum tubes or solid state devices within the transmitter are responsible for this alternating current. From the antenna, the electromagnetic waves spread out in all or in somewhat preferred directions with the velocity of light, 300,000 km per second. The distant receiving antenna intercepts only a small fraction of the wave energy that is radiated by the antenna of the transmitter. The receiving set amplifies these weak electrical signals and converts them into signals that can be audibly detected or otherwise recorded.

We shall not deal with the transmitter or the receiver itself in this book nor with the different ways in which the intelligence may be imprinted upon the high frequency waves, which act as carriers. We shall want to know only the power output from the transmitter and that the receiver is a communications equipment of high sensitivity. The limit of its receptive power will depend on the external static level of natural or man-made character upon the inevitable noise level of a receiver. Our problem starts when the waves leave the transmitting antenna and ceases when they enter the receiver.

Two types of waves travel from the transmitter to the receiver. The sky-wave travels through the atmosphere and the ground-wave skirts along the curved surface of the earth. The distance which the ground-wave travels is for many purposes very important. In this book we will deal only with the sky-wave which is particularly important for long-range communication.

The early history of sky-wave propagation is interesting. After Marconi's first experiments, it was proved by mathematical physicists that even on long waves it would be impossible to receive ground wave signals at very great distances because of the attenuation the waves would suffer when being bent round the curved surface of the earth. When, in 1901, it was found experimentally that signals could be received across the Atlantic Ocean, these mathematical results were questioned. The mathematics were, however, correct but not the assumption made that the earth was surrounded by free space. It was suggested independently by Heaviside and by Kennelly in 1902, that the earth was surrounded by an ionized layer that acted as a reflector for the waves and prevented them from escaping into free space. In these layers currents could also flow which, according to Balfour Stewart, might explain the changes in the earth's magnetic field during magnetic storms. Some years later, in 1912, Eccles showed mathematically how such an ionized layer could behave like a reflector by bending the waves round by a process of refraction and returning them to earth. The existence of waves, reflected down to earth in this way, was demonstrated experimentally some years later in 1925 by Sir Edward Appleton and his co-workers, who have also shown that, ionization of the upper atmosphere exists in the form of several layers at different heights. The name 'ionosphere' was suggested by Watson-Watt and has been universally adopted.

The first layer to be detected was the ionospheric E-layer. This is situated at a height of about 120 km which may appear to be a considerable height but, relative to the radius of the earth (6300 km), it is, on a global scale, nevertheless, very close to the earth. To-day, our interest reaches very much further out and we will have to concern ourselves with nearly the whole space between the sun and the earth when trying to understand the phenomena of long distance communications.

The aim of this book is to assist the practical communication officer and engineer in his endeavour to 'get the message through'. The hope is that, by understanding the physical processes, possibilities and limitations, he may be

able to use his equipment to the limit of its possibilities and also adapt himself in an optimum way to the whimsicality of nature.

In Chapter 1 the basic physical properties of radio wave propagation are described while the prediction of ionospheric conditions is dealt with in Chapter 2. It is commonly known that the polar ionosphere is frequently disturbed and this creates great difficulties for radio wave propagation via the polar ionosphere. In Chapter 3 both a description of the polar ionosphere and a discussion of the different types of disturbances are presented. Lastly, in Chapter 4, an attempt is made to describe the different possibilities that exist in order to predict the onset of such disturbances.

It is apparent that this book does not present any new information. On the contrary, it draws heavily on a wealth of published material but it has been the intention and hope of the authors that it will prove of value to have put this material together in a uniform style and for the particular purpose of assisting the communication officer and engineer, especially when working in polar regions.

## Summary

The aim of this book is to assist the practical Communication Officer and Engineer in radio communication, particularly when working in Polar Regions.

All signalling by means of radio, occurs as the result of waves, that travel from the transmitter to a receiver. Of the two types of waves, the sky-wave and the ground-wave, we deal here only with the sky wave, which is particularly important for long-range communication. The transmitter and the receiver themselves are not investigated, nor the different ways intelligence may be imprinted upon the high-frequency waves which act as carriers. We want only to know the power output from the transmitter and that the receiver is a communications equipment of high sensitivity. The limit of receptive power will depend on external static level of natural or man-made character or upon the inevitable noise level of a receiver. Our problem starts when the waves leave the transmitting antenna and ceases when they enter the receiver.

We are concerned here with nearly the whole space between the sun and the earth when trying to understand the phenomena of long-distance communication. The hope is that, by understanding the physical processes, equipment may be used to the limit of its possibilities.

## Acknowledgements

We would like to express our gratitude to Dr. Kenneth Davies of the Institute for Telecommunication Sciences and Aeronomy (I.T.S.A.), Environmental Science Services Administration, Boulder, Colorado, U.S.A., (formerly C.R.P.L. of the National Bureau of Standards) and to the I.T.S.A. Forecasting Center, Fort Belvoir, Virginia, U.S.A. for helpful advice in the preparation of this book. We are also grateful to a number of research workers for releasing their published and unpublished results for reproduction in this book. Finally, we acknowledge a pleasant collaboration with AGARD during the course of publication.

# 1

## Propagation of Radio Waves via the Ionosphere

### Introduction

In this chapter we will deal with the physical principles governing radio wave propagation via the ionosphere. The basic physical properties of the ionosphere will be described in a qualitative way and an attempt will be made to explain the reflecting and absorbing properties of the ionosphere.

The aim of this first chapter is to explain why radio wave propagation via the ionosphere in the High Frequency Band is possible between a higher limit, called the Maximum Useable Frequency and a lower limit called the Lowest Useful High Frequency. The upper limit is principally given by the properties of the ionosphere, chiefly the electron density, and does not depend upon transmitted power or any other factor under our influence. The lower limit is determined by the absorption in the ionosphere, but also on the radiated power, the efficiency of the aeriels involved and the noise level at and in the receiver. The lower limit is, therefore, a much more complicated limit to calculate and does not, in the same way as the upper limit, represent a physical boundary given by nature.

Our approach will be based on secondary school physics with some addition of more advanced nature, the understanding of which is not necessary for the full benefit of the book. This means that very important contributions, like the Appleton Hartree magnetoionic theory, will not be dealt with and also, that many concepts, like group and phase velocities, are only illustrated but not properly introduced.

This is, however, done deliberately in order to focus attention on the physical properties and of the needs of the communication officer in order to assist him in his practical work in getting his message through.

### Principle of radio communication via the ionosphere

In order to describe in detail the way in which the ionosphere influences a radio wave passing through it, we really ought to use a rather complicated mathematical treatment. We shall, however, content ourselves with a much simpler treatment similar to that used in geometrical optics. This will suffice for an understanding of most phenomena.

#### Wave propagation in a refractive medium

When a ray of light passes from one medium to another of greater refractive index, for example from air to glass, it is bent towards the normal, Fig. 1. 1.

The law of refraction is:

$$\mu_1 \sin i_1 = \mu_2 \sin i_2 \quad (\text{Eq. 1. 1})$$

This equation is frequently referred to as Snell's Law. The refractive index is a measure of the optical density of a material. The refractive index of vacuum is 1 and of glass approximately 1.4. This means that a ray going from air into glass is bent towards the normal since  $\sin i_2$  must be less than  $\sin i_1$ , that is  $i_2$  must be less than  $i_1$ .

It may be shown that the phase velocity of light is inversely proportional to the refractive index of a medium.

Since the velocity of light in vacuum is 300 000 km/s, the velocity of light in

a medium with refractive index  $\mu_2$  is:

$$v_2 = \frac{300\,000}{\mu_2} \text{ km/s} \quad (\text{Eq. 1. 2})$$

Suppose we have a series of slabs of materials with different refractive indices  $\mu_1, \mu_2, \mu_3$  etc. and the ray makes the angle  $i_1$  with the normal of the first slab, and an angle  $i_2$  in the second, and so on, Fig. 1. 2.

The law of refraction gives:

$$\mu_1 \sin i_1 = \mu_2 \sin i_2 = \mu_3 \sin i_3 = \dots = \text{constant} \quad (\text{Eq. 1. 3})$$

This means that, as the ray travels through the slabs, it changes its direction in such a way that the product of the local refractive index and the sine of the angle between the ray and the normal is constant. If we have a great many slabs, all relatively thin, this approximates a layer of continuously varying refractive index. The same law will apply,

$$\mu \sin i = \text{constant} \quad (\text{Eq. 1. 4})$$

and the ray will be bent towards or away from the normal as the ray travels into regions of greater or less refractive index, Fig. 1. 3. If medium 1 is free space, equation (1. 4) yields:  $\mu \sin i = \sin i_1$ . If the ray penetrates into a region where  $\mu = \sin i_1$ , this means that  $\sin i$  becomes equal to unity and  $i$  must be  $90^\circ$ . This gives us the case of reflection. If, therefore, the refractive index of a medium is less than unity, reflection of a wave may occur. At vertical incidence ( $i_1 = 0$ ), reflection can only occur where  $\mu = 0$ . At oblique incidence it suffices if the refractive index drops down to  $\sin i_1$ . Fig. 1. 3.

The ionosphere as a refracting medium

In order to understand the nature of the transmission of radio waves through the ionosphere, it is simplest at this stage to accept the fact that the upper atmosphere is ionized, that is, that it contains free negatively charged electrons and much heavier positively charged ions and neutral molecules. It is the free electrons which, on account of their light mass, can be moved in resonance with the oscillating electric field of the radio wave and thus control the refracting properties of the ionized medium. The most important parameter of an ionized layer is therefore the electron density, normally given as  $N$  electrons/ $\text{m}^3$ . We also frequently use the term electron density profile of a layer and by this we mean the electron density as a function of height  $N(h)$ . In Figure 1. 4 a typical E-layer profile is given as an example. If we want to apply the ideas of geometrical optics to the ionosphere, we will in the first instance have to make some simplifying assumptions:

- (a) The earth is flat
- (b) The effects of the earth's magnetic field may be neglected
- (c) The ionosphere is horizontally stratified

We will later do away with these assumptions, but they will make our start more manageable.

Radio waves are electromagnetic waves of the same character as light, only the wavelength is very much longer. From elementary physics we know that:

$$v = \lambda \cdot f \quad (\text{Eq. 1. 5})$$

where  $v$  is the velocity of the wave,  $\lambda$  its wavelength and  $f$  its frequency.

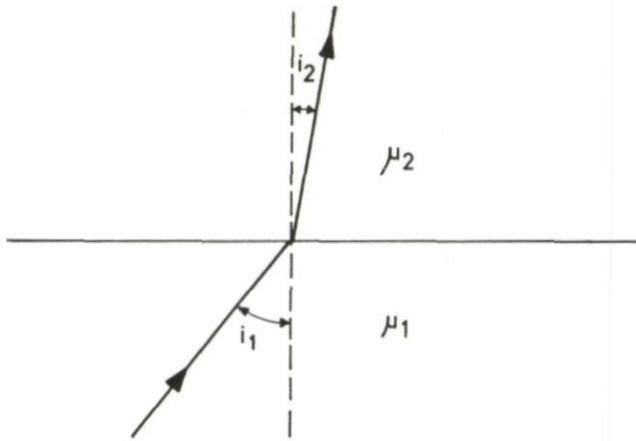


Figure 1.1 Refraction

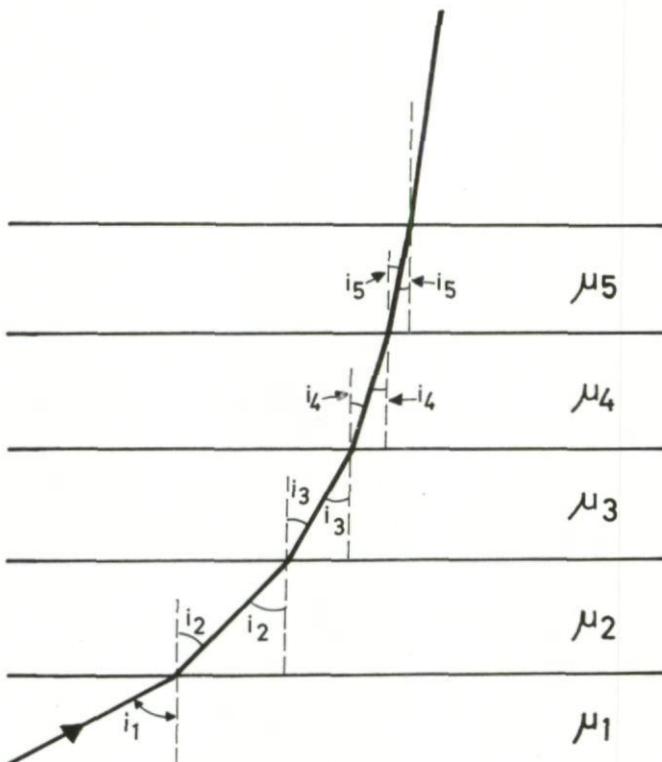


Figure 1.2 Refraction in a series of slabs

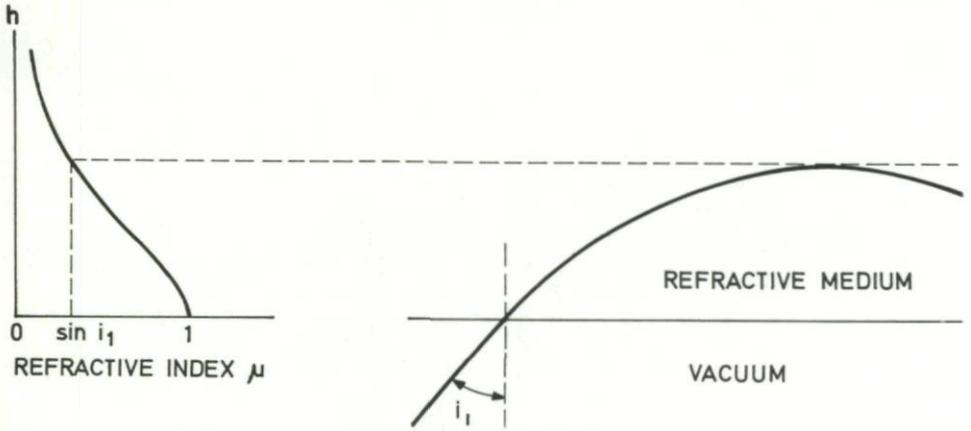


Figure 1.3 Ray path in a continuously varying medium

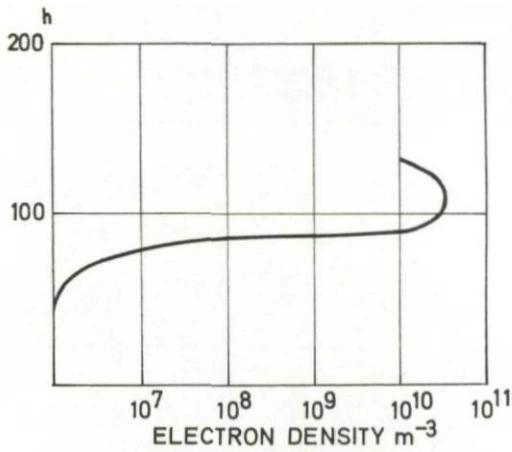


Figure 1.4 A typical E-layer electron density profile

Since  $v$ , in free space, is 300 000 000 m/s, we find, for example, that the wavelength is 300 meters if the frequency is 1 000 000 Hz (1 MHz). Light waves have very much shorter wavelengths, only of the order of 0.0005 mm. But the basic principles of geometric optics apply to all electromagnetic waves, independent of wavelength.

It can be shown that the refractive index of a medium, which contains free electrons, is normally less than unity. A ray penetrating into such a medium will therefore be bent away from the vertical towards the horizontal.

We may now study the incidence of a radio wave upon an ionized layer.

Fig. 1.5. The ray leaves the transmitter T at an angle of elevation  $\alpha$  and enters the ionosphere under the same angle of elevation, corresponding to an angle of incidence  $i_0 = 90 - \alpha$ . Below the layer the refractive index is unity. At an arbitrary point A on the trajectory, the refractive index  $\mu$  and the angle of elevation  $\theta$  is connected through Snell's law:

$$\mu \sin(90^\circ - \theta) = \mu \cos \theta = \text{constant} = \sin(90 - \alpha) = \cos \alpha \quad (\text{Eq. 1.6})$$

or

$$\mu \cos \theta = \cos \alpha \quad (\text{Eq. 1.7})$$

This again shows that as  $\mu$  decreases,  $\theta$  will also decrease. We will find it more convenient to use angles of elevation rather than angles of incidence.

Again, if  $\mu$  decreases from unity to  $\cos \alpha$ ,  $\theta$  will decrease from  $\alpha$  to zero, that is, the ray path becomes horizontal and the ray will bend over and return to earth by a symmetrical path.

If the elevation angle is  $90^\circ$ , that is, the wave leaves the transmitter vertically, the condition for reflection is that there must be enough electrons to reduce the refractive index to zero.

The frequency dependence of the refractive index

In the previous section we have noted that the refractive index in an ionized layer is normally less than unity. We will not derive the exact relationship but only qualitatively indicate why an ionized layer has this important property. When an electromagnetic wave penetrates into the layer, its electromagnetic field tries to make the electron oscillate in resonance with the wave frequency. If the electrons had been tied to the atoms, as they may be said to be in a solid, they would be oscillating in phase with the electromagnetic wave field. In a solid the refractive index is greater than unity but in the ionosphere the electrons are free and only hampered in their movement through their mass  $m$ . This makes them oscillate in resonance with the field but  $180^\circ$  out of phase. This is the basic property that makes the refractive index less than unity.

It may be shown that the refractive index  $\mu$  in an ionized medium is:

$$\mu = \sqrt{1 - \frac{Ne^2}{mf^2 4\pi^2 \epsilon_0}} \quad (\text{Eq. 1.8})$$

where

- $\epsilon_0$  - dielectric constant of free space ( $8.85 \cdot 10^{-12}$  farad/m)
- $m$  - electronic mass in kilograms ( $9 \cdot 10^{-31}$ )
- $e$  - electronic charge ( $1.6 \cdot 10^{-19}$  coulomb)
- $f$  - frequency of radio wave in Hz

By putting these values into equation (1.8) we obtain:

$$\mu = \sqrt{1 - \frac{81N}{f^2}} \quad (\text{Eq. 1.9})$$

We see, from the formula, that the refractive index depends on the electron density  $N$  and the frequency of the radio wave  $f$ . It follows that, when  $N$  increases, the refractive index is reduced. Similarly, as  $f$  increases, the refractive index increases towards unity.

It is convenient to define a parameter called the 'plasma frequency'  $f_0$  given by:

$$(f_0)^2 = 81 \cdot N \quad (\text{Eq. 1.10})$$

$f_0$  depends only on the electron density in the layer. With this definition we realize that:

$$\mu = \sqrt{1 - \frac{(f_0)^2}{f^2}} \quad (\text{Eq. 1.11})$$

which is a convenient form. Furthermore:

$$N = 0.0124 (f_0)^2 \quad (\text{Eq. 1.12})$$

If the plasma frequency is, for example,  $3 \times 10^6 \text{ Hz} = 3 \text{ MHz}$ , we find the corresponding electron density  $1.1 \times 10^{11}$  electrons per  $\text{m}^3$ .

For a given frequency, it is now possible for us to deduce the refractive index  $\mu$  as a function of height in a layer from the electron density profile. Fig. 1.6. If the maximum electron density is  $N_m$ , the corresponding plasma frequency, which in this case is called the critical frequency  $f_c$  of the layer, is given by:

$$f_c = \sqrt{81N_m} = 9\sqrt{N_m} \quad (\text{Eq. 1.13})$$

For frequencies higher than  $f_c$ , the refractive index does not become zero. For frequencies less than  $f_c$ , the refractive index becomes zero below the maximum of the layer.

From our previous discussion, it is possible to trace the path of a radio wave of frequency  $f$  through an ionized layer, when the angle of incidence  $i$  and the electron density profile are known.

In the example shown in Figure 1.7 the critical frequency of the layer is lower than the wave frequency  $f$ , but the angle of elevation  $\alpha$  is small enough to result in reflection, below the maximum of the layer, at a level where  $\mu$  is reduced to  $\cos \alpha$ .

#### Vertical incidence h'f-curves

The fact that the refractive index depends upon the frequency of the wave, means that the ionosphere is a dispersive medium. Different frequencies will propagate with different velocities. When we are interested in the velocity with which a radio impulse is transmitted in the ionosphere, we are concerned with what is known as group velocity  $v_g$  as distinct from the phase velocity  $v_p$ . We cannot enter into any detailed discussions of these two concepts but only state that in the ionosphere:

$$v_p = \frac{c}{\mu} \quad (\text{Eq. 1.14})$$

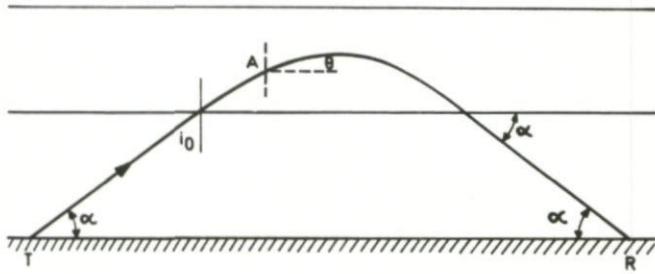


Figure 1. 5 A ray incident upon the ionosphere.

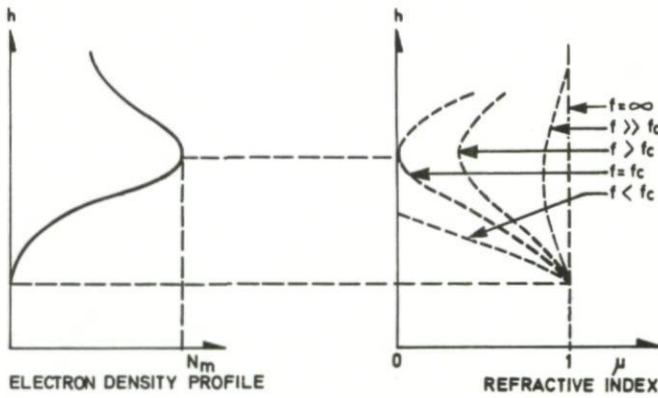


Figure 1. 6 Refractive index as a function of height in a layer

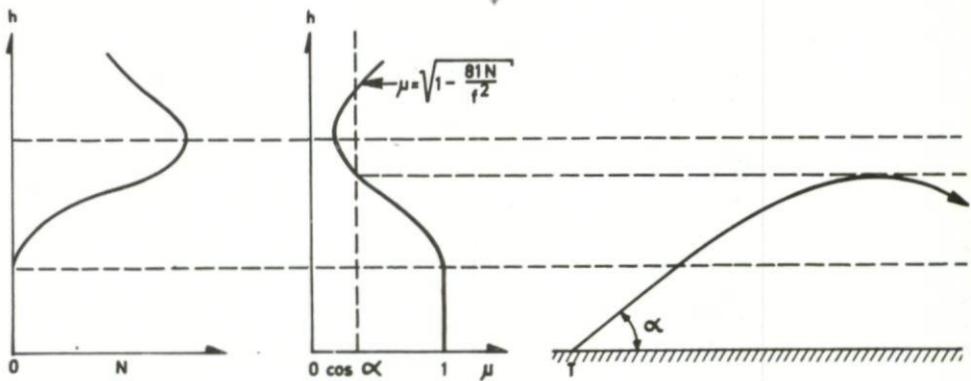


Figure 1. 7 Path of radio wave in an ionized layer

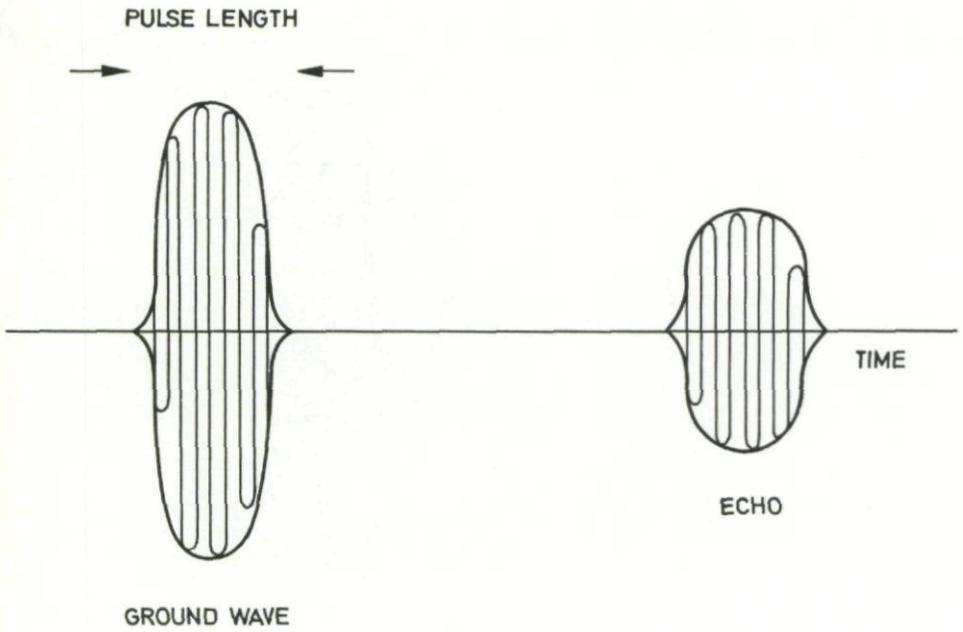


Figure 1.8 Principle of ionospheric sounding

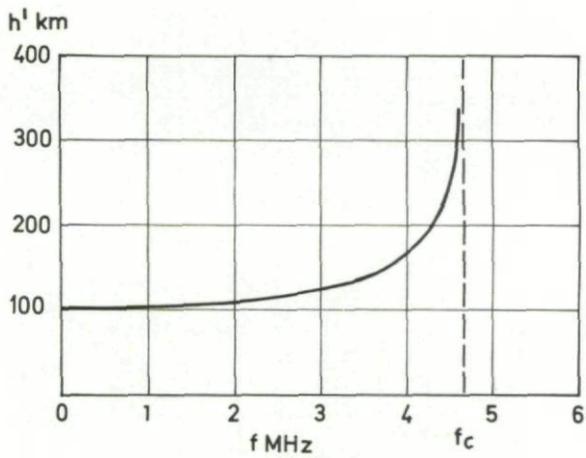


Figure 1.9 An h'f-record

and

$$v_g = c\mu \quad (\text{Eq. 1. 15})$$

These simple relationships will enable us to study what takes place when a short radio impulse is transmitted vertically on the ionosphere. From the transmitter up to the ionosphere, the wave will travel with the velocity of light. As it penetrates the ionosphere the wave packet is slowed down as it enters regions of increasing electron density. If the frequency is low enough for reflection, the group velocity at the point of reflection is reduced to zero, since  $\mu = 0$  at this point.

The basic principle of an ionospheric sounder may now be explained. We transmit a short impulse, say 100 microseconds long, of a known frequency  $f$  and record the time before it returns to a receiver located at the transmitter. Fig. 1. 8. For a higher frequency, the wave has to travel further into the ionosphere and will return later. As the frequency is increased to the critical frequency of the layer given by:

$$f_c = 9\sqrt{N_{\max}} \quad (\text{Eq. 1. 16})$$

where  $N_{\max}$  is the maximum electron density, the time delay becomes greater and for frequencies above the critical frequency the wave penetrates the layer and is reflected from a higher and denser layer or lost.

If we plot the apparent or equivalent height of reflection  $h'$  as a function of frequency given by:

$$h' = \frac{1}{2} ct, \quad (\text{Eq. 1. 17})$$

where  $t$  is the total time from transmission of an impulse to the return of the echo, we obtain the important  $h'$ - $f$ -records, Fig. 1. 9. Near the critical frequency we observe the greatest delay since the wave remains longer in a region where the group velocity is near zero. Thus, the group retardation is particularly large near the critical frequency.

An ionospheric sounder, Fig. 1. 10, is an instrument that transmits short pulses of radio waves, say fifty times per second, and records the delay photographically. The radio frequency is continuously increased. Typical characteristics of an ionosonde could be:

- Pulse power 10 kW
- Pulse repetition frequency 50 Hz
- Pulse length 100  $\mu$ s
- Frequency range 500 kHz to 30 MHz
- Sweep time 20 seconds

**Breit and Tuve's and Martyn's theorems** — Two simple theorems, which have proved useful for a description of oblique transmission of radio waves in the ionosphere, were originally presented by (a) Breit and Tuve and (b) Martyn. These two theorems describe certain relationships between the reflection of an oblique radio wave of frequency  $f_\alpha$  and a vertically incident wave of frequency  $f_{90}$ .

**Breit and Tuve's theorem** — Let us assume that a radio wave of frequency  $f_\alpha$  leaves the transmitter under an elevation angle  $\alpha$ . This ray, Fig. 1. 11, reaches the top of its trajectory at an altitude  $h$  and returns to the earth at R.

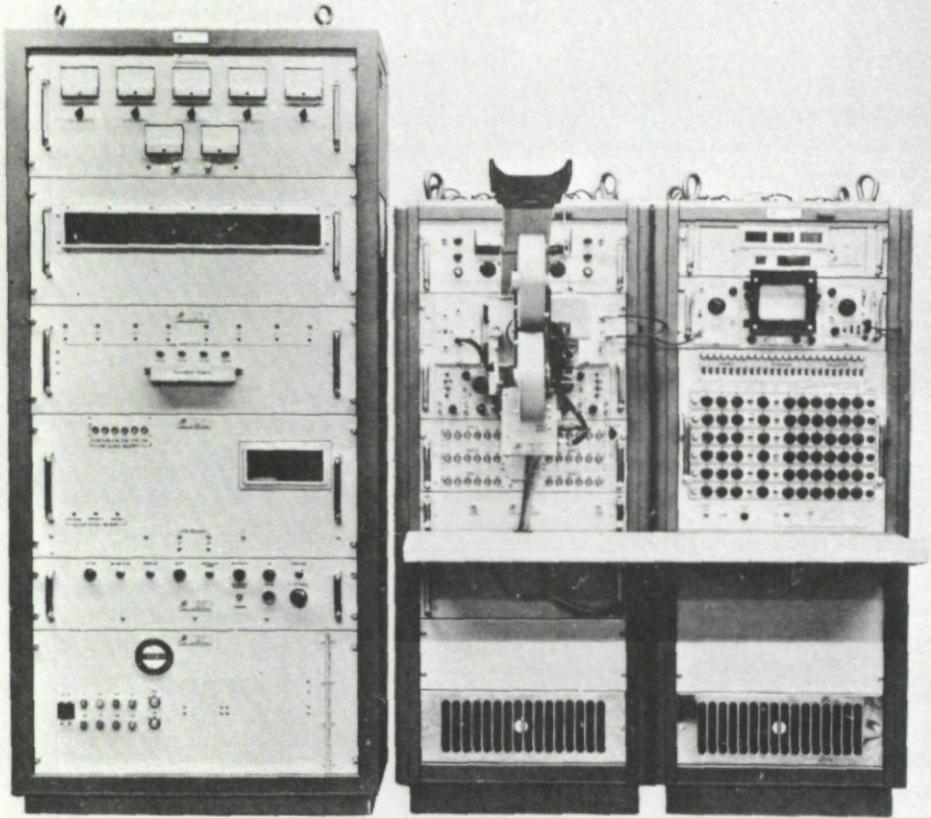


Figure 1.10 An ionospheric sounder (Model D-1, used by the Institute for Telecommunication Sciences and Aeronomy, Boulder, Colorado, U.S.A.).

The ray becomes horizontal at the point D where the refractive index  $\mu = \cos \alpha$ . The plasma frequency at this point may be deduced from equation (1.11)

$$\mu^2 = 1 - \left(\frac{f_0}{f_\alpha}\right)^2$$

$$f_0 = f_\alpha \sqrt{1 - \mu^2} = f_\alpha \sqrt{1 - \cos^2 \alpha} = f_\alpha \sin \alpha \quad (\text{Eq. 1.18})$$

A vertically incident wave is reflected at the same height  $h$ , provided its frequency  $f_{90} = f_0 = f_\alpha \sin \alpha$ .

Thus the frequency  $f_\alpha$  at oblique incidence under an elevation angle  $\alpha$  reaches the same real height as the frequency  $f_{90} = f_\alpha \sin \alpha$  at vertical incidence in the same layer. This is an important result.

If  $t$  is the time a wave takes to travel along the path TADBR, we can define an equivalent path  $P'$

$$P' = ct \quad (\text{Eq. 1.19})$$

It is not difficult to prove, the theorem known as Breit and Tuve's theorem which related  $P'$  to the horizontal range  $X$ ,

$$P' = \frac{X}{\cos \alpha} \quad (\text{Eq. 1.20})$$

This is a purely geometrical relation and does not involve either the frequency or the density in the layer.

Note: The equivalent path, Fig. 1.11, is:

$$P' = c \int_T^R \frac{ds}{v_g} = c \int_T^R \frac{ds}{\mu c} = \int_T^R \frac{ds}{\mu}$$

From Snell's law

$$\mu = \frac{\cos \alpha}{\cos \theta}$$

$$P' = \int_T^R \frac{ds \cos \theta}{\cos \alpha} = \int_T^R \frac{dx}{\cos \alpha} = \frac{1}{\cos \alpha} \int_T^R dx = \frac{x}{\cos \alpha}$$

**Martyn's theorem** — Martyn's theorem relates the behaviour of the frequency  $f_\alpha$  at oblique incidence to that of the frequency  $f_{90} = f_\alpha \sin \alpha$  at vertical incidence. It has already been shown that the two frequencies under these conditions both reach the same height  $h$ . Martyn's theorem states that the equivalent path  $P'(\alpha, f_\alpha)$  for the frequency  $f$  at the angle of elevation  $\alpha$  is related to the equivalent path at vertical incidence  $P'(90^\circ, f_{90})$  in the following way:

$$P'(\alpha, f_\alpha) \sin \alpha = P'(90^\circ, f_{90}) \quad (\text{Eq. 1.21})$$

provided  $f_{90} = f_\alpha \sin \alpha$

The proof of this theorem is:

$$P'(\alpha, f_\alpha) = c \int_T^R \frac{ds}{v_g} = \int_T^R \frac{ds}{\mu} = \int_T^R \frac{ds \sin \theta}{\mu \sin \theta} = \int_T^R \frac{dh}{\mu \sin \theta}$$

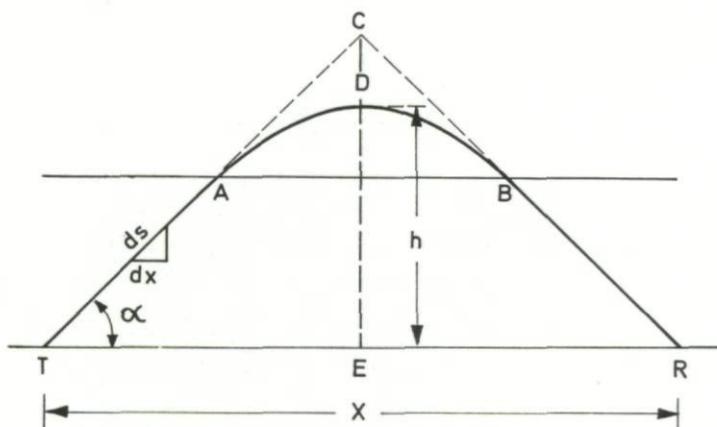


Figure 1.11 Oblique transmission

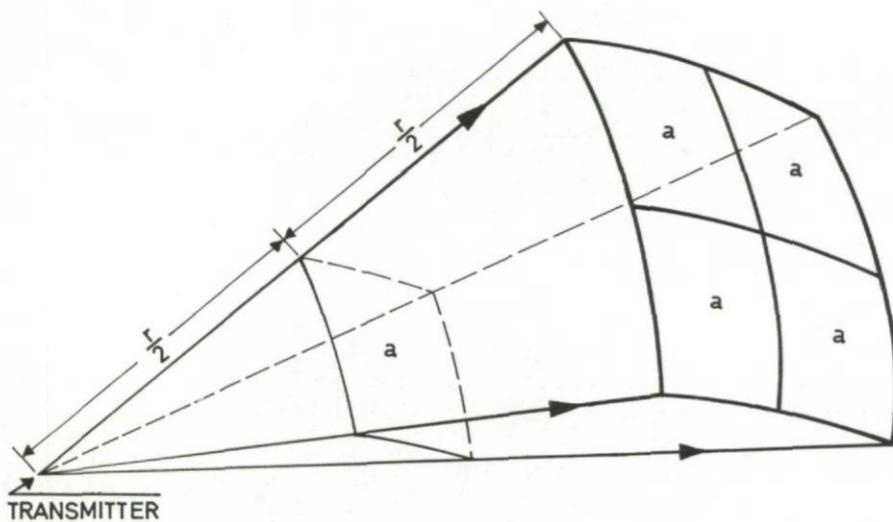


Figure 1.12 Illustrating how the energy carried by the radio wave spreads out in free space

From Snell's law, we have:

$$\mu \cos \theta = \cos \alpha$$

or

$$\mu^2 \cos^2 \theta = \cos^2 \alpha$$

We find:

$$\begin{aligned} \mu^2 \sin^2 \theta &= \mu^2 (1 - \cos^2 \theta) \\ &= \mu^2 - \cos^2 \alpha \\ &= 1 - \frac{81 N}{(f_\alpha)^2} - \cos^2 \alpha \\ &= \sin^2 \alpha - \frac{81 N}{(f_\alpha)^2} \\ &= \sin^2 \alpha \left( 1 - \frac{81 N}{(f \sin \alpha)^2} \right) \end{aligned}$$

If this is substituted above, we obtain:

$$\begin{aligned} P'_{\alpha, f_\alpha} &= \int_T^R \frac{dh}{\sin \alpha \sqrt{1 - \left( \frac{81 N}{(f \sin \alpha)^2} \right)}} \\ &= \frac{2}{\sin \alpha} \int_0^h \frac{dh}{\sqrt{1 - \frac{81 N}{(f \sin \alpha)^2}}} \end{aligned}$$

This integral is just  $P'(90^\circ, f_\alpha \sin \alpha)$  per definition since we already have seen that the frequency  $f_\alpha$  at an elevation  $\alpha$  reaches the same real height  $h$  as the frequency  $f_{90} = f_\alpha \sin \alpha$  at vertical incidence.

**Combination of the two theorems** – By combining the two theorems, we find:

$$P'(\alpha, f_\alpha) = \frac{P'(90^\circ, f_\alpha \sin \alpha)}{\sin \alpha} = \frac{X}{\cos \alpha} \quad (\text{Eq. 1. 22})$$

that is:

$$P'(90^\circ, f_\alpha \sin \alpha) = X \tan \alpha, \quad (\text{Eq. 1. 23})$$

or

$$h(90^\circ, f_\alpha \sin \alpha) = \frac{x}{2} \tan \alpha \quad (\text{Eq. 1. 24})$$

From Figure 1. 11 we find that:

$$TC = \frac{X}{2 \cos \alpha} \quad (\text{Eq. 1. 25})$$

or

$$TC + CR = \frac{X}{\cos \alpha} \quad (\text{Eq. 1. 26})$$

From Breit and Tuve's theorem, we then obtain:

$$P'(\alpha, f_{\alpha} \sin \alpha) = TC + CR \quad (\text{Eq. 1. 27})$$

Thus, we have the important result that, the time required for the wave to travel the actual path from T via D to R is equal to the time required to travel with the uniform velocity  $c$  along the triangular path  $TC + CR$ . Furthermore, Martyn's theorem tells us that, the height  $EC$  of the 'equivalent triangle' equals the apparent height of reflection of a frequency  $f_{90} = f_{\alpha} \sin \alpha$  at vertical incidence.

These relationships will be useful when we are going to relate oblique incidence transmission to the more easily observed characteristics at vertical incidence.

Absorption of radio waves — When a radio wave is emitted from an antenna, the strength of the electromagnetic field decreases with distance from the antenna. Due to geometrical spreading, the power density  $p$  will decrease inversely with the square of the distance  $r$  from the transmitter, Fig. 1. 12.

When the distance from the transmitter is doubled, the energy in the wave will spread over four times the area.

$$\text{Thus: } p = \frac{p_0}{r^2} \quad (\text{Eq. 1. 28})$$

where  $p_0$  is the power density at some reference distance, say 1 km from the antenna. Since the power is proportional to the square of the amplitude, the field strength  $E$  will be inversely proportional to the distance along the ray path, so that:

$$E = E_0 \frac{1}{r} \quad (\text{Eq. 1. 29})$$

where  $E_0$  is the electric field strength at the reference distance.

In addition to this type of attenuation, the radio wave may also suffer dissipative losses when it passes through the ionosphere. A free electron oscillating in the electric field of the radio wave will pick up energy from the wave.

If the electron is allowed to oscillate freely in the wave field, this energy will be reradiated. If, however, the electron collides with heavy particles like ions or neutral air molecules, then the electron may lose its oscillatory energy. This latter energy will not be reradiated and will therefore be removed from the wave. We will appreciate that the loss the wave suffers in this manner must depend on the electron density  $N$ , the collision frequency  $\nu$  of electrons with heavy particles and on the time the wave remains in the layer.

If the wave frequency is high compared with the plasma frequency, the refractive index  $\mu$  will be near unity and the wave will travel through the layer with a velocity near the velocity of light  $c$ . In this case the ray path will not be appreciably deviated from a straight line and the absorption suffered by the wave is called non-deviative absorption. The non-deviative absorption will decrease when the wave frequency increases. This is to be expected since, at a high frequency, an electron in the wave field may oscillate many times between collisions which will therefore, play a minor role.

If the wave frequency is near the plasma frequency, the refractive index  $\mu$  is small. The wave will then travel slowly through the medium and the ray path may be appreciably deviated. The deviative absorption the wave suffers in this case, can be very considerable and it does not depend in a simple way,

on the wave frequency. The attenuation of a wave in a dissipative medium will follow an exponential law. Including the geometrical spreading of radio waves we therefore, have:

$$E = \frac{E_0}{r} 10^{-\alpha} \quad (\text{Eq. 1. 30})$$

It may be shown that for non-deviative absorption:

$$\alpha = 0.4335 \int \kappa ds \quad (\text{Eq. 1. 31})$$

where:

$$\kappa = 5.29 \cdot 10^{-6} \frac{N\nu}{4\pi^2 f^2 + \nu^2} \text{ m}^{-1} \quad (\text{Eq. 1. 32})$$

For deviative absorption  $\alpha$  is proportional to the group delay the wave suffers in the region where  $\mu$  is small.

It is convenient to measure power, electric fields, etc. in a logarithmic measure called decibel (dB). If we have two power levels  $p_1$  and  $p_2$ , we may define the power level of  $p_1$  as being:

$$10 \log \frac{p_1}{p_2} \text{ dB} \quad (\text{Eq. 1. 33})$$

over the level  $p_2$ . If  $p_1 = 100$  watt and  $p_2 = 1$  watt,  $p_1$  is 20 dB over  $p_2$ .

Since power is proportional to voltages and field strengths squared, the appropriate formula for voltage and field strength is:

$$10 \log \frac{E_1^2}{E_2^2} (\text{dB}) = 20 \log \frac{E_1}{E_2} (\text{dB}) \quad (\text{Eq. 1. 34})$$

If  $E_1$  is 1000 volt and  $E_2$  is 1 volt,  $E_1$  is 60 dB over the level of  $E_2$ .

The level of the field from a transmitter relative to the reference field  $E_0$ , will be, equation 1. 30:

$$\frac{E}{E_0} = \frac{10^{-\alpha}}{r}, \quad (\text{Eq. 1. 35})$$

or in logarithmic measure:

$$20 \log \frac{E}{E_0} = -20 \alpha - 20 \log r = -L - 20 \log r \quad (\text{Eq. 1. 36})$$

Here  $L$  is  $8.67 \int \kappa ds$ , equation 1. 31.

In this case  $E$  is a given number of decibels less than  $E_0$ .

$20 \alpha + 20 \log r$  is the total attenuation under the reference level  $E_0$  at the reference distance which is the unit in which  $r$  is measured.

The influence of the earth's magnetic field on radio wave propagation — The magnetic field of the earth has a profound influence on the propagation of electromagnetic waves, as it makes the ionosphere doubly refracting. A radio wave entering the ionosphere is split up into two rays, the ordinary and the extraordinary ray. The ordinary ray is least affected by the magnetic field.

The ordinary and extraordinary components of a radio wave will be reflected

from different heights in an ionospheric layer and will have different group delays. Therefore, the ionosonde will normally receive two echoes at each frequency and the ionogram will show both ordinary and extraordinary traces. The two traces will look similar, but have a frequency separation of about 0.7 MHz, Fig. 1.13. The critical frequency of the ordinary wave is not influenced by the presence of the earth's magnetic field, and is given by:

$$f_c = 9\sqrt{N_{\max}} \quad (\text{Eq. 1.37})$$

as in the no-field case.

In many cases the effect of the earth's magnetic field can be taken into account in the simple formulas for absorption and refraction by replacing the wave frequency  $f$  with an effective frequency  $f \pm f_L$  where the plus sign refers to the ordinary wave and the minus sign to the extraordinary wave.  $f_L = f_H \cos \theta$  where  $\theta$  is the angle between the direction of propagation and the earth's magnetic field. The gyro frequency  $f_H$  is associated with a natural resonance of the free electrons in the geomagnetic field.  $f_H$  varies, from about 0.8 MHz at the geomagnetic equator to about 1.6 MHz over the geomagnetic poles, appendix 1.

For example, equation 1.32 for the absorption index becomes:

$$\kappa = 5.29 \cdot 10^{-6} \frac{N\nu}{4\pi^2 (f \pm f_L)^2 + \nu^2} \text{ m}^{-1}$$

when the magnetic field is included. We note that the extraordinary wave will have a larger absorption index than the ordinary wave.

### Variations of the ionospheric layers

As shown in the previous section, the reflecting properties of an ionospheric layer depend on the distribution of free electrons in the layer. For convenience, an ionospheric layer is described in terms of three parameters: (i) the height  $h$  of the base of the layer above the ground, (ii) the maximum density of free electrons  $N_{\max}$  in the layer, and (iii) the 'thickness'  $d$  of the layer.

In order to predict the conditions for radio communication over a certain distance on a given day, it is important to know how these three parameters vary over the day, over the year and over longer periods. As the two former  $h$  and  $N_{\max}$  are the most important for radio communication, only these will be discussed in the present section.

For a fuller understanding of the time variations of the ionospheric layers we have to discuss how the different layers are produced and how they disappear.

How free electrons are produced and how they are lost

In a gas which is completely in 'darkness', which means not exposed to any kind of external radiation, most of the electrons are bound in low-energy orbits in the gas atoms. If now, for some reason, a bound electron gains sufficient energy, it may leave its natural orbit in the atom and (i) move to a more energetic orbit in the atom or (ii) it may leave the atom. In the first case the atom is excited, whereas in the latter it is ionized. In this book, only the ionization processes are of interest.

The electron may gain sufficient energy to leave the atom in various ways:

- (a) By absorbing electro-magnetic radiation of a proper wavelength. The sun is a source of electro-magnetic radiation of energies sufficiently high to ionize the atoms in the upper atmosphere. Of particular importance are soft x-rays and ultraviolet light. Most of the regular ionization in the earth's upper atmosphere may be ascribed to solar radiation of these wavelengths.
- (b) Fast, electrically charged particles from the sun or from the galaxies may collide with bound-electrons and transfer sufficient energy to the electrons. This type of ionization is of particular importance for some of the irregular ionospheric phenomena, namely in the lower ionosphere at high latitudes.

Free electrons may be lost in two different ways: (i) by falling back into an atomic shell which is not filled up with bound-electrons, recombinative loss of free electrons, or (ii) the electron may attach to a neutral atom or molecule thus forming a negative ion, attachment loss of free electrons.

At low altitudes, below 80 km, most of the free electrons are lost by attachment, whereas the recombination loss types dominate at higher altitudes.

How the different ionospheric layers are sustained

When the ultraviolet radiation or the x-rays from the sun penetrate the earth's atmosphere, they are gradually absorbed and the absorbed energy is partly used in the production of free electrons. The different parts of the solar spectrum are not absorbed with the same ease in the earth's atmosphere. Hence, the penetration depths into the atmosphere depend on the wavelength of the radiation. Thus, the ionization at various altitudes in the atmosphere is caused by different parts of the solar spectrum.

The shape of the ionospheric layers depends on (i) the spectrum of the solar radiation, and (ii) the air density in the upper atmosphere. The average atmospheric density and temperature are shown in Figure 1.14. The solar spectrum consists of a 'background' radiation with certain strong 'lines' superimposed. Some of these lines are of particular importance for the formation of the ionized layers. The number of ion pairs produced in a gas, when exposed to radiation, is directly proportional to the number of ionizable atoms in the gas. Hence, above the altitude where the solar radiation of a particular wavelength is completely absorbed, the ionization rate falls off with altitude in the same way as the air density.

An ionospheric layer produced by a particular frequency of the solar spectrum will therefore have a fairly sharp lower boundary, below which no significant radiation can penetrate. Above the peak of the layer the electron density decreases slowly in accordance with the altitude decrease of the air density. This rather crude picture, which essentially, is identical to the theoretical 'Chapman-layer', is in good agreement with the electron density profile in the upper regions of the ionosphere.

The altitude distribution of the ionization rate depends critically on the solar zenith angle, Fig. 1.15. When the sun is directly overhead, the solar radiation may penetrate deeper into the atmosphere than when the sun is a few degrees above the horizon. Furthermore, for decreasing elevation, the solar energy is spread out over a larger horizontal area of the atmosphere and the total number of ions produced in a vertical column decreases. We should, therefore, expect that the maximum electron density peaks near noon and that, at the same time, the height of the layer is at a minimum.

The three regular layers of ionization in the upper atmosphere are the D-, E- and F-layers. These are produced by electromagnetic radiation from the sun. The D-layer is in the height range from about 70-90 km, the E-layer is situ-

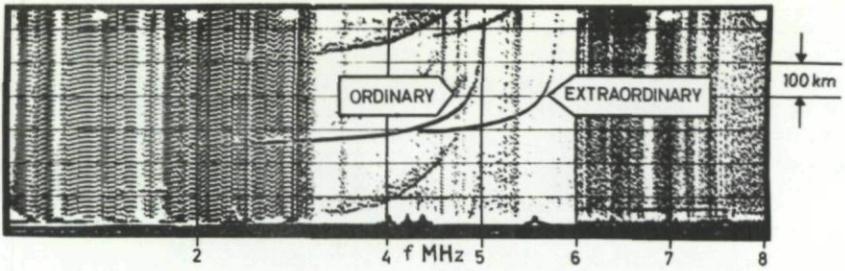


Figure 1.13 Ionosonde recording showing ordinary and extraordinary traces (Courtesy Institute for Telecommunication Sciences and Aeronomy, Boulder, Colorado, U.S.A.).

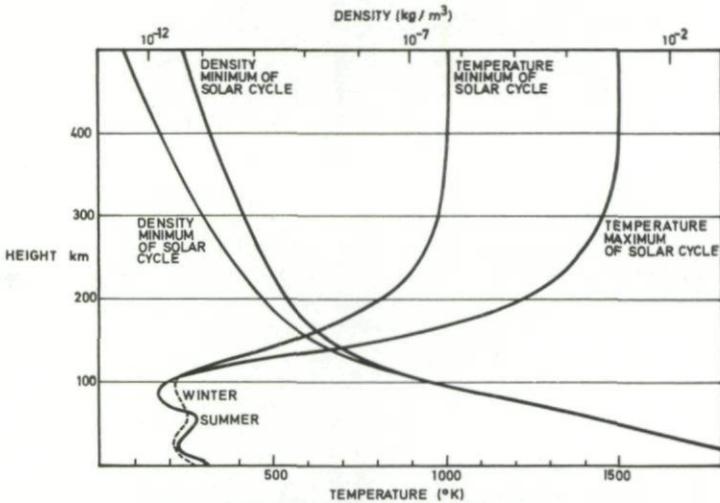


Figure 1.14 Average temperature and air density distribution in the upper atmosphere near sun spot maximum and sun spot minimum (Courtesy F.S. Johnson, Satellite Environment Handbook, Stanford University Press, 1961)

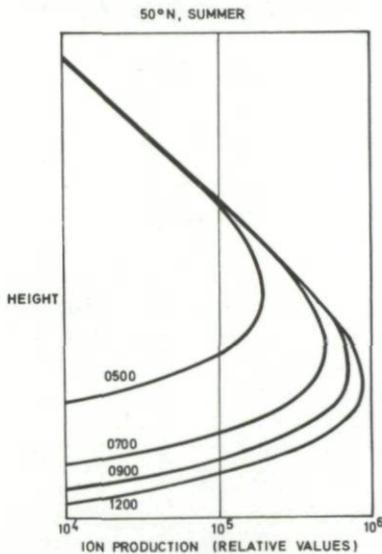


Figure 1.15 Diurnal variation in the ionization rate of an ionospheric layer (mode)

ated around 100 km and the F-layer around 250 km. This last layer is often split into two layers, the F1- and the F2-layers.

According to recent theories the F-layer is produced by ultra-violet radiation, the E-layer by soft x-rays, whereas the D-layer is produced by x-rays and ultraviolet radiation.

The 'shape' of the regular ionospheric layers is controlled by two processes: (i) different wavelengths of the solar spectrum are absorbed at different heights in the atmosphere, therefore, the shape of the layer of ionization depends on the shape of the solar spectrum, and (ii) the loss rate of free electrons in the atmosphere depends on the altitude. In the upper part of the ionosphere the lifetime of a free electron is many minutes or hours, whereas in the lower D-region a free electron is lost within a much shorter time period. Hence, in order to maintain a layer of ionization below 100 km, it is necessary to have a very high production rate. These two processes have to be taken into account when we want to explain the height and time variations in the ionospheric layers.

#### Methods for observing the regular ionospheric layers

The main tool for observations of the regular F- and E-layer ionization in the ionosphere is the ionosonde, which was briefly described on page 6 for the case when only one layer was present in the ionosphere. If two or more ionized layers exist in the upper atmosphere, the ionosonde record becomes more complex. A typical ionogram obtained during normal conditions at a middle latitude station is shown in Figure 1.16. The altitude difference between two adjacent height marks is 50 km, and the apparent bases of the E- and F-layers are situated at 105 and 240 km, respectively.

The duplication of the traces may be ascribed to the doubly refractive properties of the ionosphere. The frequency difference between the two 'sets' of traces ( $\approx 0.7$  MHz) agrees well with theory. The critical frequencies, ordinary components, of the two layers are 3.3 and 8 MHz, respectively. By a complicated procedure the corresponding electron density profile may be derived from the ionogram, Fig. 1.16. This Figure shows that the ionization increases fairly steadily with altitude, up to the F2-layer peak at 300 km.

In contrast to the regular traces shown in Figure 1.16, an ionogram obtained during disturbed conditions in the auroral zone is shown in Figure 1.17. This ionogram is not easy to interpret. The reduction of an ionogram to an electron density profile normally requires an electronic computer. A good indication of the state of the ionosphere can, however, be obtained by reading some simple parameters off the records. In a normal routine analysis the following parameters are taken from the ionograms, Fig. 1.16.

- (a) The critical frequencies  $f_oE$ ,  $f_oF1$ ,  $f_oF2$  of the ionospheric layers which directly give the maximum electron density in each layer
- (b) The minimum virtual heights  $h'E$ ,  $h'F1$  and  $h'F2$
- (c)  $f_{min}$ , the lowest frequency at which a trace can be seen on the record
- (d) The factor  $M(3000)F2$ , see page 18

The ground based ionosonde has a serious limitation in that it cannot give information about the whole height range of interest in the ionosphere. Firstly, it cannot observe the ionosphere above the maximum of the F2-layer and secondly, it gives little information about the important region below 100 km.

A number of methods have been devised to study the parts of the ionosphere where the conventional ionosonde cannot be used.

With regard to the 'topside' ionosphere the most successful experiment in

recent years has been the launching of a satellite-borne ionosonde. This instrument sounds the ionosphere vertically from above and telemeters the ionograms down to earth. An example of a topside ionogram and the electron density profile derived from it is shown in Figure 1.18. The region between 50 and 100 km is important because it strongly absorbs high frequency radio waves. The absorbing properties of this part of the ionosphere will be further discussed on page 20

The parameter  $f_{min}$  gives an indication of the intensity of the absorbing layer, but special techniques must be used to find the structure of the D-region. Such techniques are for example, field strength measurements of signals reflected from the E- and F-layers, or studies of the properties of very low or low frequency waves reflected from the D-region.

In recent years rockets and satellites have made it possible to carry instruments through the ionosphere and the *in situ* measurements thus obtained have proved very valuable for the understanding of the physics of the ionized regions of the atmosphere.

#### Time variations in the ionospheric layers

The regular ionospheric layers D, E and F2 undergo three types of regular time variations:

- (a) Diurnal variations. The density of free electrons in the regular ionospheric layers shows a clear local time control. Near noon the electron density has a maximum and a minimum is observed at night.
- (b) Seasonal variations. The day-time electron density in the regular E- and D-layers is significantly higher during the summer than during the winter, whereas the opposite is true for the F2-layer.
- (c) Sun spot cycle. The regular layers undergo an eleven year periodical variation in phase with the cycle of solar activity. Near sun-spot maximum, the electron density in the regular layers is significantly higher than the density near minimum activity.

The diurnal variations in the electron density of the regular layers can grossly be ascribed to variations in the solar elevation. A similar explanation can be offered for the seasonal variations in the E- and D-layers. The seasonal variations of the F2-layer is, however, somewhat more complicated and the variations of the solar elevation alone cannot account for the behaviour of this layer. The eleven year periodicity in the regular layers is due to long term variations in the ionizing radiation from the sun. In addition to the time variations in the regular layers, there is a variation in the electron density with geographical latitude and longitude. The longitudinal variations are mainly due to the difference in local times, and the general decline of electron density in the regular layers towards the geographical poles is ascribed to the changes in solar elevation with latitude. Near the equator, the F2-layer behaves somewhat irregularly, as the ionization here is lower than at middle latitudes. This phenomenon is known as the 'equatorial anomaly'.

The different types of regular variations in the ionosphere will be discussed at some length below.

**Time variations in the F-region** — The most important ionospheric layer for radio wave communication is the F-layer. Historically, this region is separated into an F1-layer and an F2-layer. The F1-layer is normally an insignificant 'hump' in the F-region electron density profile, it is probably of minor importance for radio wave communication via the F-region. The F2-layer is in many respects the most irregular of the regular ionospheric layers. There is no simple law between the time variations in the peak electron density of

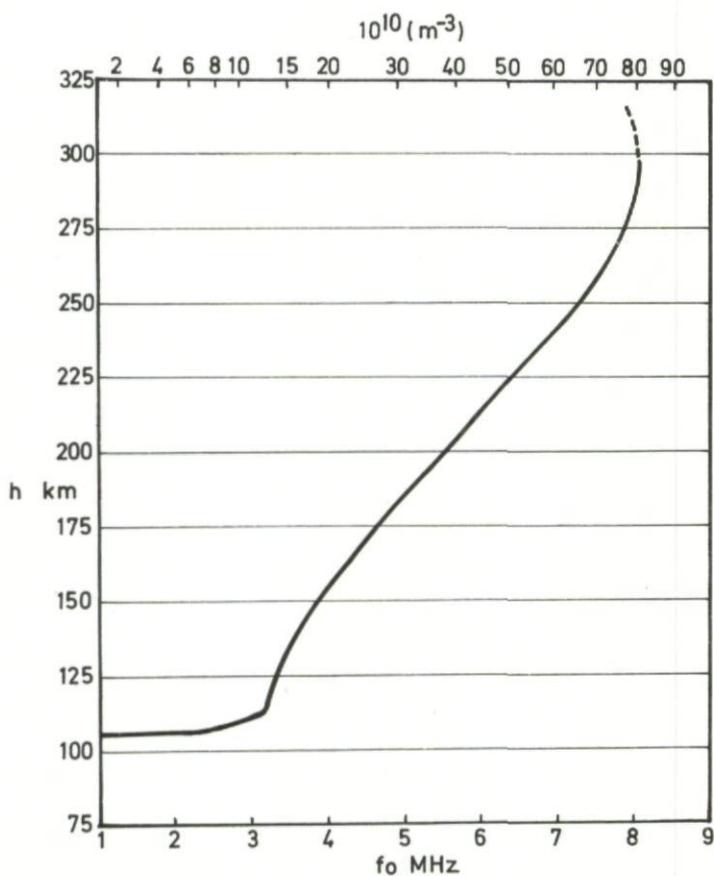
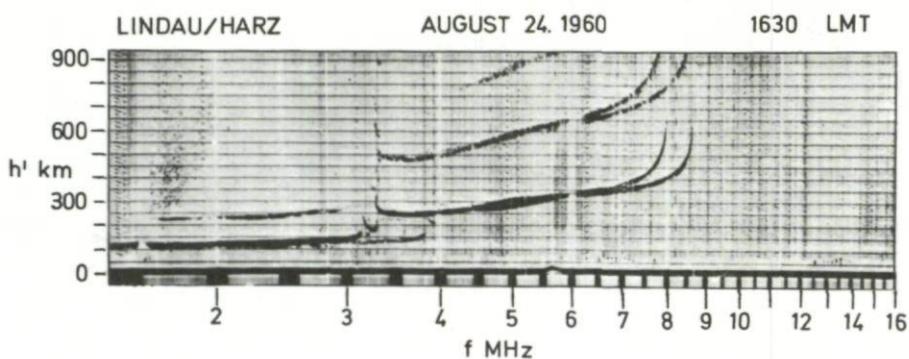


Figure 1.16 Quiet day ionogram and corresponding electron density profile  
 (Courtesy W. Becker, Max-Planck Institut für Ionosphären-  
 Physik, Lindau, West-Germany)

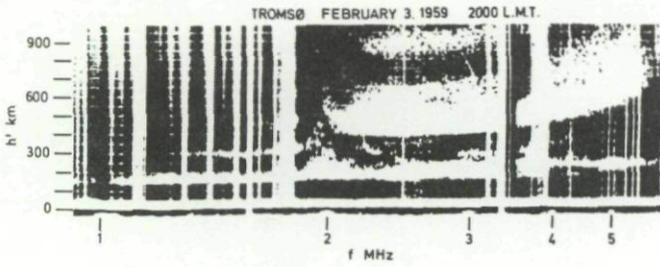


Figure 1. 17 Ionogram obtained during disturbed conditions in the auroral zone

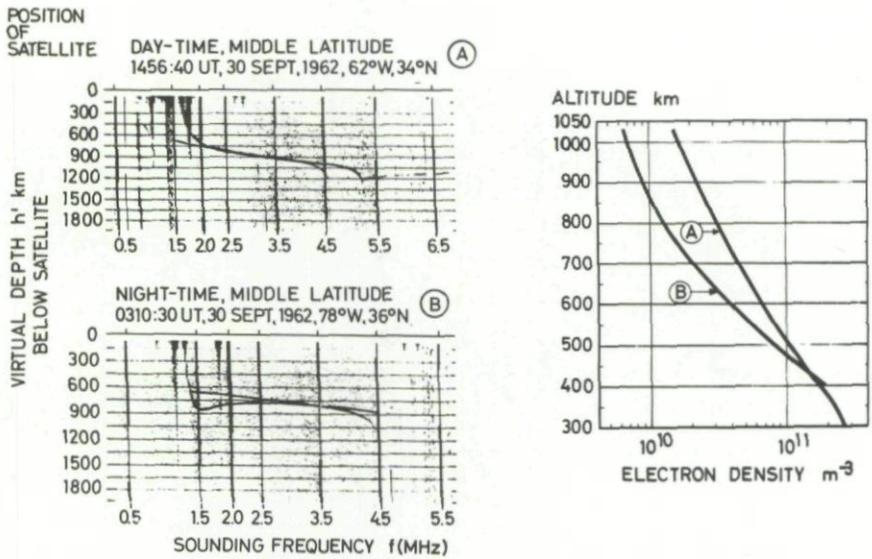


Figure 1. 18 Topside ionogram and corresponding electron density profile  
(After J. E. Jackson, Electron density distribution, North-Holland Publishing Company, 1964)

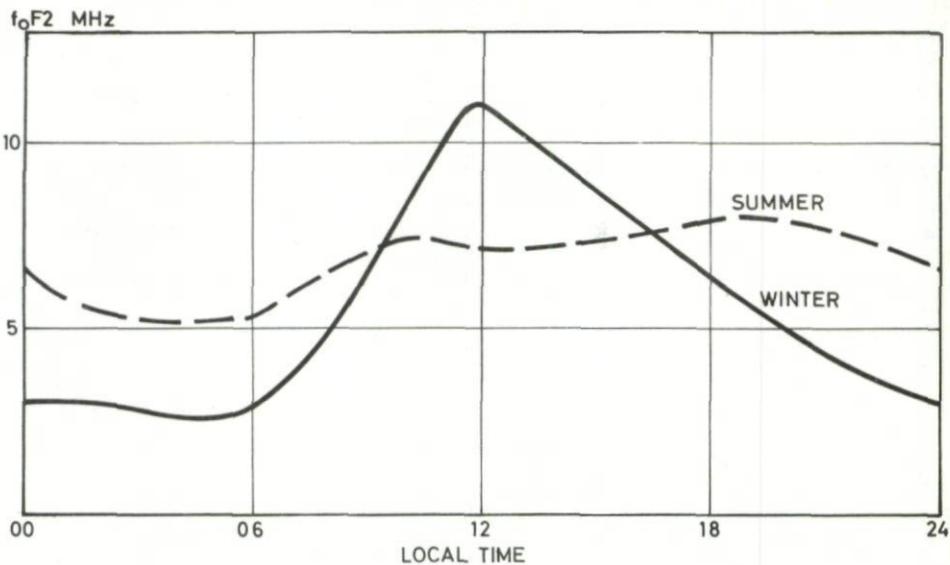


Figure 1.19 Typical diurnal variation of  $f_oF2$  at middle latitudes

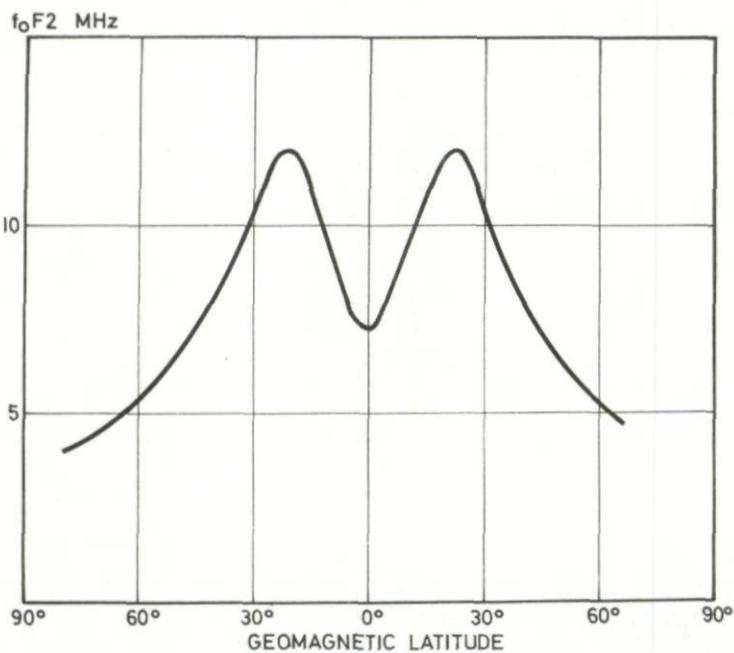


Figure 1.20 Average noon values of  $f_oF2$  (equinoctial conditions) at various geomagnetic latitudes

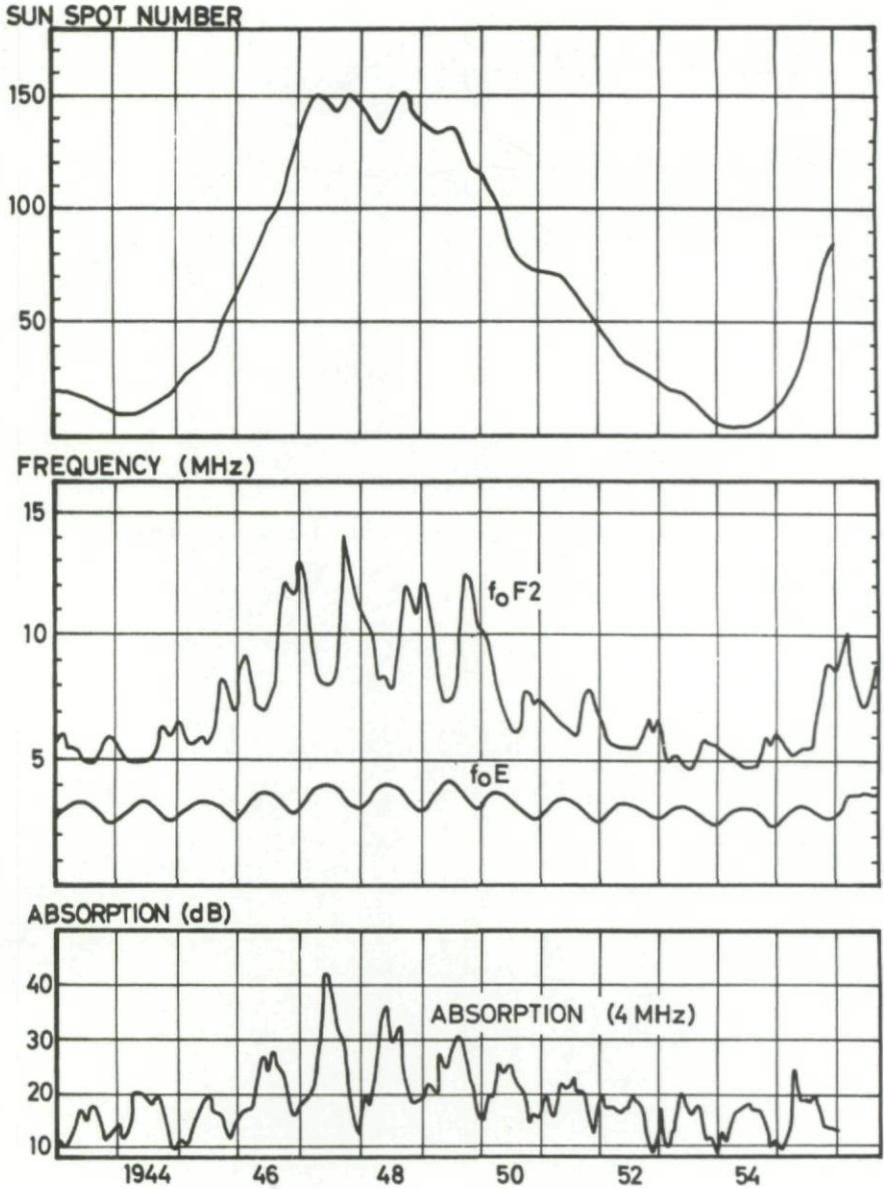


Figure 1.21 Long term variations in R and in the average noon values of  $f_oF2$ ,  $f_oE$  and D-layer absorption at middle latitudes  
 (Courtesy W.R. Piggott, Radio and Space Research Station, Slough, Bucks, England)

the layer and the variations in the solar elevation.

Typical diurnal variations of the critical frequency of the layer (foF2) during summer and winter conditions at middle geographical latitudes are shown in Figure 1. 19. The individual observations for each day are indicated by black dots. The day-to-day variation of foF2 is of the same order of magnitude as the diurnal variation. The average diurnal variations are indicated by the continuous lines. Average values of foF2, vary from 2.5 MHz to 11 MHz during the winter and from 5 MHz to 7 MHz during the summer. The winter curve shows a variation similar to the diurnal variation in the solar elevation, although a slight difference in the time of maximum is found. During summer-time, however, foF2 is almost constant over the day. In spite of the fact that the sun is much higher above the horizon during the summer-time than in the winter, the noon value of foF2 in the winter is twice as high as the summer noon value. This has been explained as being due to a seasonal variation in the upper atmosphere temperature. During the summer the upper atmosphere is warmer than during the winter. Hence the atmosphere, and the ionosphere, is expanded, which causes a lowering of the number of free electrons per unit volume at these altitudes. At higher latitudes the F2-layer ionization is significantly lower than indicated in Figure 1. 19. Generally speaking, the critical frequency decreases about 1 MHz per 10° latitude interval between 20° and 70° latitude near the equinoxes. The corresponding summer and winter values are somewhat smaller and larger, respectively. Near the equator the maximum electron density in the F2-layer is lower than at slightly higher latitudes. The 'equatorial anomaly' is illustrated in Figure 1. 20. The long term variations in the F-2-layer ionization is dominated by the eleven year periodicity in the ionizing radiation from the sun. A convenient parameter for solar activity is the Wolf's sun spot number, R. The time variation of this number for the period 1930-1965 is shown in Figure 1. 21. The monthly averages of the mid-day foF2 show a similar variation over the same period, as shown in the same illustration. Near maximum of solar activity the winter noon values may exceed 15 MHz at a middle latitude station, whereas the corresponding frequency near sun spot minimum is only about 6 MHz. The ionization in the F2-layer during summer-time is less sensitive to variations in the solar radiation. Irregular variations in the solar radiation may strongly affect the ionization in the F2-region. During 'geomagnetic storms', associated with irregular solar activity, the electron density in the F2-layer may temporarily drop by a factor of two or more. In particular, at high geographical latitudes, where the largest ionospheric disturbances occur, many irregular variations in the F-region may be observed, some of which will be discussed in a later section. In addition to the general, short-lived declines in the F-layer ionization during disturbed periods, the F-region is often somewhat 'rugged' at high latitudes, Fig. 1. 17. The radio signals reflected from this layer often consist of a large number of rapidly fading components, which are not well suited for regular HF radio wave communication. This type of reflection is similar to the twinkling of light from a corrugated ocean surface.

**Time variations in the E-region**—The most regular ionospheric layer, which is the E-layer, is situated in the altitude range 100-150 km. The vertical extent of this layer is not well defined. There is no sharp transition between the E-layer and the D-layer below and most observations indicate that the ionization above the peak of the E-layer shows no clear minimum but increases with height into the F-region, Fig. 1. 16.

The time variations in the E-region are regular over the day and over the year. In fact, all the variations can be attributed to the time variations in the solar elevation and to the solar radiation by the following formula:

$$foE = 0.9 [(180 + 1.44R) \cos \chi]^n \text{ MHz} \quad (\text{Eq. 1. 38})$$

where  $R$  is the Wolf's sun spot number, and  $\chi$  is the angle between the zenith at a certain point on the earth and the sun-earth line. This angle is often called the solar zenith angle. The exponent  $n$  varies slightly with geographical latitude from 0.1 at high latitudes to 0.4 near the equator. An average value of  $n$ , which is about 0.25, gives values of  $f_oE$  in fair agreement with observations for all latitudes. In fact, the theoretical value of  $f_oE$  deduced from this equation and the observed value, generally agree within 0.2 MHz.

The noon, equinoctial value of  $f_oE$  varies from about 4.5 MHz to about 3.5 MHz near the equator from sun-spot maximum to sun-spot minimum. At higher geographical latitudes the long term variations in the  $f_oE$  values are somewhat smaller. Corresponding variations in the critical frequency of the E-layer observed at middle latitudes may be represented by maximum and minimum values of 4.0 and 3.2 MHz, respectively, Fig. 1.21.

In addition to the time variations in the critical frequency of the E-layer, the radio communication via this layer is also affected by time variations in the height of the ionization. In agreement with theory, Fig. 1.15, the ionization starts at high altitudes just after sunrise. As the solar elevation increases, the E-layer ionization gradually descends and the minimum altitude of the layer is reached near local noon. This lowering of the E-layer ionization, together with the maximum in the  $f_oE$  near local noon, implies that the high frequency radio communication is most often controlled by the E-layer ionization near mid-day, as will be shown later.

**Time variations in the D-region** — The regular E- and F2-layers were described in terms of the time variations in the peak electron density and the height of the layers. The time variations in the regular ionospheric D-region will be treated from a different point of view. High frequency radio waves cannot be reflected from the D-region, but they penetrate the layer and may, at times, be heavily attenuated. Hence, the time variation in the integrated absorption through the layer is the important property for radio wave communication.

The attenuation of a radio wave in an ionospheric layer depends on the number of free electrons per unit volume and the number of collisions each free electron makes per second with the neutral gas molecules. The collision frequency, which mainly depends on the gas pressure, is not known to vary significantly. Most of the regular variations in the D-layer absorption can therefore be ascribed to changes in the electron density of the layer. At medium and low latitudes during summer-time the D-layer absorption undergoes a fairly regular diurnal variation in phase with the E-layer ionization. Near mid-day there is a clear maximum in the absorption and the layer is practically absent during night-time.

This diurnal variation can be expressed as:

$$\text{D-layer absorption} = K \cos^n \chi \text{ (dB)} \quad (\text{Eq. 1.39})$$

The exponent  $n$  at medium and low latitudes is in the range 0.8-1.0, and the parameter  $K$  depends on the solar activity. A typical variation in the absorption over the day at middle geographical latitudes during the summer is shown in Figure 1.22. The diurnal variations during a winter day differ significantly from the variations observed in the summer. During winter-time the absorption varies considerably from one day to the other and the absolute value of the absorption is generally higher than the corresponding summer values.

In Figure 1.23, the observed absorption is plotted versus the solar zenith angle  $\chi$ . The figure clearly demonstrates that the absorption  $L$  in dB, varies linearly with  $\cos \chi$  during the summer and equinox months but that the absorption in winter can be appreciably higher than expected from the simple linear relationship, equation 1.39. The abnormally high absorption in winter is not a regular

$$D = L(f + f_L)^2$$

60°N, SUMMER 1956

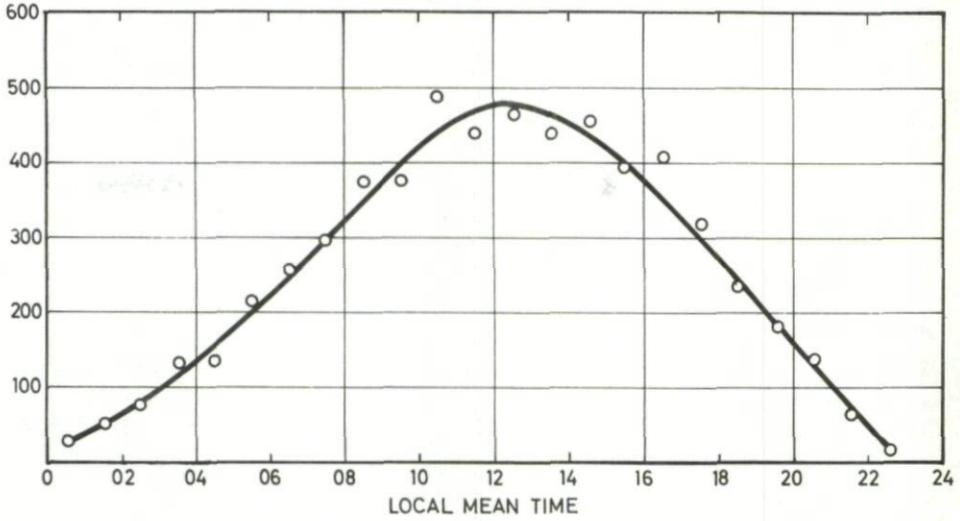


Figure 1.22 Typical diurnal variation of the D-layer absorption at middle latitudes during summer time

$$L(f + f_L)^2 \text{ dB(MHz)}^2$$

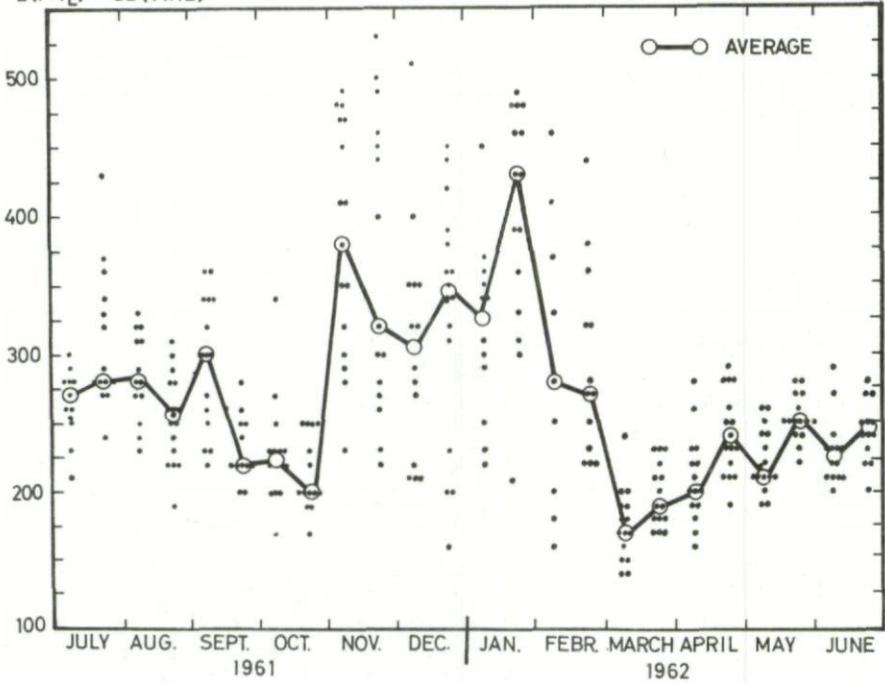


Figure 1.23 Seasonal variation of the D-layer absorption at middle latitudes (noon values)

feature but occurs on groups of days. No satisfactory physical explanation of this phenomenon has yet been given. The winter anomaly is only present at middle latitudes, but it can be observed as far south as Southern Europe.

The D-layer absorption at high geographical latitudes is much less sensitive to variations in the solar elevation than the absorption nearer to the equator. Typical values of the exponent  $n$  for the regular D-layer at high latitudes range from 0.0 to 0.2. As the diurnal variations in the solar elevation at high latitudes is small and the absorption does not depend critically on the solar elevation, one should expect that the regular variation of the absorption over the day is small. This is in agreement with observations. The diurnal variation of the absorption at high latitudes is, in fact, controlled by irregular variations in the radiation from the sun rather than the regular time variation in the solar elevation. The irregular occurrences of high absorption at high latitudes, which are often associated with visual auroral events, will be discussed later.

The time variation of the absorption of a radio wave, 4 MHz reflected from the ionosphere during one sun spot cycle is given as an example in Figure 1. 21. This illustration refers to middle latitude conditions. The average absorption changes by 10-25 dB from sun spot minimum to sun spot maximum.

#### Irregular time variations in the E- and D-regions

In addition to the regular time variations in the ionosphere, most of the layers undergo various irregular variations. The majority of these variations will be discussed in some detail in later sections. For completeness some of these phenomena are listed below.

**Irregular D-layer ionization** — (a) SID (Sudden Ionospheric Disturbance) — Short-lived increases in the D-region ionization, due to enhanced solar x-ray radiation, may cause break down in high frequency radio wave communication simultaneously over the entire sun-lit part of the earth.

(b) Auroral Absorption Events — During periods when visual aurora is present at high geographical latitudes, the D-region ionization at these latitudes is often significantly enhanced. This enhancement is caused by the fast, electrically charged particles from space which also cause visual aurora.

(c) Polar Cap Absorption — Sporadically there is a significant increase in the D-layer ionization over the polar caps. This enhancement is caused by positively charged particles which are expelled from the sun during periods of strong solar activity.

**Irregular E-layer ionization** — Irregular E-layer ionization, named the sporadic E-layer, is observed at all geographical latitudes. The behaviour of the layer differs at different latitudes. At times, this layer is extremely thin (2 - 3 km).

The earth is generally separated into three 'zones' of Es occurrence:

(a) The Equatorial Zone — in which the Es-layer is observed regularly every day during the sunlight hours. The layer is partly transparent for radio signals reflected from higher ionospheric layers and the critical frequency of the layer shows a clear diurnal variation with a peak value as high as 13 MHz near mid-day. Because of the regular behaviour of this layer it is not quite logical to refer to it as a sporadic layer.

(b) The Middle Latitude Zone — Between geographical latitudes 10° and

60° the sporadic E phenomenon is mainly observed during summer time, and the maximum ionization in the layer is found in the sunlight hours. The critical frequency of the layer over North America is higher than 3 MHz in more than 60% of all hours and in about 10% of all hours the  $fEs$  exceeds 7 MHz. At other geographical longitudes the middle latitude Es-layer is less developed.

- (c) The Auroral Zone — At geographical latitudes near the visual auroral zone, the sporadic E phenomenon is exclusively observed during the night hours. Normally the critically frequency of the layer is lower than 4-6 MHz but values of  $fEs$  as high as 10 MHz have occasionally been observed. The sporadic E-layer observed at high geographical latitudes, differs from the corresponding type of ionization at middle and low latitudes in many respects. Frequently, the auroral zone Es is completely opaque for high frequency radio waves. The importance of this layer for radio wave communication in polar regions will be discussed in Chapter 3.

### Maximum useable frequency

#### Skip distance

Breit and Tuve's and Martyn's theorems can assist us in understanding the basic properties of oblique transmission of radio waves. Let us assume that we have an ionospheric layer at a height  $h$ , and that the critical frequency of the layer is  $f_c = 5$  MHz. At vertical incidence, frequencies below 5 MHz are reflected back to the earth, while frequencies above 5 MHz penetrate the layer and are lost. Assume we have a transmitter, transmitting at a frequency of say  $f = 10$  MHz. This frequency would penetrate the layer at vertical incidence and be reflected when the angle of elevation,  $\alpha$ , is reduced from  $90^\circ$  to some critical value,  $\alpha_c$ , for which  $f \sin \alpha_c = f_c$ , that is

$$\sin \alpha_c = \frac{f_c}{f} = \frac{1}{2}.$$

In our case, we obtain  $\alpha_c = 30^\circ$  which then means that the frequency 10 MHz will not be reflected at a higher angle of elevation than  $30^\circ$ , Fig. 1. 24. At this critical angle, the height of the triangular path is infinite since the equivalent vertical incidence frequency  $f \sin \alpha_c$  equals the critical frequency  $f_c$  of the layer. This frequency is infinitely delayed at vertical incidence or, in other words, the equivalent height is infinite. Therefore, the triangular path of the frequency  $f$  at elevation angle  $\alpha_c$  is also infinite. As the angle of elevation is further reduced, case 2, the equivalent vertical incidence frequency is reduced below  $f_c$  of the layer and the height of the triangular path is reduced so that it equals the apparent reflection height of the equivalent frequency  $f \sin \alpha$  at vertical incidence. The range of the wave is similarly reduced. As the elevation is still further reduced, case 3, the range again becomes smaller since the equivalent frequency is now considerably below the critical frequency.

As the elevation is reduced, a compensating effect also appears. The finite height of the layer causes the range of the wave to increase again. Reducing the elevation from  $\alpha_c$  case 1, we note therefore, that the oblique range of the wave moves in from infinite to a minimum distance,  $d_s$ , and so moves out again. No point within the skip zone can be reached in this case ( $f > f_c$ ) and at the skip distance itself, we have a focusing effect with a stronger signal. The corresponding angle of elevation  $\alpha_s$  is called the skip angle. Points outside the skip distance may be reached by a high angle ray, which, due to group delay and deviative absorption, is weak, and a stronger low angle ray. The two rays take different times to reach a receiver and this may cause disturbances and fading. At the skip distance we may say that the two rays become one, since

they are equally delayed and in phase. It is this property that causes the strong build up of the signal at the skip distance.

### Transmission curves

Martyn's theorem relates the properties of the frequency  $f$  at an angle of elevation  $\alpha$  to the properties of the frequency  $f \sin \alpha$  at vertical incidence. Normally, we do not know the angle  $\alpha$  but only the range  $x$  over which we want to communicate and we will have to determine  $\alpha$  from knowledge of the reflecting layer and of the distance  $x$ . Let us assume that we know  $\alpha$ , then, we also know the equivalent frequency  $f \sin \alpha$  at vertical incidence, and from the observed  $h'f$ -curve we can find the value of  $h'$  for the frequency  $f \sin \alpha$ . We have seen that it is a direct relation between  $h'$  and  $\alpha$  for a given range  $x$ , namely:

$$h' = \frac{x}{2} \tan \alpha$$

This is a general relationship independent of the shape of the  $h'f$ -curve. We can draw a curve of  $h'$  as a function of  $\alpha$  for different values of  $x$ . We will, in accordance with a technique described by Newburn Smith, G. Millington and others, make a plot of  $h'$  not as a function of  $\alpha$  but of  $\frac{1}{\sin \alpha}$  and furthermore, use a logarithmic scale for the abscissa values of  $\frac{1}{\sin \alpha}$ . Such a curve is known as a transmission curve for the distance  $x$  and is shown in Figure 1.25 a.

We may also plot the  $h'f$ -curve on a logarithmic scale of frequency of the same size as the logarithmic scale for  $\frac{1}{\sin \alpha}$  on the transmission curve, and with the same scale for  $h'$ . This is shown in Figure 1.25 b. We can place the two curves on top of each other in the way shown in Figure 1.25 c. The points at which the curves intersect will give us the values of  $h'$  we need. By the property of the logarithmic scales for  $\sin \alpha$  and  $f$ , the ordinate for  $\sin \alpha = 1$  ( $\alpha = 90^\circ$ ) will be on top of the ordinate for the frequency  $f \sin \alpha / \sin \alpha$ , that is, the frequency  $f$  we are using at oblique incidence.

By reversing the process we can now solve our problem graphically in the following elegant way. We place the ordinate for  $\sin \alpha = 1$  on the transmission curve, over the ordinate for the frequency we want to use at oblique incidence over the distance  $x$ , for which the transmission curve has been made. The intersection of the transmission curve with the  $h'f$ -curve gives at once the  $h'$  relating to our problem and normally two sets of corresponding values of  $h'$  and  $\alpha$ , that is a high angle and a low angle mode. As we increase the frequency, the two modes will converge and when the transmission curve is tangent to the  $h'f$ -curve we have the highest possible frequency for the distance in question. This frequency is the Maximum Useable Frequency (MUF) for the distance. This distance again is the skip distance for the frequency in question. In Figure 1.26 the limiting case is shown. From an observed  $h'f$ -curve and with transmission curves for different distances, we may find the MUF(d) and the skip angle  $\alpha_s(d)$ , both as indicated as a function of distance.

### Determination of factor M

From the preceding section, we have succeeded in determining the MUF for a given distance from the  $h'f$ -curve. The MUF is really an equivalent frequency  $f \sin \alpha$  divided by  $\sin \alpha$ , that is  $f$ . But since we are recording and tabulating the critical frequency  $f_c$  of a layer, it is common to try to write:

$$\text{MUF}(d) = M(d) \cdot f_c \quad (\text{Eq. 1.40})$$

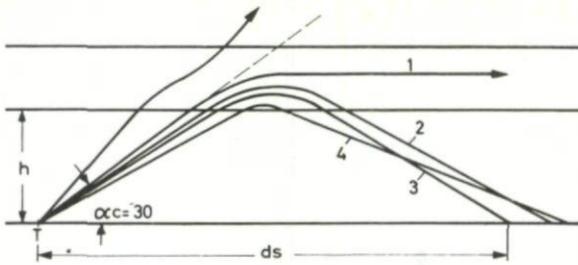
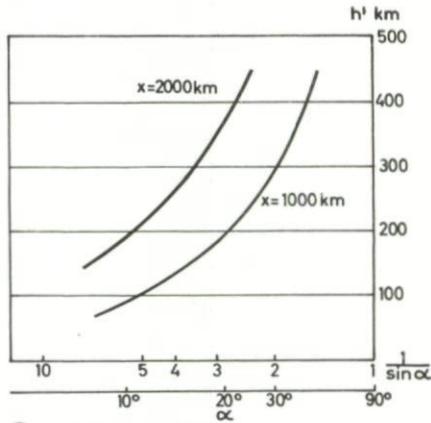
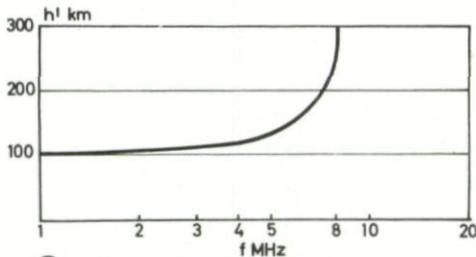


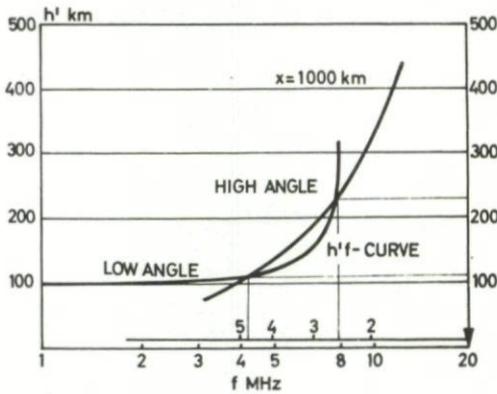
Figure 1.24 Skip distance



(a) THE SKIP SLIDER



(b) h'f ON LOGARITHMIC SCALE



(c) USE OF SKIP SLIDER

Figure 1.25 The skip slider

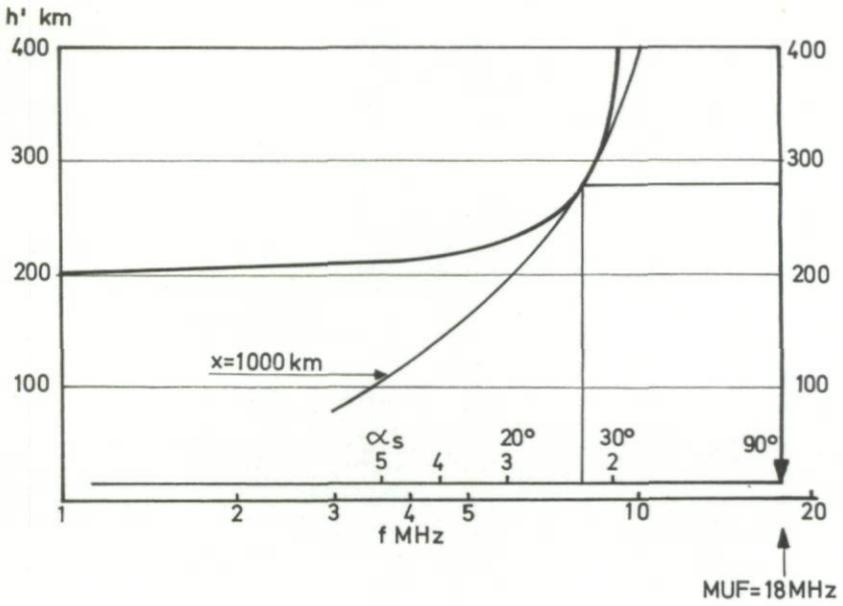


Figure 1.26 Determination of MUF

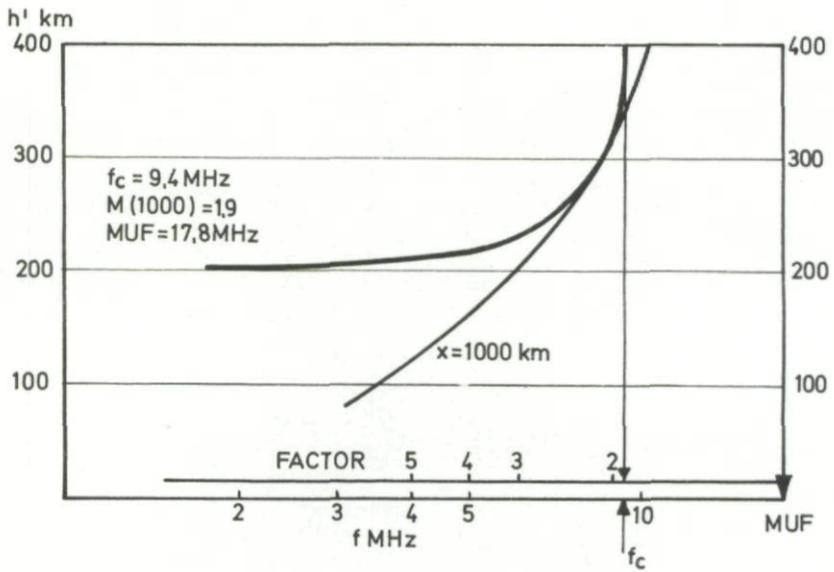


Figure 1.27 Reading off factor  $M$

From the skip slider, which due to the logarithmic scales is really a slide-rule, the factor  $M$  for a given distance may be directly read off as indicated in Figure 1. 27. We observe that the factor will depend on distance and in order to be able to find the MUF for all distances, we ought to read off factors for the distances in question. Experience shows that for the F-layer it is enough to read off  $M(3000)$ , that is the factor for 3000 km. Using this value, factors for other distances from 0 to 4000 km may be read off an empirical conversion chart, Fig. 1. 28. If the Factor  $M(3000)$  is, say 3. 4 we find that the factor of 1500 km is 2. 3.

For the regular layers E and F1, which may be assumed to be thin, the factor  $M(d)$  may with sufficient accuracy be regarded as being independent of  $f_c$  and given as  $1/\sin \alpha$  where  $\alpha$  is given from the geometry of the problem. The E layer is here assumed to be at a constant height of 120 km and of zero thickness.

Referring to page 15, we know that the geomagnetic field influences propagation of waves at vertical incidence. Such is also the case at oblique incidence but we may apply the skip-slider as before on the ordinary wave. The MUF is, as at vertical incidence, somewhat increased when the magnetic field is included and may be written:

$$\text{MUF}(d) = f_o \cdot M(d) + \text{correction} \quad (\text{Eq. 1. 41})$$

Here  $f_o$  is the critical frequency for the ordinary magnetoionic wave. For the extraordinary component the suffix  $x$  will be used. The correction is approximately 0. 7 MHz at vertical incidence and is actually somewhat reduced at greater distances.

In a later chapter we will in more detail give the practical techniques by which it is possible from prediction of  $f_o$  for the layers in question and  $M(3000)$  for the F-layer to determine the MUF for the distance and the time of interest. Principally, the procedure ought to be clear from this chapter, but a great number of graphs and nomograms have been constructed to ease the numerical work.

#### Effect of the earth's curvature

The techniques developed in the preceding sections, including the principles of the skip slider, are based on our assumption of a flat earth. Fortunately the picture derived, with the concept of skip distance, skip angle, high and low angle ray etc, still applies for a curved earth, only details are modified. The earth's curvature will not enter our considerations for propagation at vertical incidence and to points at distances less than 300 km. As transmission becomes more oblique, the modification becomes more important. We do remember that with a flat earth, the angle  $\alpha$  at the transmitter equals the angle of elevation at the layer. From Figure 1. 29 it is clear that with curved earth this is not so. When  $\alpha$  at the transmitter is zero, the angle  $\gamma$  at the layer is not zero but is approximately given by:

$$\sin \gamma = \sqrt{\frac{2H}{r_o}}$$

where  $r_o$  is the radius of the earth. The frequency corresponding to the limiting ray will not be infinite, as with flat earth, but given as:

$$f_c \sqrt{\frac{r_o}{2H}} \quad (\text{Eq. 1. 42})$$

This factor  $\sqrt{\frac{r_0}{2H}}$

is given in Figure 1.30. For an Es-layer, for example with critical frequency 7 MHz, we find, assuming  $h = 100$  km that the factor is 5.6 and the highest possible frequency is therefore, in this case, approximately 40 MHz. It may be shown that the skip slider can be modified for curved earth. This is not an exact modification, but an approximation. With this modification, we might use the techniques which we have developed for determining MUF(d) as described.

One feature of oblique incidence which is introduced by the curvature of the earth, is that we have a practical limit to the distance to which a wave may be transmitted by reflection from a given layer by a single 'hop'. Practical limits for the F-layer is 4000 km, for the F1-layer 3000 km and for the E-layer 2000 km. Furthermore it is worth noticing that at great distances outside the skip distance, the low angle ray may be screened off while the high angle ray may be the only ray present but this ray is normally very weak and of little practical importance.

### Absorption, radio noise, and lowest useful high frequency

We have, in the preceding section, dealt in a general way with the problem of finding the maximum useable frequency. It is, of course, important for the operator of a high frequency radio station to know the range of frequencies in which he can work and we will now turn to the problem of whether there is a lowest useful high frequency and, if so, how this lower frequency limit can be determined.

The quality of reception is determined by the ratio of the field strength of the wanted signal to the noise level present at and in the receiver. That a lowest useful high frequency must exist can be understood from Figure 1.31 where typical frequency variations of received field strength and of radio noise are sketched. In the high frequency range the strength of the received signal will normally increase with the working frequency up to the MUF, while the intensity of radio noise will decrease. Therefore, at a certain frequency the received field strength will be equal to the minimum field strength necessary for the receiver to detect the transmitted information. This frequency is called the Lowest Useful High Frequency (LUHF). It will, normally, be determined by the point A at which the two curves in Figure 1.31 cross.

At frequencies lower than the LUHF the signal will drown in the noise. The letter H is included in LUHF to indicate that transmission over the same path might still be possible using frequencies in the VLF or LF bands. This possibility is illustrated in Figure 1.31 where the received field strength exceeds the noise at the lowest frequencies. The problem of finding the LUHF for a given transmission path and for a given time, therefore falls naturally into two parts, (a) to estimate the received field strength as a function of frequency and (b) to determine the radio noise level at the receiving site in the frequency range of interest.

In this section we will discuss the principles which form the basis for the determination of the LUHF, while the actual technique used in preparing the predictions will be treated in Chapter 2.

The power radiated from a transmitter

In order to find the amplitude of the received signal we first need to know how much power the transmitter radiates in the direction of importance for our particular transmission mode. In other words, we must find the field strength

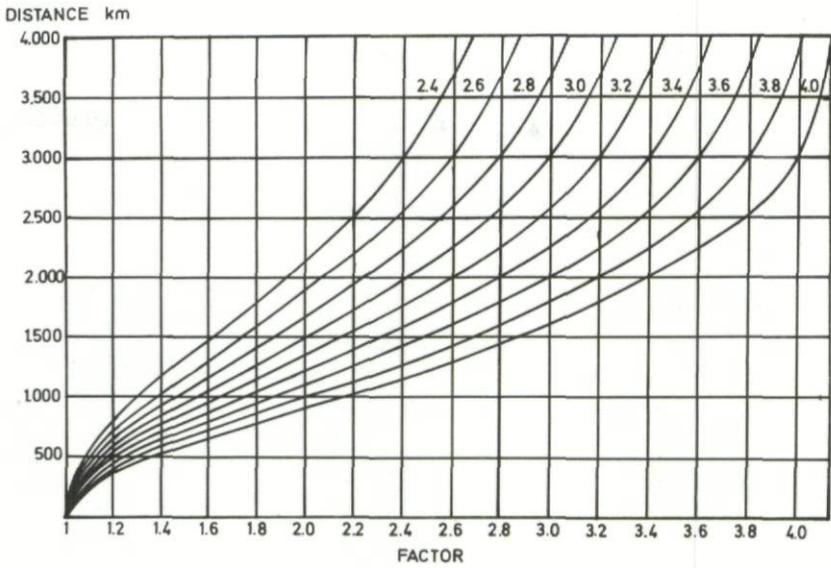


Figure 1.28 Factor conversion curves

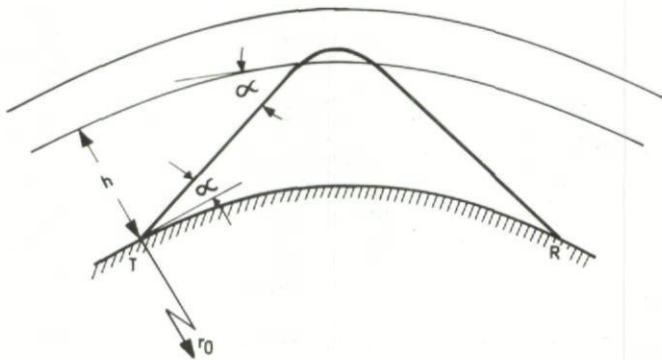


Figure 1.29 Curved earth

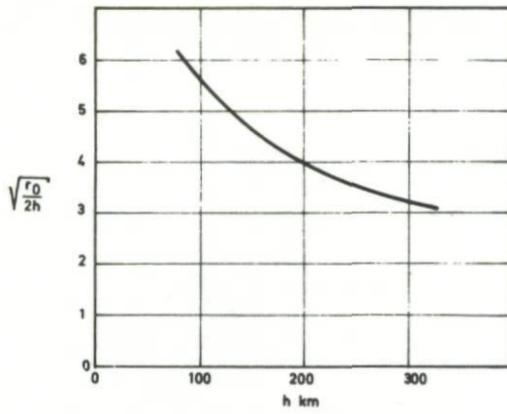


Figure 1.30 Limiting factor

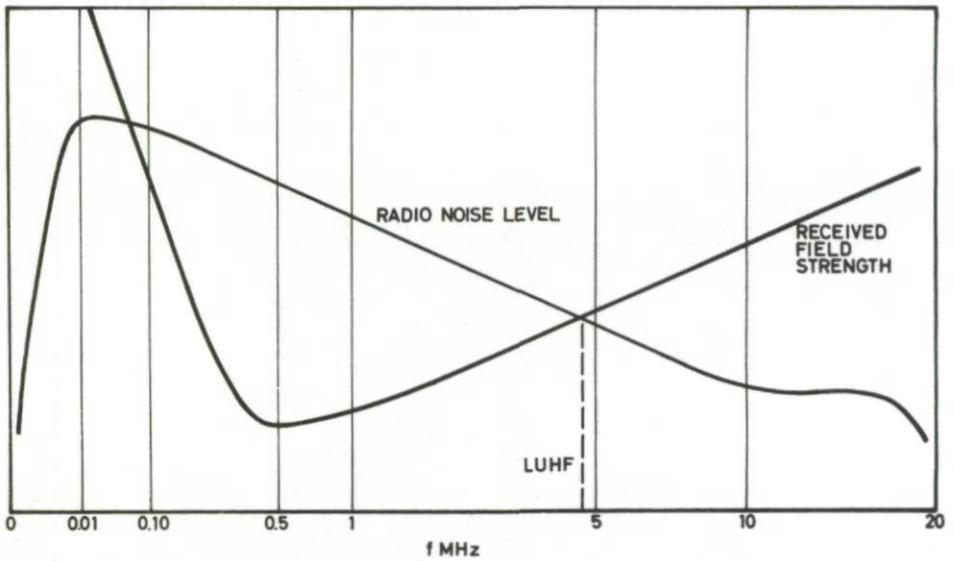


Figure 1.31 The frequency variation of received field strength and of radio noise

at some reference distance, say 1 km, from the transmitter along this direction.

A transmitting antenna can be built to concentrate the radiated power along a particular direction, and it is convenient to give the resulting field strength in terms of a reference. A useful reference antenna is a short vertical wire, one end of which is on the ground. If this antenna radiates a total power of 1 kW, the field strength at a distance of 1 km along the ground from the antenna will be 0.3 V/m. Thus, if our actual antenna radiates a total power of 1 kW, the gain in a particular direction can be defined as:

$$G_{\theta} = 20 \log \frac{E_{\theta}}{0.3} \text{ dB}$$

where  $E_{\theta}$  is the field strength in V/m at a distance of 1 km along this direction. If our antenna radiates  $P$  kW, the field strength at a distance of 1 km will be:

$$10 \log P + G_{\theta} \text{ dB over } 0.3 \text{ V/m}$$

or

$$10 \log P + G_{\theta} + 109.5 \text{ dB over } 1 \mu\text{V/m} \quad (\text{Eq. 1. 43})$$

The gain factor  $G_{\theta}$  for radiation in a given direction must either be calculated from the geometry of the antenna or measured.

The free space attenuation

A high frequency radio wave will not be significantly influenced by the neutral atmosphere and below the ionosphere the wave will therefore behave as if propagated in free space.

On page 8, it was shown that under such conditions the field strength  $E$  of the wave will decrease with distance  $r$  along the ray path from the transmitting antenna according to the law:

$$E = E_0 \frac{1}{r},$$

where  $E_0$  is the field strength at some reference distance.

If we assume that the ionosphere is a plane, perfect mirror, it is therefore easy to calculate the field strength received at a given distance along the ground from the transmitter.

Assuming that the field strength at a distance of 1 km from the transmitter is  $E_0$  and that the height of the reflecting layer is  $h$ , it follows from Figure 1. 32 that the field strength at a distance  $d$  along the ground from the transmitter will be:

$$E = E_0 \frac{1}{\sqrt{d^2 + 4h^2}}$$

When  $d \gg h$ , this becomes:

$$E = E_0 \frac{1}{d} \quad (\text{Eq. 1. 44})$$

The attenuation of the signal in dB will be:

$$A_s = 20 \log \frac{E_0}{E} = 10 \log (d^2 + 4h^2) \quad (\text{Eq. 1. 45})$$

If the curvature of the earth and ionosphere is taken into account, the calculation of the free space attenuation is somewhat more complicated. At small angles of elevation the ionosphere can then act as a focusing mirror as illustrated in Fig. 1. 33. This effect is, however, not important in practice because most transmitters radiate little power at low angles of elevation.

The free space attenuation will, of course, depend on the height of the reflecting layer and on the mode of propagation. To give an idea of the orders of magnitude involved some typical values are listed below.

**Table 1. 1**

**Some values of free space attenuation**

Height of reflecting layer	Mode of propagation	Distance from transmitter	Attenuation in dB
110 km	1 hop	300 km	51. 4
110 km	2 hop	1000 km	115
300 km	1 hop	300 km	55. 3
300 km	2 hop	1000 km	57. 8
200 km	3 hop	2000 km	174

The ground losses

For more than one hop propagation the loss at the ground reflection may be important. This loss depends on the angle of incidence, the wave frequency and the type of ground or sea in the area where the wave is reflected. The loss for sea reflection is small, of the order of a fraction of a decibel except at vertical incidence, while the loss when the wave is reflected from dry ground may be considerable, normally between 1 and 10 dB. In the following, the ground loss will be denoted  $A_g$ . Figure 1. 34 a and b shows typical values of  $A_g$  as a function of frequency and of angle of incidence.

Ionospheric absorption

The next problem is to estimate the magnitude of the losses suffered by a radio wave in the ionosphere and how these losses depend on the wave frequency. On page 8 it was described how a radio wave is absorbed when it passes through an ionized medium. It is convenient to discuss the loss in terms of an absorption index, that is the absorption in decibels per unit length along the ray path. The total loss, see equations 1. 31 and 1. 36, can be expressed as:

$$L = 8. 67 \int_s \kappa ds \text{ (dB)} \quad (\text{Eq. 1. 46})$$

where  $ds$  is an element of path, and the integral indicates a summation of the losses along the ray path. The absorption index  $\kappa$ , depends in a complicated way on the wave frequency  $f$ , the density of free electrons  $N$ , and the frequency  $\nu$  of electron collisions with neutral molecules in the air. It also depends on the earth's magnetic field. A simplified expression was given in equation 1. 32.

An accurate calculation of the total ionospheric absorption for a particular circuit, presents a formidable problem, since we need to know the exact path of the wave through the ionosphere as well as the values of  $N$ ,  $\nu$  and the earth's magnetic field at each point along this path. Fortunately, it is possible by combining theory and practical experience, to find certain simple rules which can

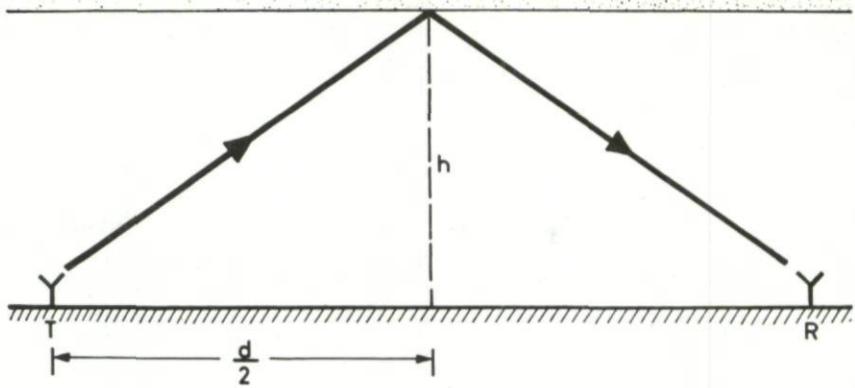


Figure 1.32 Reflection from a plane, perfectly mirroring ionosphere

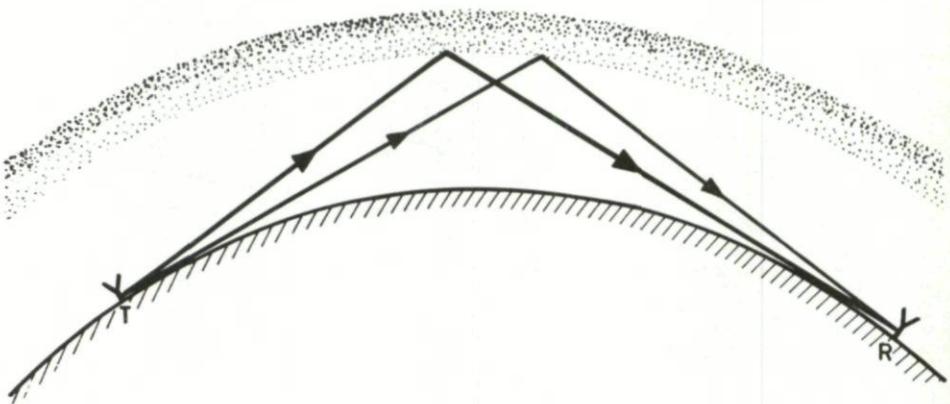


Figure 1.33 The focussing effect of a curved ionosphere

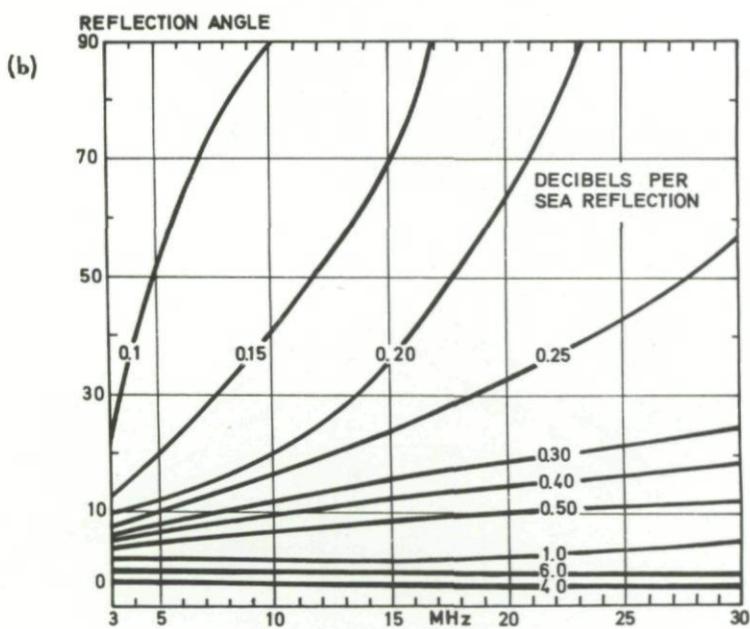
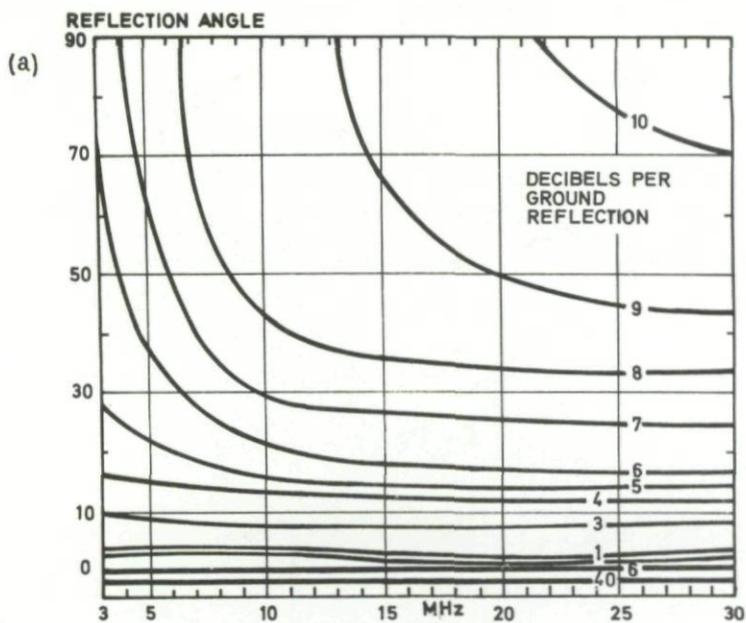


Figure 1.34 (a) Ground losses for reflection from the sea  
 (b) Ground losses for reflection from dry land.

be used to make fairly good estimates of the total ionospheric absorption. The most important of these rules will be discussed in this section. Some of the experimental techniques used to study ionospheric absorption are described in Chapter 3. We will now proceed to discuss the variations of ionospheric loss with frequency, angle of incidence, solar zenith distance and sun spot cycle.

**The frequency dependence of ionospheric absorption — Deviative and non-deviative absorption —** On page 8 a somewhat arbitrary, but useful distinction was made between the non-deviative and deviative absorption, that is between the absorption when the refractive index  $\mu$  is near unity, and the absorption when  $\mu$  is very much less than unity. We now want to ask which of these forms of absorption is likely to be most important in practice, and in what part of the ionosphere most of the absorption occurs.

The absorption index for non-deviative absorption is given by equation 1.32. This equation becomes

$$\kappa = \text{const} \frac{N\nu}{f^2} \quad (\text{Eq. 1.47})$$

when the collision frequency  $\nu$  is much smaller than the frequency  $f$  of the wave. This condition is normally fulfilled in the ionosphere for high frequency waves.

The total non-deviative absorption can be found by summing up all the contributions to the absorption along the parts of the ray path where:

$\mu \approx 1$ , so that:

$$L_{\text{non dev}} = \text{const} \frac{1}{f^2} \int_s N\nu \, ds \quad (\text{Eq. 1.48})$$

Thus the total non-deviative absorption varies as  $\frac{1}{f^2}$  and is proportional to  $\int_s N\nu \, ds$ .

Figure 1.35 shows typical variations of  $N$ ,  $\nu$  and the product  $N\nu$  with height in the ionosphere. Since the collision frequency  $\nu$  falls off exponentially with height, while the electron density normally increases with height, the product  $N\nu$  will have a maximum in the D-region. This is true for day-time conditions. At night most of the ionization in the E- and D-region will disappear and only a very weak maximum in  $N\nu$  will be present at E-region heights. We conclude, therefore, that during the day nearly all of the non-deviative absorption will take place in the D-region, since the product  $N\nu$  is small above and below this region.

The deviative absorption will be important if the wave passes regions when (a) the wave frequency is near the plasma frequency (b) the collision frequency  $\nu$  is appreciable. The first condition is fulfilled near the level of reflection when the angle of incidence is small and, for frequencies, just above the penetration frequency of a layer. The second condition is true in the lower ionosphere.

Thus a radio wave will suffer considerable deviative loss for short distance communication via the E-layer but for long range transmission via the F-layer on frequencies well above the foE the deviative loss will be negligible. Since the latter case is of greater importance for high frequency communication, we conclude that, as a first approximation, the absorption of high frequency radio waves in the ionosphere is non-deviative. This means that nearly all

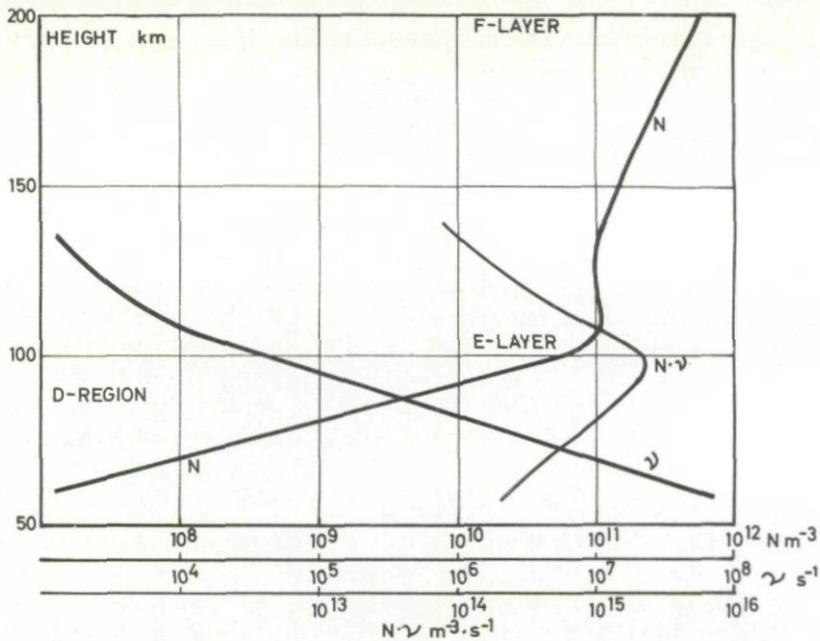


Figure 1.35 Height variations of  $N$ ,  $\nu$  and  $N \cdot \nu$  (day-time conditions)

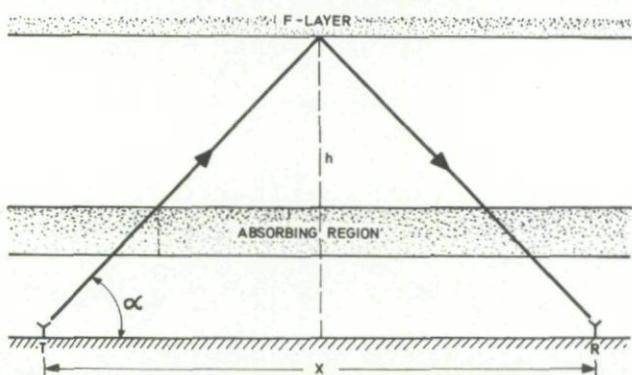


Figure 1.36 Illustrating how the absorption depends on distance

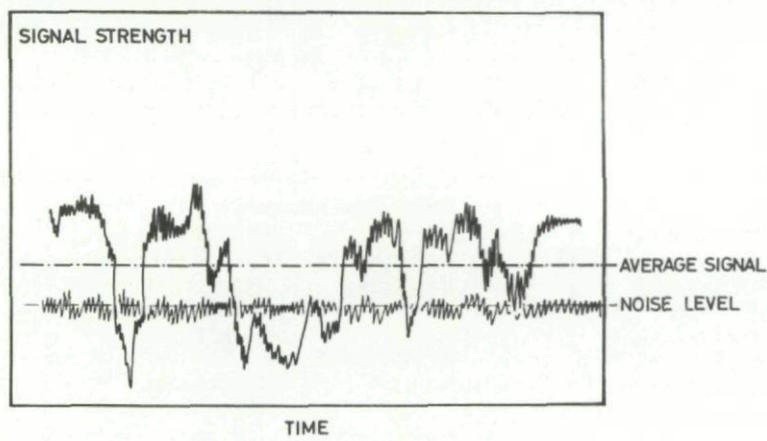


Figure 1.37 Illustrating the fading of a radio signal

the absorption takes place in the D-region and that the total ionospheric loss in decibels can be expressed as:

$$L = \text{constant} \frac{1}{f^2} \int_s N \nu ds \quad (\text{Eq. 1. 49})$$

in the following we need therefore only consider the variations of D-region absorption.

**The variation of ionospheric absorption with distance** — A rough picture of how the absorption for a wave of a given frequency will vary with distance between transmitter and receiver can be obtained from Figure 1. 36. Let us assume that the wave is reflected from the F-layer, and that the wave frequency is well above foE so that there is little deviation of the ray in the E-region. The ray path will then be close to the triangular path drawn in the figure. As explained on page 23 most of the absorption will be non-deviative and will take place in a narrow height region. The absorption will therefore be roughly proportional to the length of the ray path through this region. Thus, if the absorption at vertical incidence is  $L_v$  dB the absorption  $L$  at an angle of elevation  $\alpha$  will be:

$$L = \frac{L_v}{\sin \alpha} = L_v \sqrt{\left(\frac{x}{2h}\right)^2 + 1} \text{ (dB)}. \quad (\text{Eq. 1. 50})$$

when  $h$  is the height of the reflecting layer and  $x$  is the distance between transmitter and receiver. An example will serve to illustrate the use of equation 1. 50. Suppose  $h = 200$  km,  $x = 500$  km and the ionospheric absorption is  $L_{500}$ . How much will the ionospheric absorption increase when the distance is increased to  $x_2 = 1500$  km? From equation 1. 50, we obtain  $L_{1500} = L_{500} \cdot 2.54$  dB.

The simple approach just described cannot be used for waves reflected from the E-region at large angles of elevation, because, in this case, the deviative absorption will be appreciable.

On page 30 it was shown that a radio wave of frequency  $f_\alpha$  and at an angle of elevation  $\alpha$  will be reflected at the same real height in the ionosphere as a wave of frequency  $f_{90} = f_\alpha \sin \alpha$  transmitted vertically. It may also be shown that the corresponding total absorptions  $L_{f_\alpha}$  and  $L_{f_{90}}$  suffered by these waves are related as:

$$L_{f_\alpha} = L_{f_{90}} \cdot \sin \alpha \quad (\text{Eq. 1. 51})$$

This expression is not exact for the actual ionosphere, but it is a useful guide which enables us to estimate the oblique incidence absorption for a given frequency  $f_\alpha$  if we know the absorption of the 'equivalent vertical wave' of frequency  $f_\alpha \sin \alpha$ .

Note that the deviative absorption is included in equation 1. 51.

**Total ionospheric absorption** — We have found that, the ionospheric absorption for high frequency radio waves, mainly takes place in the lower ionosphere and is non-deviative in character. The dependence of this loss on frequency and angle of elevation has been discussed. The time variations in the absorption were described on page 15 where it was found that the losses depend on solar zenith angle and sun spot cycle.

For middle latitudes where the absorption behaves fairly regularly, it is possible to find a formula, which includes all the factors discussed, to give the

average, normal behaviour of the total ionospheric absorption:

$$L = 430 (1 + 0.0035R) \cos^{0.75} \chi \frac{1}{\sin \alpha} \frac{1}{(f + f_L)^2}, \quad (\text{Eq. 1.52})$$

decibels per hop. Here R is the sun spot number,  $\chi$  the solar zenith angle,  $\alpha$  the angle of elevation and f the wave frequency. The formula holds for the ordinary ray and the effect of the earth's magnetic field is included through the parameter  $f_L$ , which, at middle latitudes, is about 1 MHz. This parameter is associated with a natural resonance of the free electrons due to the geomagnetic field. The formula does not include the effects of the winter anomaly, page 20. In winter, therefore, the values obtained from equation 1.52 may, on some days, have to be multiplied by as much as a factor of two.

#### Attenuation due to fading

So far we have only discussed the long term variations in the average field strength of the received signal, such as diurnal, seasonal variations etc. In fact the amplitude of the signal reflected from or passing through the ionosphere can fluctuate very rapidly about this average. Amplitude variations of the order of ten to one may occur within a few seconds but less rapid fluctuations with periods of as much as thirty minutes to one hour can also sometimes be observed. This fading of the signal can be caused by a number of different mechanisms of which some of the most important will be discussed.

The effect of fading on the quality of reception is illustrated in Figure 1.37. Even if the average field strength received is above the radio noise level, the instantaneous field strength may well be below the noise level for a great percentage of the time. This will cause a poorer quality of reception or even make communication impossible. In order to avoid or lessen the effects of fading, one, therefore, needs a stronger signal than would have been necessary if no short term fluctuations were present. One way of obtaining a stronger signal is, of course, to increase the wave frequency. Thus the presence of fading will mean an increase in the LUHF. In the predictions of the LUHF, allowances are made for fading by increasing the estimated total absorption by different amounts for the different kinds of services.

The fading can be caused by interference between waves reflected from different levels in the ionosphere, for example, interference between high angle and low angle rays, page 29. It can also be caused by irregularities in ionization density near or below the level of reflection. Movements of ionization can give rise to variations in the signal amplitude when clouds of large electron density cross the ray path. Another cause of fading may be local curvatures in the reflecting layer so that the layer acts as a focusing or defocusing mirror. Near the skip distance, a deep fading may be observed. This is caused by interference between the high and low angle rays, page 29. The period of the fading depends largely on the acting mechanism. Thus the fading period for interference fading can be from a fraction of a second to several seconds, while focus fading can have a period of the order of half an hour.

For signals passing through the ionosphere, such as satellite transmissions and cosmic radio noise, a very rapid fluctuation can often be observed. These 'scintillations' are believed to be caused by irregularities in the F-region.

#### Total loss and received field strength

We should now be in a position to add up the losses a radio wave suffers on its way from transmitter to receiver and to give an expression for the received field strength. The different types of losses discussed in the preceding sections are:

- (a) Free Space attenuation  $A_s$

- (b) Ground loss  $A_g$
- (c) Ionospheric absorption  $L$
- (d) Loss due to fading  $A_f$

Suppose that the transmitter antenna radiates a total power of  $p$  kW and has a gain  $G$ . Then the field strength at the receiving antenna must be:

$$F = 10 \log p + G + 109.5 - (A_s + A_g + L + A_f), \quad (\text{Eq. 1. 53})$$

dB over  $1 \mu\text{V/m}$ . If the receiving antenna in the direction towards the transmitter has a gain  $G_r$ , the apparent field strength will be:

$$F' = F + G_r \quad (\text{Eq. 1. 54})$$

The radio noise level

Having made an estimate of the total losses and of the field strength of the received signal, we now turn to the problem of finding the intensity of radio noise. The quality of reception will depend on the ratio of the field strength of the wanted signal to the intensity of the inevitable background of unwanted radio signals created outside and in the receiver. This background of noise will be present at all frequencies but its intensity will vary with frequency.

The minimum signal to noise ratio that can be tolerated will depend on the character of the noise and on the type of communication service, and can be difficult to predict. However, some general rules can be given. The noise will nearly always have a broadband character and it is important to realize that the noise power, that is mixed with the wanted signal, will be proportional to the bandwidth of the receiver. The amplitude of the noise signal will therefore increase with the square root of the bandwidth. Radio noise must be measured in power per frequency unit, for example in  $\text{W/Hz}$ . The radio noise consists of five different types of noise:

- (a) Atmospheric noise
- (b) Cosmic noise
- (c) Man-made noise
- (d) Thermal noise in the antenna and the receiver
- (e) Precipitation noise

Here we shall mainly be concerned with the first three types of noise because they are most important to high frequency communication.

The thermal noise in the receiver and the antenna is generated by the random motions of the electrons in the tubes and the other components. These motions, which become more rapid with increasing temperature, create weak fluctuating noise currents in the conducting materials. The thermal noise however, is not normally the limiting factor in the high frequency bands but becomes extremely important above 300 MHz.

The last type of noise, precipitation noise, is generated by snow, rain, sea spray or dust falling on the antenna and inducing voltages in the wires. This type of noise can be very serious under some conditions.

The noise power is measured in terms of a quantity  $F_a$ , which is defined as the external noise power per unit bandwidth available from a lossless antenna, expressed in decibels above a certain reference value. The following is a more complete definition:

If  $F_a$  and the bandwidth  $b$  of a system is given and if the external noise is inci-

dent uniformly in all directions on the antenna, then the total noise power available at the terminals of a lossless antenna is given by:

$$P_n = F_a + B - 204 \text{ dB above 1 W} \quad (\text{Eq. 1. 55})$$

here  $B = 10 \log b$

Note: The term radio noise must be properly defined in order to be useful, and the noise power generated in a thermal source, for example a resistor, can be used as a reference for measurements of noise power. If the resistor is  $R$  ohms, the mean square noise voltage over the resistor will be:

$$\overline{E^2} = 4 RkTb \quad (\text{Eq. 1. 56})$$

where:

$k$  = Boltzmanns constant =  $1.38 \cdot 10^{-23}$  joule/°K

$T$  = absolute temperature in degrees Kelvin

$b$  = bandwidth in Hz

We have here assumed that the thermal noise has a uniform distribution over the bandwidth  $b$ . The term  $\overline{E^2}/4R = kTb$  is called the available noise power from the thermal source. A reference temperature  $T_0 = 288.63^\circ\text{K}$  ( $63^\circ\text{F}$ ) is often chosen and the available noise power per Hz from the reference source will then be:

$$p_0 = kT_0 = 4 \cdot 10^{-21} \text{ W} \cdot \text{s}$$

Suppose that radio noise is incident uniformly in all directions on a lossless antenna and that the noise power available at the terminals of the antenna is  $p_n$  watts. If the bandwidth is  $b$ , the power available per unit bandwidth will be  $p_n/b$  watts per hertz (or  $\text{W} \cdot \text{s}$ ).

The noise power per unit bandwidth in decibels above the reference  $p_0$  will be:

$$F_a = 10 \log \frac{p_n}{b \cdot p_0} \quad (\text{Eq. 1. 57})$$

This gives:

$$P_n = F_a + B - 204 \text{ dB over 1 watt} \quad (\text{Eq. 1. 55})$$

where  $P_n$  is the total available noise power in decibels over 1 watt,  $B = 10 \log b$  and  $10 \log p_0 = -204$ .

**Atmospheric noise** — A stroke of lightning is an extremely strong but short-lived electric current resulting from high voltages between the atmosphere and the earth or between parts of the atmosphere. Such a discharge will emit a strong pulse of electromagnetic radiation. The frequency spectrum from a lightning discharge is shown in Figure 1. 38. We will see that there is a broad maximum around 10 kHz and that the intensity of the emitted radiation decreases steadily towards higher frequencies. The electromagnetic energy from an electrical discharge in the lower atmosphere can propagate to great distances by multiple reflections from the earth and the ionosphere. It has been estimated that about 50 000 thunderstorms occur every day and that on the average there are about 100 strokes of lightning every second. Thus, it is not difficult to understand that, atmospheric electricity can create a background of radio noise that can seriously influence radio communications.

The principal centers of atmospheric noise generation are the active thunderstorm areas in Equatorial Africa, Central America and the East Indies. The radio noise will of course follow the same general laws of propagation as other types of electromagnetic radiation. The noise propagating from a distant source will, therefore, be absorbed in the ionosphere and the magnitude of the absorption will depend on the solar zenith angle over the transmission path.

Extensive measurements of radio noise are currently being made and in order to predict the intensity of atmospheric noise, world wide radio noise data are published in the form of maps giving the expected geographical distributions of the parameter  $F_a$  at a frequency of 1 MHz. The maps are issued by CRPL for different seasons and times of day, and Figure 1.39 gives an example of such a map. No significant long term variations in the radio noise level have been found.

To find  $F_a$  at frequencies different from 1 MHz, charts, showing the expected frequency variations for this parameter, are also issued. Examples of such curves, derived for day- and night-time conditions respectively, are shown in, Figure 1.40 a and b. The numbers on the curves in Figure 1.40 refer to the values of  $F_a$  at 1 MHz.

The types of graphs shown in Figures 1.39 and 1.40 are based on measurements made at many geographical locations and at many frequencies. The curves represent the expected average behaviour of the noise level and, at a particular time and place, the actual values of  $F_a$  can of course deviate from the predictions.

Note: The frequency variation of required field intensity shows some interesting features, which illustrate some of the remarks made above. We notice that, at night, most of the curves show a steady decrease with increasing frequency up to a frequency that corresponds to the MUF for the noise transmission path and then the field strength falls off very rapidly with frequency. Obviously, no noise on frequencies above the MUF will reach the receiver. The steady decrease with frequency at night corresponds to the variation to be expected from the power spectrum from a lightning flash shown in Figure 1.38. During the day the field intensity curves show a minimum in the medium frequency band. In this band therefore, the noise is strongly absorbed in the ionosphere and a smaller field strength is necessary to obtain good communication.

**Cosmic radio noise** — At frequencies above the foF2, radio noise of extraterrestrial origin is known to be the main source of interference under many circumstances. This noise has the same characteristics as the thermal noise generated in the components of the receiver. The cosmic noise originates in certain areas in the sky, mainly near the center of the galaxy. The intensity of the noise, therefore, varies regularly at any particular place on the earth's surface, as the earth rotates. Very little is known about the possible mechanism which generates this type of noise, although some tentative explanations have been put forward.

The intensity of cosmic radio noise decreases with increasing frequency and the normal level of this type of noise is indicated by the dashed line at the higher frequencies in Figure 1.41.

**Man-made noise** — Man-made noise, that is noise generated by electrical machinery, switches and so on, can be a serious problem. By selecting the site for the receiver far away from the noise sources and by using suitable filtering, this type of noise can be reduced to a minimum.

Even at very quiet receiving sites, however, there will always be a background of man-made noise and the expected intensity of man-made noise at such a site

is indicated by the dot-dashed line in Figure 1.41. As can be seen this type of noise can be comparable to the atmospheric noise under some circumstances.

### Resultant LUHF

We can now sum up our knowledge. As mentioned previously the problem of determining the Lowest Useful High Frequency falls into two parts, (a) to determine the received field intensity as a function of frequency, and (b) to determine the intensity of radio noise at the receiving site and find the frequency dependence of this noise.

On pages 34 to 44 we discussed how the first part of the problem could be solved in principle. After estimating the power radiated from a transmitter antenna in different directions, it was shown how the various types of loss will reduce the signal strength on the way from the transmitter to the receiver. The most important loss mechanisms were free space loss, ionospheric absorption, ground loss and loss due to fading. Subtracting these losses from the radiated field strength it was possible to arrive at values of the field strength at the receiving antenna. The methods used are only approximate but give, nevertheless, useful results. In order to find the LUHF for a particular circuit, the computations must be made for a number of frequencies and for all the possible propagation modes between the two points. The mode giving the highest field strength will give the lowest LUHF.

Having obtained a frequency variation of the received field strength, the next step is to compare these results with the field intensity of radio noise versus frequency for the receiving station. The latter curve can be derived using the methods described earlier.

If the available signal power is  $p_s$  and the noise power is  $p_n$  the signal to noise ratio

$$\text{ratio } r = \frac{p_s}{p_n}.$$

The signal power will then be:

$$P_s = P_n + R = F_a + B + R - 204 \text{ dB over 1 W} \quad (\text{Eq. 1.58})$$

where  $R = 10 \log r$  and  $P_n$  is given by equation 1.55. Thus, if the minimum signal to noise ratio that can be tolerated is known, the minimum required signal power available at the terminals of a lossless antenna can be found from equation 1.58. If the properties of the actual antenna are known, the minimum required field strength at the receiving site can be deduced. Plotting the minimum required field strength and the received field strength as a function of frequency, the LUHF will be determined by the point at which the two curves cross. The results are in general of the type sketched in Figure 1.41.

The LUHF derived in this manner is not always well defined. For example, this can happen when the two curves in Figure 1.41 coincide over a frequency range.

The LUHF can also sometimes be multivalued, since it may happen that communication is impossible in one part of the HF band but possible on high frequencies above and below this band.

Note: It should be pointed out that, besides noise, the efficiency of radio communication systems can be limited by interference produced by other radio-communication systems. Interference can be a serious problem, and it is of course difficult to predict for any particular time and frequency. The minimum required signal to interference ratios are listed by CCIR. (CCIR (1959), Documents of the IX Plenary Assembly, Los Angeles, Recommendation 240).

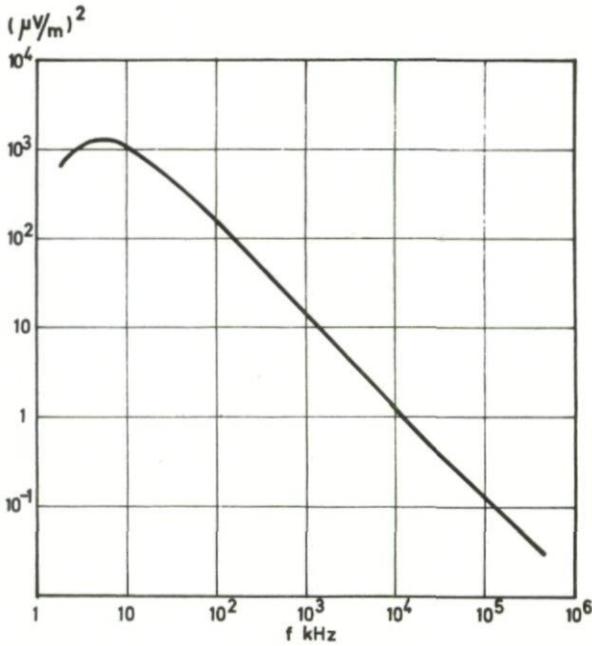


Figure 1.38 Frequency spectrum of radio noise from a lightning flash

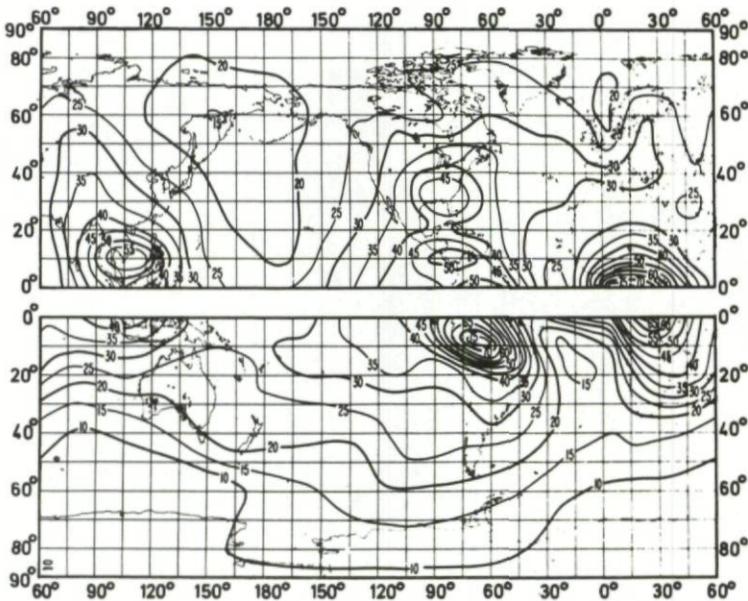


Figure 1.39 Map showing expected values of atmospheric radio noise  $F_a$  (dB above  $k T_O b$  at 1 MHz). Spring 0800 - 1200 LMT (After CCIR Report No. 322, Documents of Xth Plenary Assembly 1963)

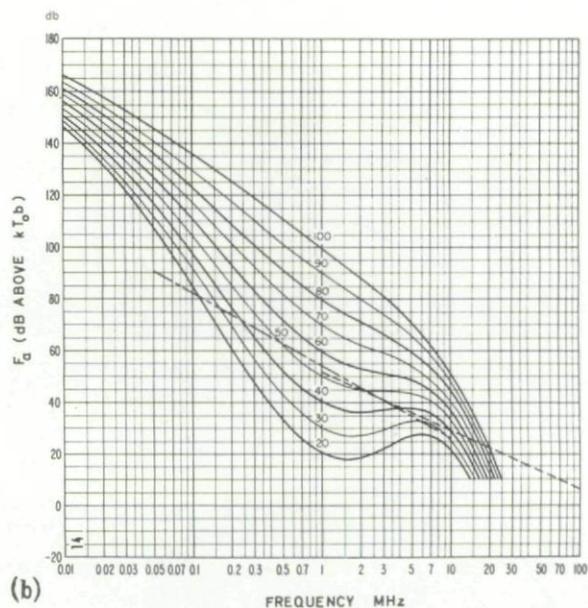
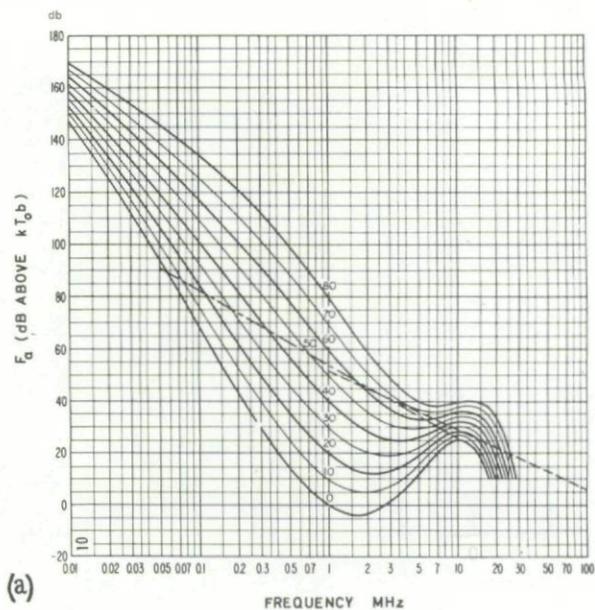


Figure 1.40 Typical variations of the radio noise parameter  $F_a$  with frequency  
 (a) Day-time conditions (Spring 0800 - 1200 LMT)  
 (b) Night-time condition (Summer 0000 - 0400 LMT)  
 (After CCIR Report No. 322, Documents of Xth Plenary Assembly 1963)

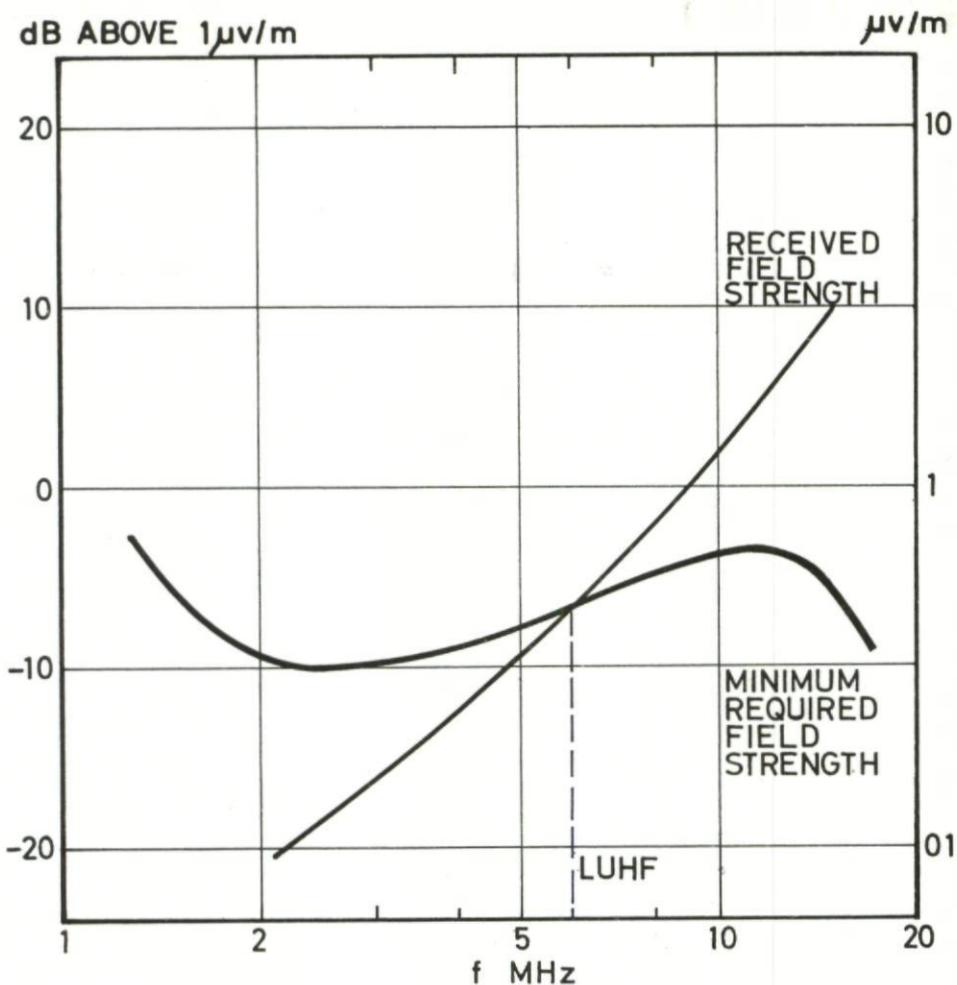


Figure 1.41 Typical frequency variations of minimum required field strength and received field strength

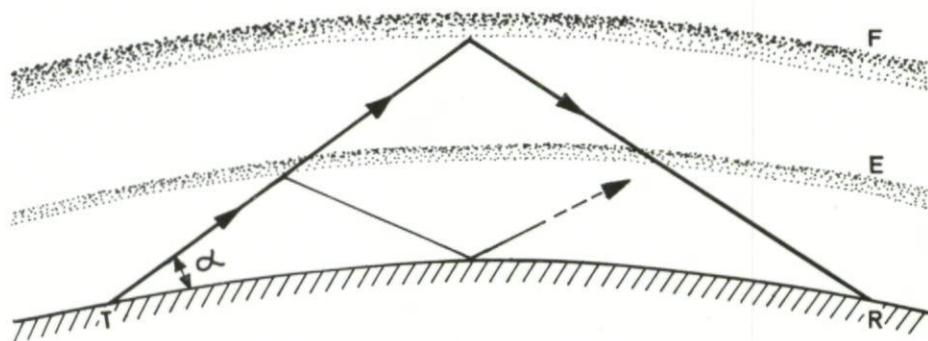


Figure 1.42 Screening by the E-layer

There is one important factor in the determination of the LUHF which we have so far ignored. In some cases the LUHF is not determined by the losses and by the noise level but by the geometry of the path. This will be explained in the following paragraphs.

**Screening by the E-region** — The LUHF can sometimes be determined by the geometry of the transmission path. This is illustrated in Figure 1.42. Suppose we have transmission from T to R at a frequency  $f$  which penetrates the E-layer and is reflected from the F-layer. If the working frequency  $f$  is decreased, then below a certain frequency  $f_s$  the wave will be reflected from the E-layer instead of the F-layer. The receiver can then be outside the range of the one hop E-layer reflection and the signal reflected twice from the E-layer will normally be too weak to be detected because it has passed four times through a strongly absorbing D-region. The LUHF in this case will therefore be the lowest frequency at which transmission via the F-layer is possible. For all frequencies lower than this the transmitter will be screened from the receiver by the E-layer.

Assuming, as in Figure 1.42, that the ray is not deviated when passing through the E-region, the screening off frequency will be given by  $f_s = foE/\sin \alpha$  where  $\alpha$  is the angle of elevation. However, the ray will normally follow a path more like the dotted line in the figure because of the deviation in the E-region. This means that in practice the screening off frequency will be near the  $foE$ .

## 2

# Predictions of Ionospheric Conditions

### Introduction

In Chapter 1 of this book, we dealt with the basic principles of radio communication via the ionosphere. We discussed how the ionospheric layers are formed and how they vary over the day, with season and with sun spot activity. We showed that, for a given circuit, at a given time, there must be a maximum useable frequency, MUF, given as:

$$\text{MUF} = f_o \cdot M(d) + \text{correction} \quad (\text{Eq. 2.1})$$

where  $f_o$  is the critical frequency (ordinary ray) of the ionospheric layer in question and  $M(d)$  the factor over the distance  $d$  between the transmitter and the receiver. The critical frequency is determined by the maximum electron density for the ionospheric layer, while the factor  $M(d)$  is determined by the shape and height of the layer. The correction is approximately 0.7 MHz at zero distance, and is reduced to zero at 4000 km. For a given circuit, at a given time, there must also be a lowest useful high frequency, LUHF, which is determined by the equipment parameters, the power losses along the path, and the radio noise level at the receiving site.

In the present chapter we shall be concerned with how the knowledge gained from world studies of the ionospheric layers can be used in predicting the future behaviour of the ionosphere, in order that the best possible use can be made of this for practical communication purposes. In principle, there are two types of predictions that can be made:

- (a) Predictions of frequency limitations months or years ahead
- (b) Predictions of disturbances at relatively short notice

The first type of predictions are normally made 2 or 3 months beforehand. These predictions can therefore not give the exact MUF or LUHF for a given circuit at a given time. In fact, the different prediction services have chosen to predict the diurnal variation of the monthly average values of the parameters involved.

The second type of predictions are also being made, see page 107. These are of importance mainly for arctic regions where radio communication is often very difficult. We shall discuss the arctic problems in some detail in Chapters 3 and 4 of this book, and therefore limit the discussion in the present chapter to the first type of predictions.

### Methods of predictions

The problem of predicting the future behaviour of a physical parameter is basically one of extrapolation. In order for an intelligent extrapolation to be possible, a good knowledge of the time variations of the physical parameter is essential. In the following paragraphs we shall describe the methods used in predicting the frequency limitations imposed by the ionosphere on HF radio communication, and the way in which these predictions are being presented.

#### Physical parameters to be predicted

In planning a high frequency sky-wave circuit the following characteristics need to be determined in order to estimate the required transmitter power:

- (a) a useful frequency

- (b) the possible modes of propagation
- (c) the power losses along the paths
- (d) the noise level at the receiving site
- (e) the equipment parameters, such as antenna gain and type of service required

A useful frequency must be selected between the MUF and the LUHF. The MUF is determined, as we have seen, by the maximum electron density in the ionospheric layer represented by the critical frequency  $f_o$  and by the factor  $M(d)$ . It should be determined for each of the possible modes of propagation.

The LUHF is determined by the power losses along the path, the noise level at the receiving site and the equipment parameters. It must be derived for all the possible modes of propagation. For each of these modes, a useful frequency band can be determined, and the resultant MUF and LUHF can be deduced by combining the results for all modes.

For practical purposes, the number of possible modes for which the above exercise must be made, can often be considerably reduced. For example, if a one-hop mode is possible for a particular layer, the two hop and higher order modes can normally be disregarded. Similarly, if the distance is too long for a one-hop E- or F1-mode, and an F2-mode is possible, the E and F1 multiple hop modes can be disregarded.

For distances up to 4000 km, it is possible to take into account, separately, all the important modes over which communication is possible. For longer distances, two or multiple hop modes are the only modes of importance, and the situation becomes more complex. For these long-distance circuits, averaging methods of prediction are used. These will be explained in detail later.

#### Basic principles of predictions

As already stated, the problem of predicting the future behaviour of a physical parameter is basically one of extrapolation. This is illustrated in Figure 2.1 where we have shown the monthly median noon values of the critical frequency for the E-layer,  $f_oE$ , as observed in Washington DC, in the period January through December 1962, and where we have extrapolated the curve through these points in order to predict the values for January through March 1963. In the illustration we have also shown the average yearly variation of the parameter as deduced from observations during the years 1953 to 1962, and it is shown how the established seasonal variation may help the extrapolation. In order to make intelligent predictions of the parameters of importance, a good knowledge of the diurnal, seasonal, and sun spot cycle variation of the parameters is essential. If we start with the parameters of importance for the MUF, and consider the E- and F1-layers first, the problems are fairly simple. These layers are well controlled by the solar radiation, and their critical frequencies can be given in terms of analytical expressions as follows:

$$f_oE = A(\cos\chi)^{1/4} \quad (\text{Eq. 2.2})$$

$$f_oF1 = B(\cos\chi)^{1/5} \quad (\text{Eq. 2.3})$$

where  $\chi$  is the solar zenith angle.

Note: According to theory, in which it is assumed that only the radiation from the sun is of importance, a factor of  $1/4$  in the exponent is to be expected. For the F1-layer a better agreement is found with an exponent of  $1/5$ . This means that the variations are not only due to those of the sun, but that also other processes, mainly diffusion, are of importance.

CRITICAL FREQUENCY IN MHz

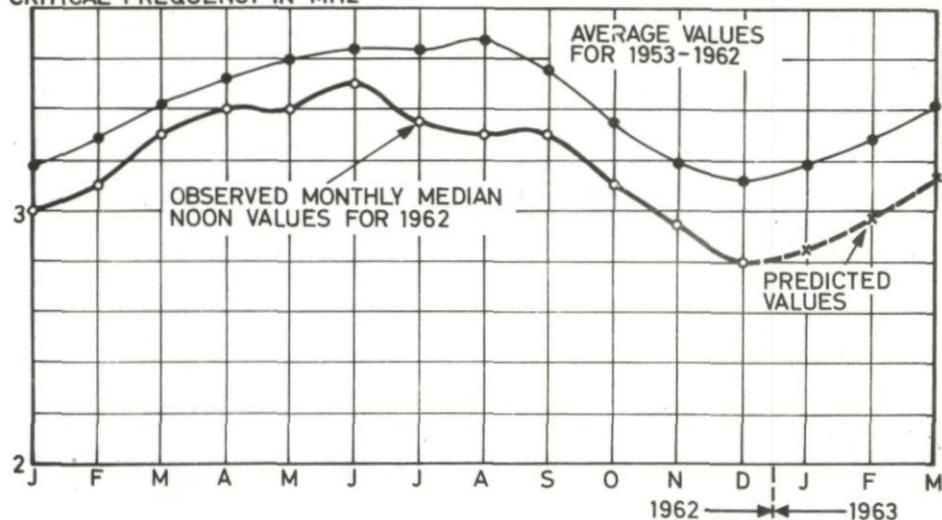


Figure 2.1 Extrapolation of monthly median noon values of the critical frequency foE

CRITICAL FREQUENCY IN MHz

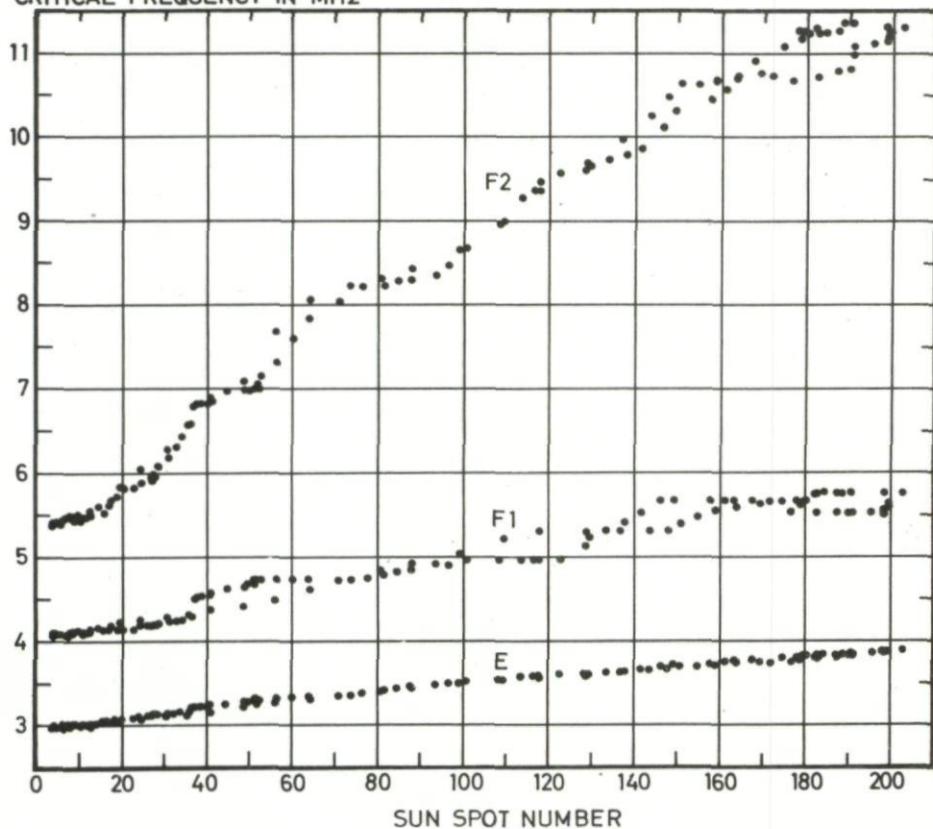


Figure 2.2 12-month running averages of foE, foF1 and foF2 versus 12-month running average sun spot number R

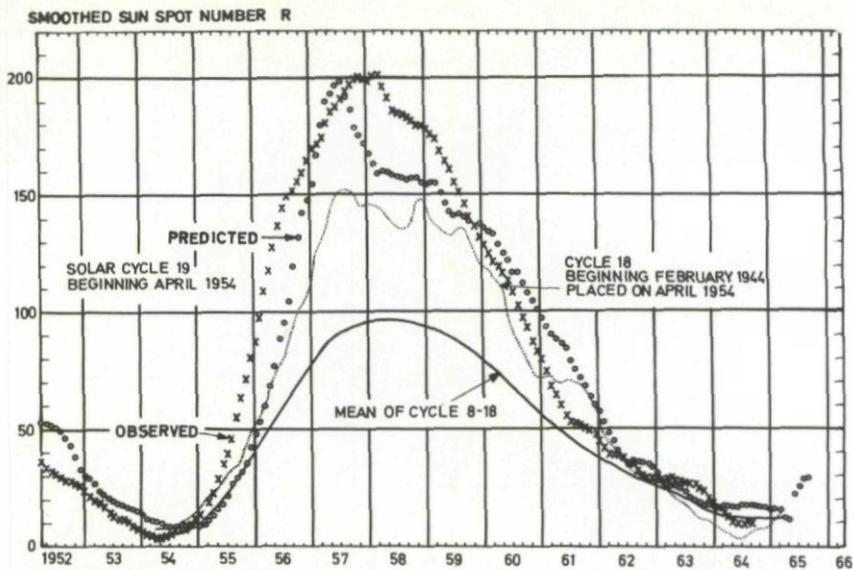


Figure 2.3 Predicted and observed sun spot numbers  
 (After CRPL, National Bureau of Standards, Ionospheric  
 Predictions for September 1965)

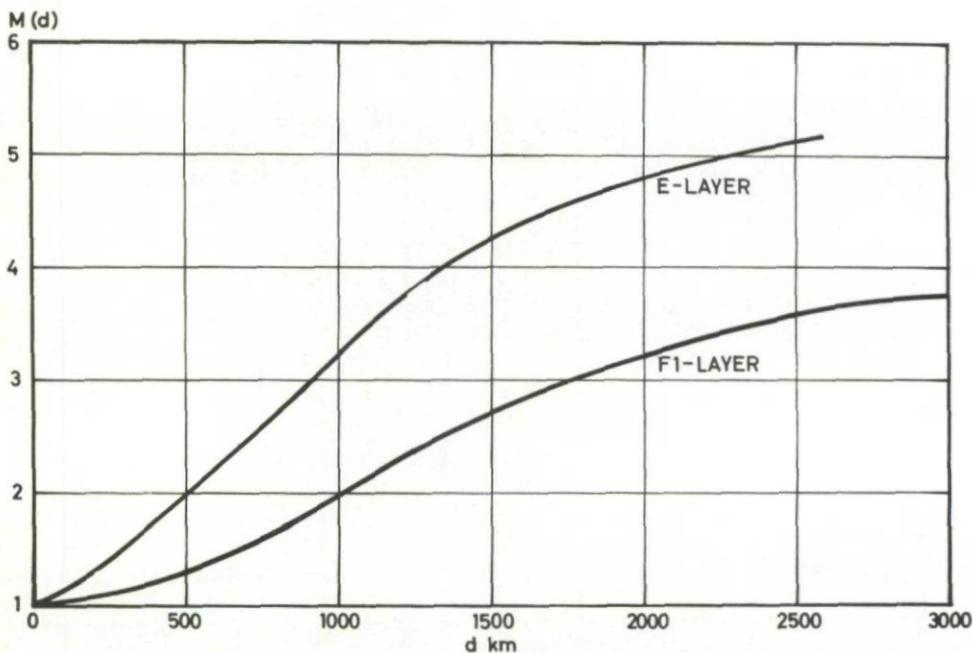


Figure 2.4 The factor  $M(d)$  plotted as a function of distance for the E- and F1-layers

A and B in equations 2.3 and 2.4 are constants for a particular value of the sun spot number R. The way in which they vary with R is illustrated in Figure 2.2 where the observed 12-month running average values of foE, foF1 and foF2 are plotted against the 12-month running average sun spot number. The ionospheric observations were made at Washington DC. The 12-month running average for a given month is the mean of the monthly median values taken over one year centered on this month.

It can be seen from Figure 2.2 that foE, foF1 and foF2 are roughly linearly related to the average sun spot number R. Although slight variations in these relations with time and geographical position have been observed, it is possible for prediction purposes to use one set of relations. The analytical expressions used by the ITSA (Institute for Telecommunication Sciences and Aeronomy) in the United States in their prediction service for the variations of foE and foF1 are the following:

$$foE = 0.9 [(180 + 1.44 R)\cos\chi]^{1/4} \quad (\text{Eq. 2.4})$$

$$foF1 = (4.3 + 0.01 R)\cos\chi^{1/5} \quad (\text{Eq. 2.5})$$

The problem of predicting the critical frequencies for the E- and F1-layers is therefore reduced to that of predicting the average sun spot number R.

Note: The way in which foE is related to R can be studied by expanding  $0.9 [(180 + 1.44 R)\cos\chi]^{1/4}$  in a power series of  $1.44 R/180$ . We obtain:

$$0.9 [(180 + 1.44 R)\cos\chi]^{1/4} = 3.29 \cdot \cos^{1/4}\chi \left( 1 + \frac{1}{500}R - \frac{3}{500000}R^2 + \frac{7}{25000000}R^3 \dots \right)$$

For values of R smaller than about 100, second or higher order terms in the series can be disregarded and for such values of R the linear relationship holds. For higher values of R, the second order term becomes important and the foE will increase more slowly with R.

In predicting the average sun spot number, use can be made of a long series of observations of this quantity. In Figure 2.3 we have compared values that were predicted for R with those that were actually observed for sun spot cycle number 19, lasting from 1954 to 1964. In predicting the numbers, use was made of the values observed during cycle number 18 together with the average behaviour presented at the mean of cycle 8 to 18. In order to derive the MUF for the possible E- and F1-layer modes, one must also know the values of the factor M(d). As the height and shape of these layers do not vary significantly, one can use fixed curves that give these quantities as a function of distance. These curves are shown in Figure 2.4. We see that the factor for the E-layer varies from 1 to 5 for distances between zero and 2000 km, and for the F1-layer between 1 and 3.75 for distances between zero and 3000 km.

In the case of the F2-layer the problems of predicting the MUF is more difficult than for the E- and F1-layers mentioned previously. The critical frequency cannot be represented by simple analytical expressions like those given in equations 2.4 and 2.5, and the height and shape of the layer varies both with time and position so that the factor must also be predicted. There is a clear linear relationship between the foF2 and sun spot number, Fig. 2.2, but the coefficients vary from one station to the other and are also different at one station for different sun spot cycles.

To deal with the problem of predicting the highly irregular variations of foF2 and M(d)F2, world maps of the parameters are prepared for a given instant of universal time giving contours for constant values of foF2 and M(4000)F2. Such maps have traditionally been drawn by hand, advantage being taken of the cartographer's knowledge of the appropriate ionospheric features. In recent

years however, computer methods have been developed to make the best use of the available data.

If the factor  $M(4000)F_2$  for a distance of 4000 km is known, the shape and height of the layer is roughly known, and it is possible to use graphs to convert the value of  $M(4000)F_2$  into the correct factor for any distance of communication. Page 18.

In order to make an estimate of the LUHF for a particular circuit, predictions of the ionospheric absorption and knowledge of the noise level at the receiving site are needed, page 34. The ionospheric absorption will depend on the solar zenith angle  $\chi$  in much the same way as the critical frequencies for the E- and F1-layers.

Prediction of the radio noise level is based upon results from actual radio noise measurements. The diurnal, seasonal, and geographical changes have been established from such measurements carried out at a number of selected stations. Page 45.

### **Form and presentation of predictions**

A number of different organizations in various countries are preparing predictions. The way in which these are made, and the form in which they are presented are by and large the same.

We shall limit the discussion to the predictions presented by the ITSA of the Environmental Science Services Administration in the United States. As already stated, this organization is preparing the predictions by means of an electronic computer. With a computer the predictions can be quickly and easily revised as new data become available.

The numerical map form of the predictions is particularly useful when large numbers of propagation path computations are required. In many countries organizations are set up that can provide predictions for different kinds of services involving large numbers of radio circuits. Such organizations use punched cards from the CRPL and prepare the predictions by means of their own computers. The CRPL also present their predictions in a form suited for manual evaluation. This method is described in detail in CRPL Handbook 90, Handbook for CRPL Ionospheric Predictions based on numerical methods of mapping. In the following paragraphs we shall describe briefly the method of preparation and presentation of the predictions.

### **Mode considerations**

Up to distances of 4000 km one hop F1- and F2-modes are possible, while one hop E-modes are not possible for distances greater than 2000 km. F1-layer ionization depends upon the solar zenith angle in much the same way as the E-layer ionization. In fact, the E-region ionization may control the propagation up to distances of 2000 km and the F1 ionization for distances between 2000 and 3000 km. During the day time, particularly in the summer and during periods of low solar activity, the regular E-F1-layer MUF (for longer distances) is often higher than the F2-MUF.

For distances greater than 4000 km only multiple hop modes are of importance, and the preparation of the predictions would become complex if a straight-forward analysis based on all the possible modes should be carried out. A simpler method of prediction is therefore used for such distances. This is a two control point method in which it is assumed: (i) that there are two F2-layer control points on the great circle, 2000 km from each terminal, (ii) that communication is possible at the frequencies that can arrive at or leave the respective terminals, and (iii) that the resulting MUF is the lowest of the two MUF-4000 which are determined for the two control points.

The method is illustrated in Figure 2.5. It does not take into account the way in which the waves are propagated between the two control points. This is an empirical method, and it has proved in practice, to give reasonably correct predictions.

In the same way as described above the MUF-E can be taken into account. In this case control points at distances of 1000 km from the terminals are used, because 2000 km is the maximum distance for one hop E transmission.

For very long distances it is often advisable to consider propagation along both the short and long arcs of the great circle.

There is one important mode-consideration which is often overlooked. In order for an F-layer mode to be possible, the frequency must be so high that the wave is not reflected from the E-layer, Fig. 2.6. Suppose that a ray path with an angle of elevation  $\alpha$  intersects the E-region at a point  $P_E$  and that the ray is reflected from the F-layer at a point  $P_F$ . There will obviously exist a lowest frequency  $f_{SE}$  which will be able to penetrate the E-layer at  $P_E$  under this elevation angle. An approximate expression  $f_{SE} = foE/\sin \alpha$  for this screening off frequency was mentioned on page 30. A more reliable result can be obtained if we realize that the screening off frequency must be equal to the MUF-E for the distance TC. The distance TC has been calculated as a function of elevation angle  $\alpha$ , and the results are shown in Figure 2.7. In the case presented in Figure 2.6 an F-layer mode is only possible for frequencies between the MUF( $d_F$ )F2 controlled at the point  $P_F$  and MUF( $d_E$ )E controlled at the point  $P_E$ . This screening effect of the E-layer may introduce a lower frequency limit.

#### Determination of related parameters

In order to prepare the curves giving the variations of MUF and LUHF over the day for particular circuits, the following related information is needed:

- (a) The distance along the great circle path and the location of the control points
- (b) The variation of the solar zenith angle over the day at the control points
- (c) The radio noise level at the receiving site

The great circle distance and the positions of the control points can be deduced simply by graphical methods. In the CRPL, Ionospheric Predictions Handbook for example, charts are given which can be used for this purpose. In Figure 2.8 these charts are shown.

- (a) shows a world map drawn in a modified cylindrical projection, and
- (b) the great circle lines and distance between these in thousands of kilometres. By means of the two charts, the wanted quantities can easily be determined by a transparency technique.

Note: The procedure is as follows: Place a transparent sheet of paper over the map and mark the equator and the two control points. Place the transparent sheet over the great circle chart so that the equatorial lines coincide and slide the sheet horizontally until the control points lie on the same great circle. The distance between these points can then be found by counting the number of the dashed lines crossing the great circle path between the points.

The variation over the day of the solar zenith angle  $\chi$  can also be read off charts that are published in the Prediction Handbook. An example of the charts given in the CRPL Prediction Handbook is shown in Figure 2.9. From these charts the  $\chi$ -values can be read off for any values of geographical latitude and the local time, and it is thus possible to construct the  $\chi$ -variations over the day at any control point.

In order to make an estimate of the noise level at the receiving site, maps are available that give the expected noise level at different places, time of day and season. These maps were discussed on page 45.

For charts of radio noise and a more detailed discussion, reference is given to Reports No. 65 and No. 322 of the International Radio Consultative Committee (CCIR).

MUF for distances less than 4000 km

For such distances, only one hop modes are of importance. The MUF-E (and MUF-F1) is determined by the sun spot number R and the values of the solar zenith angle  $\chi$  at the control point. The value of R is given in the prediction issue, while the value of  $\chi$  is given in the CRPL Ionospheric Prediction Handbook. Knowing R and  $\chi$  the MUF for the E-layer at a distance of 2000 km can be read off a nomogram as shown in Figure 2. 10. In order to determine the MUF for the actual distance d, another nomogram as shown in Figure 2. 11 is used. This nomogram is made in such a way that the MUF for both the E- and F1-layers are included, and the distance axis is therefore extended to 3000 km.

In order to determine the MUF controlled by the F2-layer, a map giving contours for constant values of foF2 is first used. A similar map could then be used to read off values of the factor M(4000)F2, (or the M value for any chosen distance) and this value could be converted into M for the actual distance. The MUF(d)F2 could then be obtained as:

$$\text{MUF}(d)\text{F2} = \text{foF2} \cdot \text{M}(d) + \text{correction} \quad (\text{Eq. 2. 6})$$

A somewhat different method is used in the ITSA predictions. Values of MUF(0)F2 and MUF(4000)F2 = foF2 · M(4000)F2 are read off maps, and the MUF for the actual distance is determined by a nomogram. In Figures 2. 12 and 2. 13 the maps and the nomogram are shown. The maps of Figure 2. 12 refer to a given universal time, and different maps must therefore be used for different times of the day.

MUF for distances greater than 4000 km

For distances greater than 4000 km, the two control point method must be used. For the E-layer the MUF is determined by the control point for which the highest value of  $\chi$  is found, and is given as the MUF(2000)E determined from that value, by means of the nomogram given in Figure 2. 10.

The F-layer controlled MUF is given at any time as the lowest of the MUF(4000)F2 determined for the two F-layer control points.

Experience shows that for distances greater than 4000 km, the screening frequency  $f_{gE}$  discussed previously, see page 52 is roughly equal to the MUF-E. The MUF-E thus represents a lower frequency limit for transmission via the F-layer.

The optimum traffic frequency FOT

Because the absorption of the waves is inversely proportional to the square of the frequency f, one should normally choose a working frequency as high as possible. The E- and F1-layers are well behaved layers and in the case of control by these layers, a frequency close to the MUF can be selected.

If there is F2-region control, however, communication at a frequency equal to the MUF would be possible only during 50% of the time, and it is advisable to select a frequency somewhat lower than the MUF. For this reason an optimum traffic frequency, FOT (Fréquence Optimum de Travail) is defined

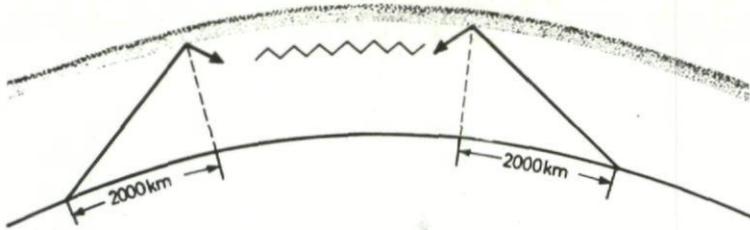


Figure 2.5 The two control point method

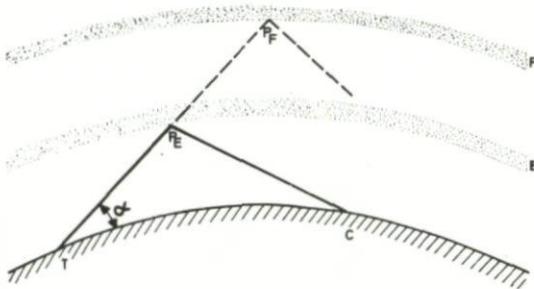


Figure 2.6 The effect of screening by the E-layer

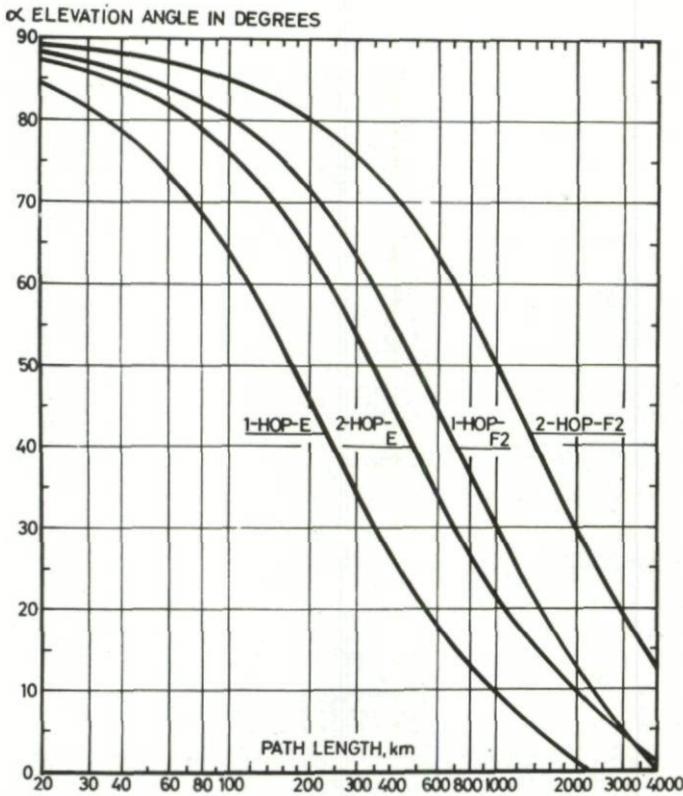


Figure 2.7 Radiation angle versus path length, based on virtual reflection heights (E-layer 105 km; F2-layer 320 km)  
 (After K. Davies, Ionospheric Radio Propagation, National Bureau of Standards, Monograph 80, 1965)

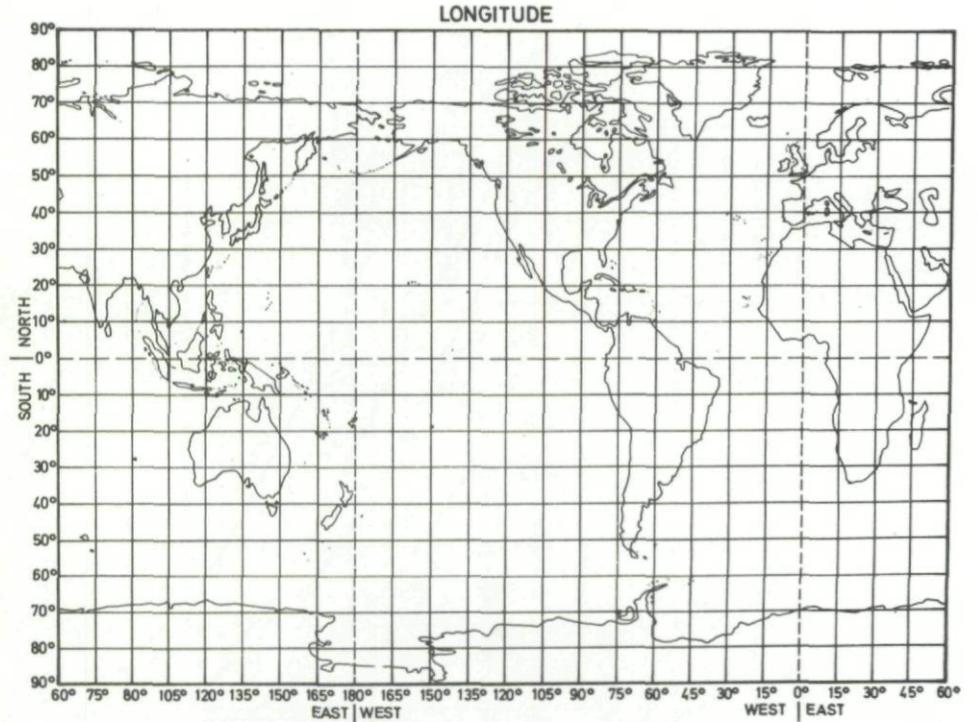
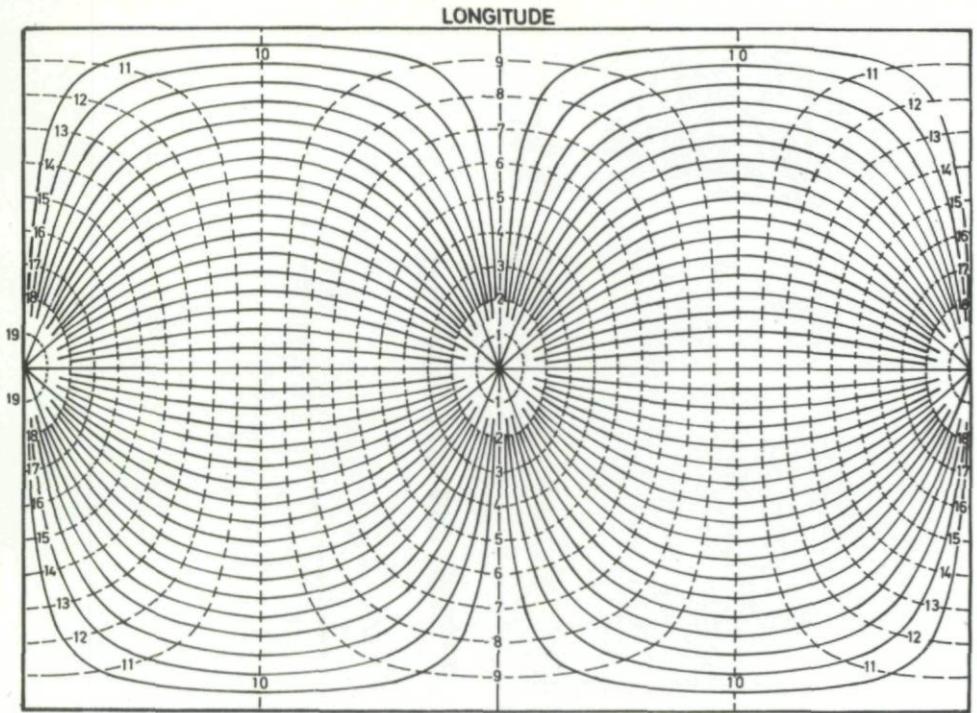


Figure 2.8 World map and great circle chart centered on equator for modified cylindrical projection  
 (After K. Davies, Ionospheric Radio Propagation, National Bureau of Standards, Monograph 80, 1965)

LATITUDE DEGREES

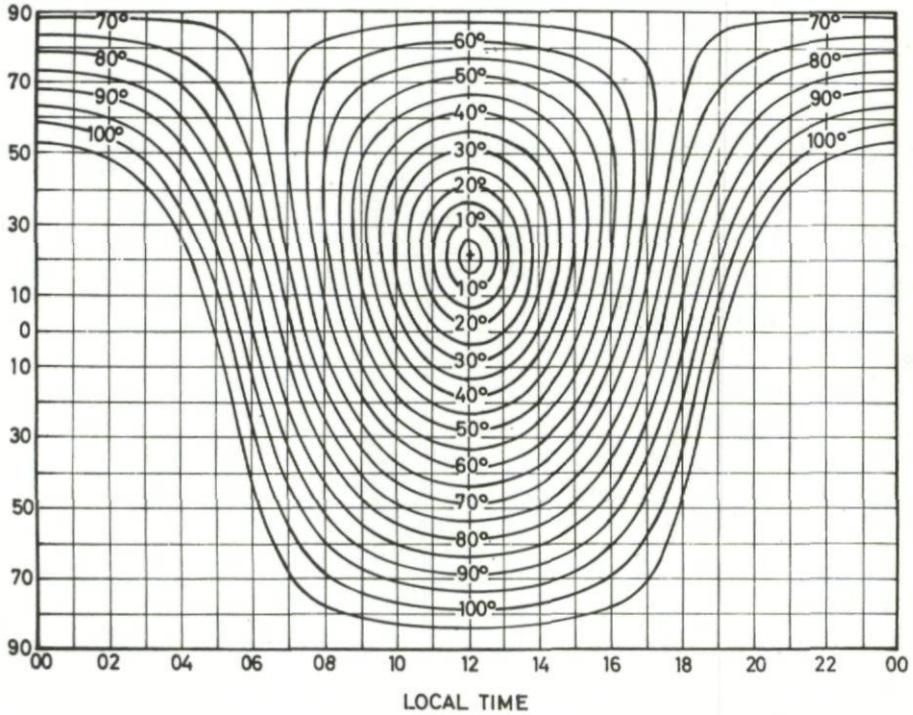


Figure 2.9 Values of the solar zenith angle for July  
(After S.M. Ostrow, Handbook for CRPL Ionospheric Predictions, National Bureau of Standards, Handbook 90, 1962).

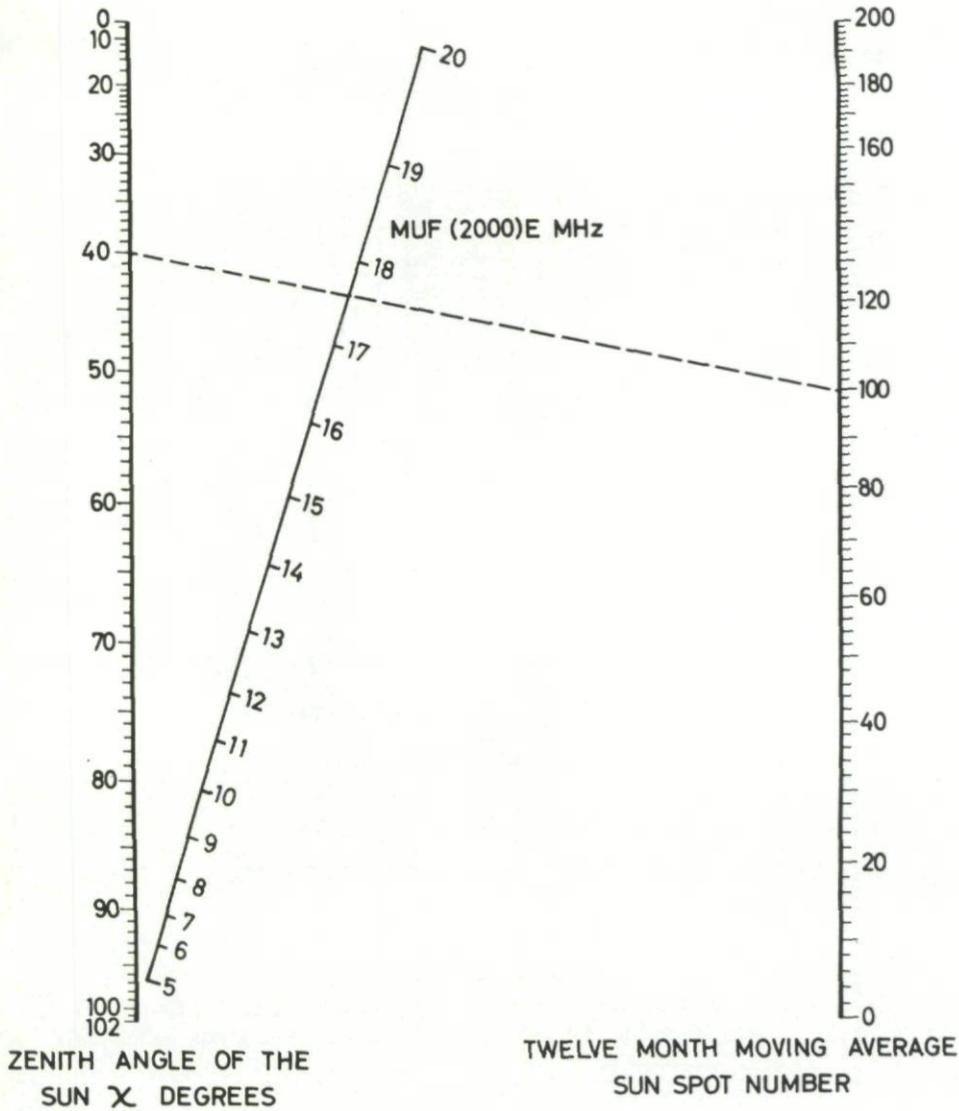


Figure 2.10 Nomogram for determining MUF(2000)E  
 (After S.M. Ostrow, Handbook for CRPL Ionospheric  
 Predictions, National Bureau of Standards, Handbook 90,  
 1962)  
 Example: When solar zenith angle is  $40^{\circ}$  and sun spot number is  
 100, MUF(2000)E is 17,6 MHz

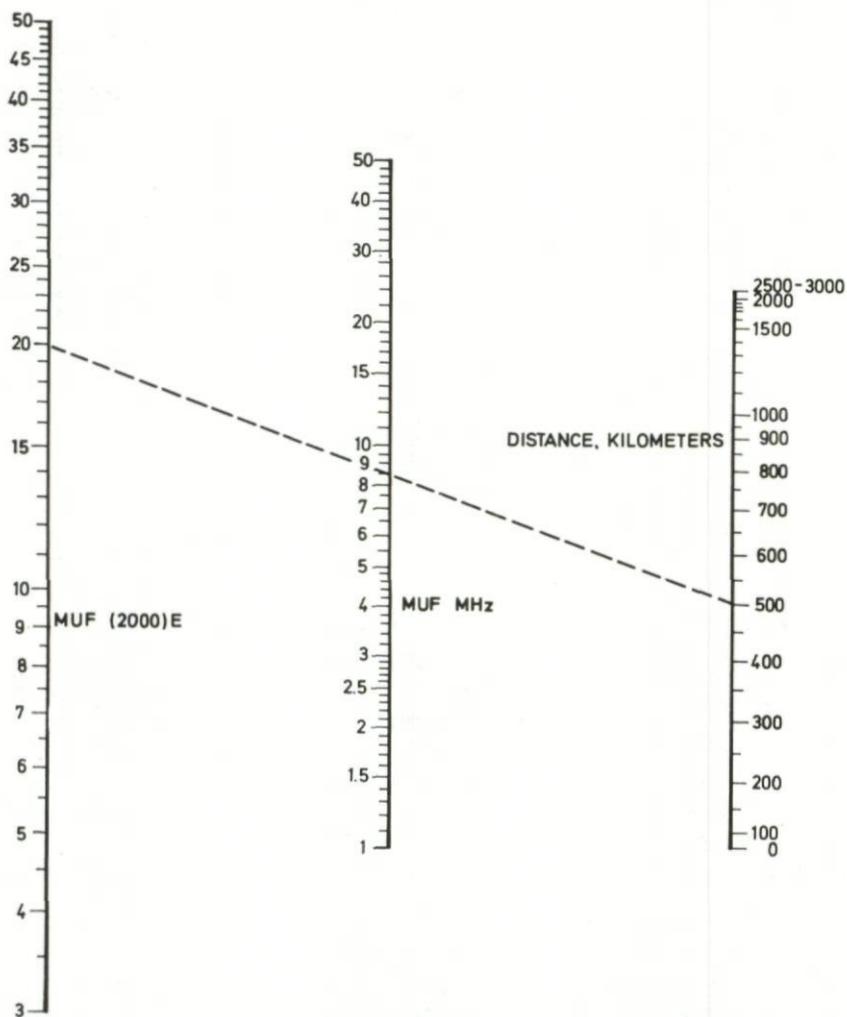


Figure 2.11 Nomogram for conversion of MUF(2000)E into combined E- and F1-layer MUF for distances between 0 and 3000 km

(After S.M. Ostrow, Handbook for CRPL Ionospheric Predictions, National Bureau of Standards, Handbook 90, 1962)

Example: When distance is 500 km and MUF(2000)E is 20 MHz, combined E- and F1-layer MUF is 8,4 MHz

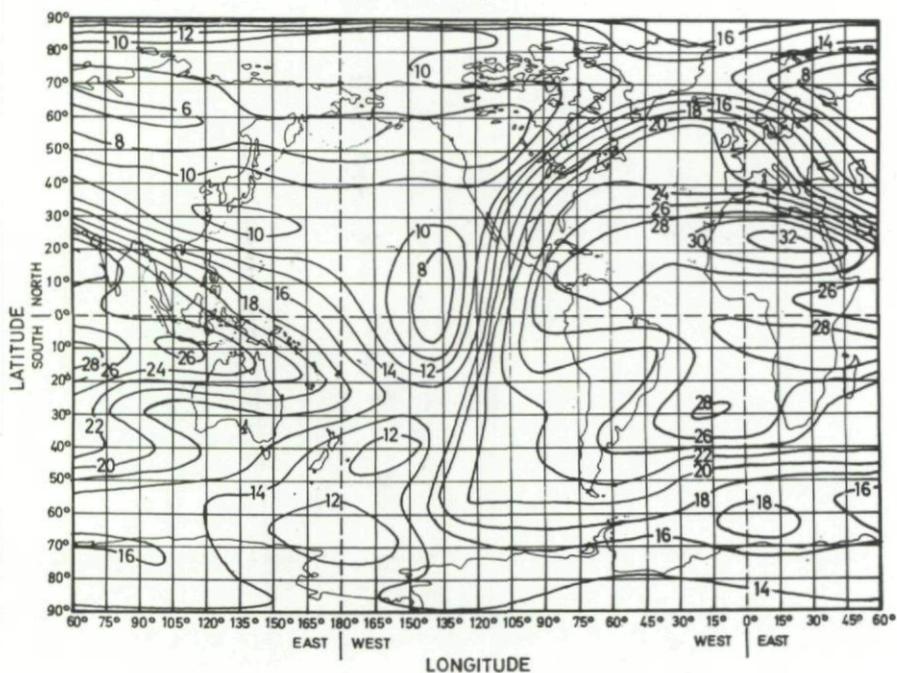
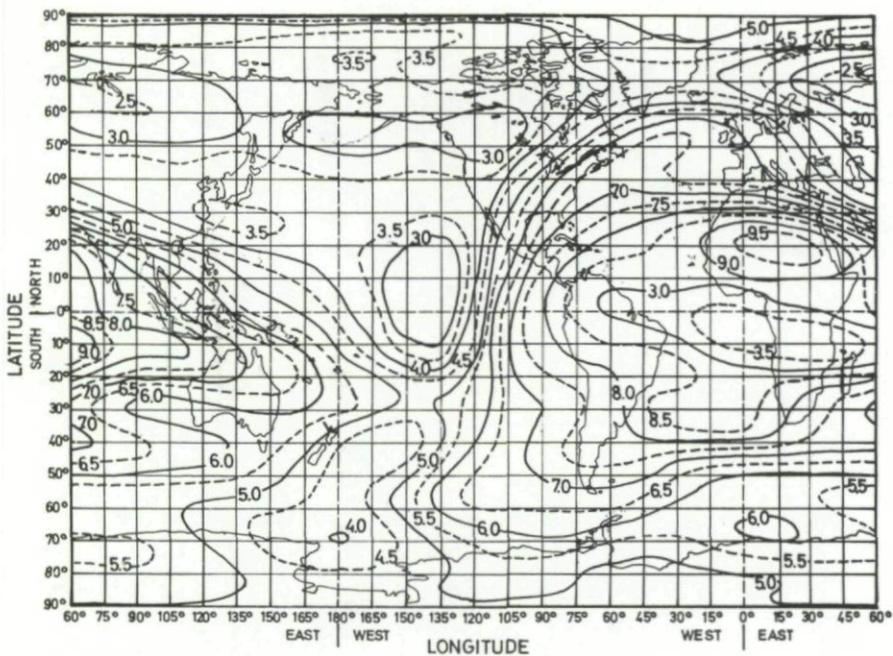


Figure 2. 12 Maps giving contours of MUF(0)F2 and MUF(4000)F2 for 1400 GMT December 1963

as either:

$$\text{FOT} = \text{MUF-E (or MUF-F1)} \quad (\text{Eq. 2. 7})$$

or:

$$\text{FOT} = 0.85 \cdot \text{MUF-F2} \quad (\text{Eq. 2. 8})$$

The spread of the foF2 values (and therefore the MUF-F2 values), is such that about 10% of the hourly observed values are smaller than 0.85 times the median value. This means that communication at the optimum traffic frequency will be possible during about 90% of the time.

The lowest useful high frequency LUHF

For most services, the LUHF is not predicted. The quantity is of particular importance when planning the equipment parameters for a circuit. The available data for parameters, like the noise at the receiving site and the various causes of power losses along the path are not well known and the uncertainty in a possible prediction of the LUHF is therefore high.

Most of the parameters involved in the determination of the LUHF are not quantities that change from year to year and therefore need to be predicted. This is true for example, for the noise level which only depends on geographical location, season, and local time.

The quantities that have to be predicted are the losses along the path. On page 34 we saw how the free space attenuation  $A_s$  and ground losses  $A_g$  can be estimated. Allowance for fading must be made, using values of  $A_f$  for different types of services listed by CCIR, (CCIR (1959), Documents of the ninth Plenary Assembly, Los Angeles, Recommendation 164). The way in which the ionospheric absorption  $L$  depends upon the solar zenith angle  $\chi$ , the angle of elevation  $\alpha$ , and the sun spot number  $R$  was also discussed.

The angle of elevation  $\alpha$  for a particular mode can be determined in terms of the great circle distance  $d$  and the height of the reflecting layer. If an F-layer mode is considered, the M(4000)F2 may be used instead of the F2-layer height. In Figure 2. 14 we have shown how the elevation angle  $\alpha$  is related to  $d$  and M(4000)F2, the latter being a reasonably accurate measure of effective layer height. M(4000)F2 is defined as the ratio of MUF(4000)F2 and foF2. MUF(4000)F2 has already been determined in connection with calculation of the MUF, foF2 may be determined by subtracting 0.7 MHz from MUF(zero)F2, equation 2. 1.

Knowing the values of  $\chi$ ,  $R$  and  $\alpha$ ,  $L$  can be determined by means of two nomograms. From the first, Fig. 2. 15, an absorption index  $I$  is determined in terms of the solar zenith angle  $\chi$  and the sun spot number  $R$ . From the second nomogram, Fig. 2. 16, the absorption  $L$  in decibels is determined in terms of the radiation angle  $\alpha$ , the absorption index  $I$ , and the effective wave frequency,  $f + f_H$ .

Having determined the total path loss and the field strength at the receiver as a function of frequency, the next step is to find the frequency variation of the intensity of radio noise at the receiving site. This can be done by the methods described on page 45. World wide noise data for different seasons and times of day have been published in CCIR Report No. 322, and the minimum required signal to noise ratio for different kinds of services have been agreed on internationally and are listed by CCIR. (CCIR (1959), Documents of the ninth Plenary Assembly, Los Angeles, Recommendation 161). The LUHF can then be found by comparing the minimum required field strength versus frequency with the frequency variation of predicted received signal strength. The E-layer cut-off frequency must be taken into account in this procedure.

## Examples of selected circuits

Under this heading we will discuss some examples of predictions in order to illustrate how the predictions are presented and how to interpret the predicted parameters.

**Oslo—Rome (1950 km)** — The predictions for the month June 1965 for the circuit Oslo—Rome are shown in Figure 2.17. In the illustration we have shown the predicted MUF-E, FOT-F2,  $f_sE$ , and resultant FOT. The resultant FOT is the higher of FOT-F2 and MUF-E, and the illustration shows how the E-region takes control in the hours 04-18 GMT. This is typical for short distance circuits during the summer.

The LUHF has not been calculated. The parameter  $f_sE$  in this prediction gives the effect of screening of F-transmission by the normal E-layer. This layer is solar controlled in much the same way as the ionospheric absorption.  $f_sE$  can therefore be used as a rough estimate of LUHF.

In Figure 2.18 only FOT and  $f_sE$  are given for the circuit Oslo—Rome. These parameters define the frequency range useful for the circuit. FOT giving the upper frequency limit, and  $f_sE$  the lower.

The radio operator should select a working frequency as high as possible between these frequency limits. In Figure 2.18 we have shown an example of selection of working frequencies. Frequency bands from which the working frequencies are selected, are allocated to the different types of radio services. The working frequencies must therefore be changed in steps. For example the ship to shore services have their frequencies in the 4, 6, 8, 12, 16, 22 and 24 MHz bands. The illustration, Fig. 2.18, shows how a 24-hour service is made possible using three working frequencies. For this particular circuit possibly only the lower two of the three recommended frequencies are necessary. The highest frequency is E-layer controlled, and since the path length is 1950 km, the circuit is just on the limit for a one hop E-layer mode.

As a rule it can be said that an operator should stick to a frequency as long as good signals are coming through, and not change the frequency even if the predictions indicate a change.

The predictions for the same circuit, Oslo—Rome, for December 1965, are shown in Figure 2.19. Comparing this graph with Figure 2.17 we note that the MUF-E is lower than in summer, but the MUF-F2 during the sunlight hours is considerably increased relative to the summer values. Since the highest possible working frequency should be chosen, the F2-layer mode will offer the best possibilities for communication during December.

That the noon value of foF2 and therefore MUF-F2 is higher in winter than in summer was discussed in chapter one. For the winter month, Fig. 2.19, both the cut-off frequency  $f_sE$  and the LUHF have been calculated.

It is clear that in this particular case the LUHF is at all times somewhat higher than the  $f_sE$ . Thus the lower frequency limit is not determined by  $f_sE$ , because the absorption is too strong near the cut-off frequency for communication to be possible. On the other hand, LUHF and  $f_sE$  are not very different, and Figure 2.19 clearly demonstrates that the two parameters vary in the same way with the solar zenith angle. It is therefore not unreasonable to use the  $f_sE$  as a rough estimate of the LUHF.

**Oslo—Johannesburg (10 000 km)** — In Figure 2.20, the predictions for the circuit Oslo—Johannesburg for the month October 1965 are shown. The parameter MUF is determined as the highest frequency that can be received at either site by the F2-layer or the E-layer. In the case shown, FOT is controlled by the F2-layer for the complete 24 hours and is therefore  $0.85 \cdot \text{MUF-F2}$ .

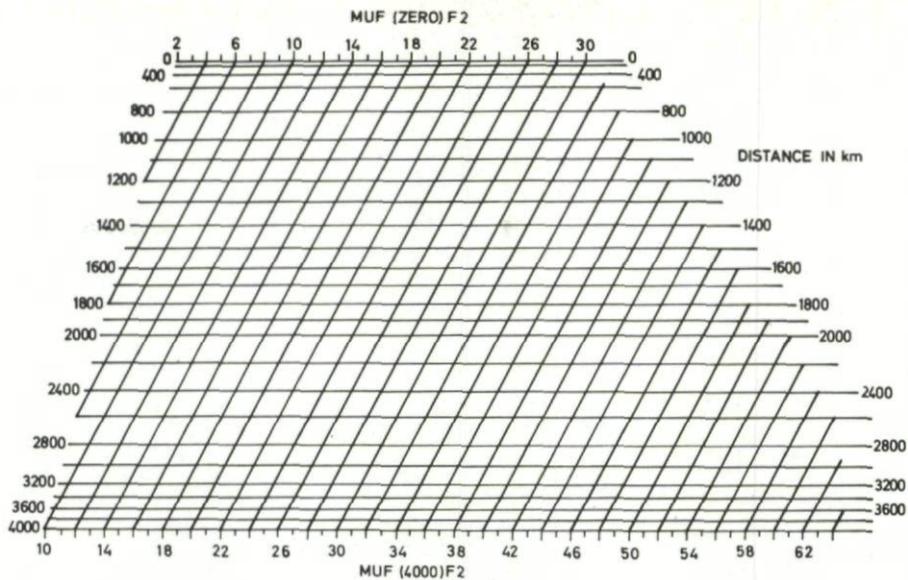


Figure 2.13 Nomogram for transforming MUF(zero)F2 and MUF(4000)F2 to equivalent MUF at intermediate transmission distances

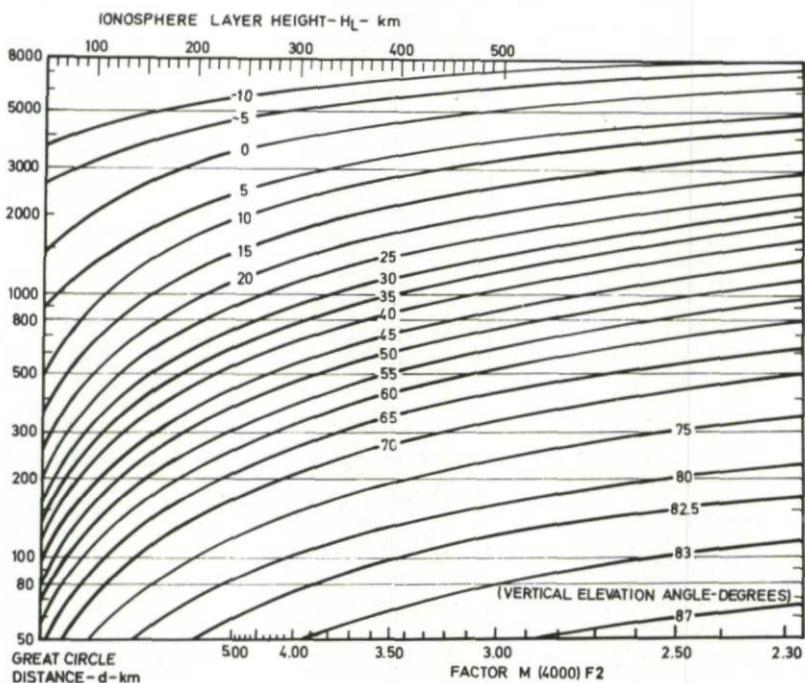


Figure 2.14 Determination of elevation angle  $\alpha$  in terms of the great circle distance  $d$  and the factor  $M(4000)F2$  (or the layer height)  
 (After K. Davies, Ionospheric Radio Propagation, National Bureau of Standards, Monograph 80, 1965)

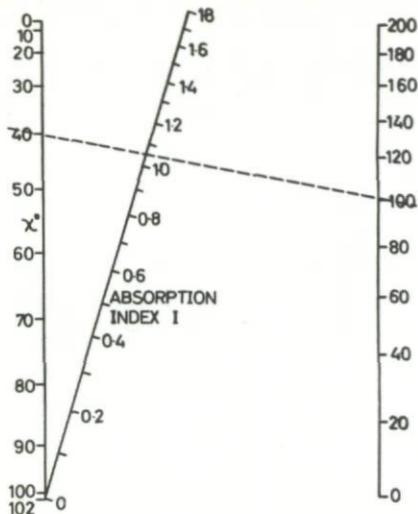


Figure 2.15 Nomogram giving the absorption index  $I$  in terms of solar zenith angle  $\chi$  and sun spot number  $R$

(After K. Davies, Ionospheric Radio Propagation, National Bureau of Standards, Monograph 80, 1965)

Example: When solar zenith angle  $\chi$  is  $40^\circ$  and sun spot number  $R$  is 100, the absorption index  $I$  is 1.06

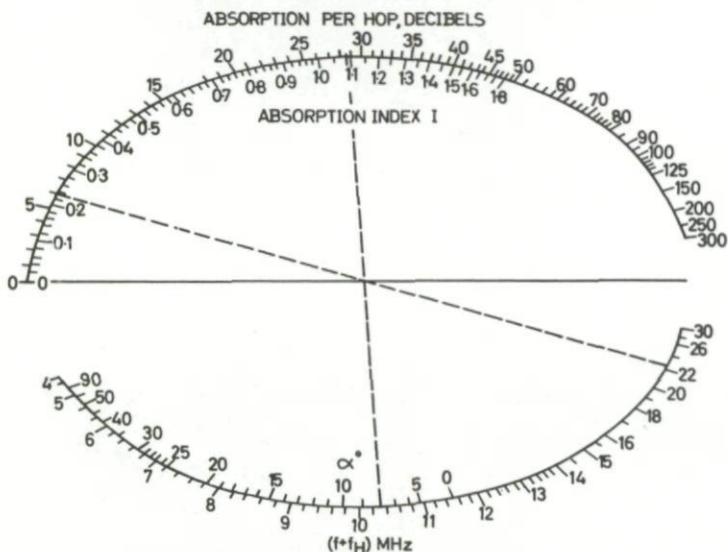


Figure 2.16 Nomogram giving the absorption  $L$  in decibels in terms of the absorption index  $I$ , the elevation angle  $\chi$ , and the effective wave frequency  $f + f_H$ . A value of 1.4 MHz is assumed for the gyro frequency  $f_H$

(After K. Davies, Ionospheric Radio Propagation, National Bureau of Standards, Monograph 80, 1965)

Example: (1) Enter with the absorption index  $I = 1.09$  and the elevation angle  $\chi = 8^\circ$ . (2) Mark the reference point of intersection with the center line. (3) With reference point and effective frequency  $f + f_H = 22.4$  MHz draw a straight line as far as the curve marked absorption per hop (= 6 dB)

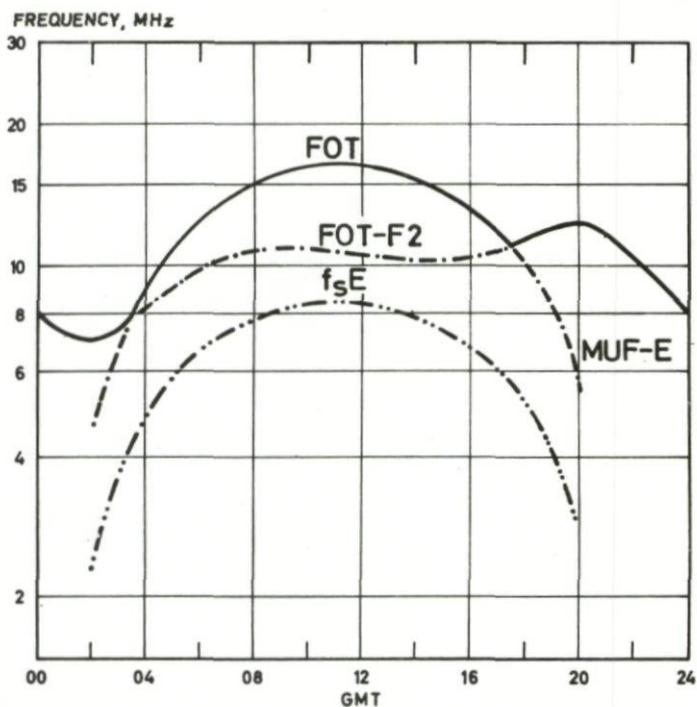


Figure 2.17 Predictions for the circuit Oslo - Rome June 1965

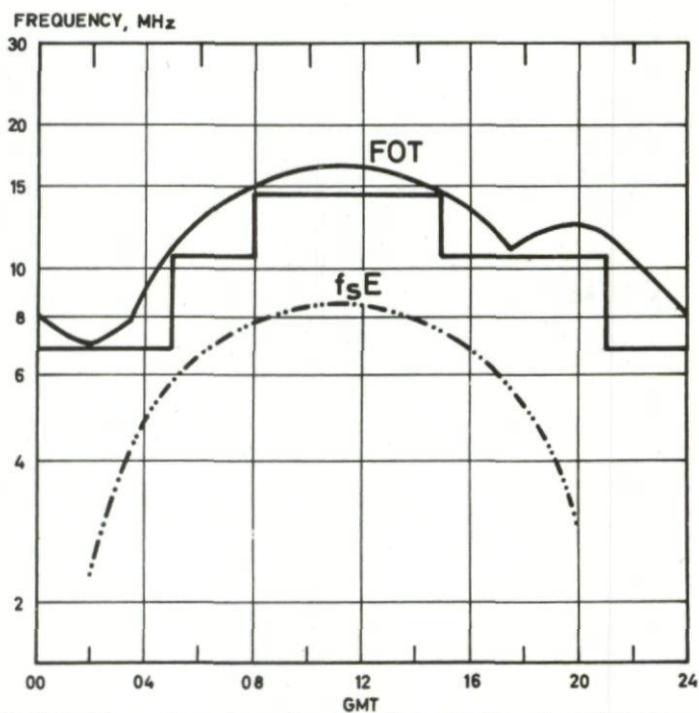


Figure 2.18 Selection of working frequencies for the circuit Oslo - Rome June 1965

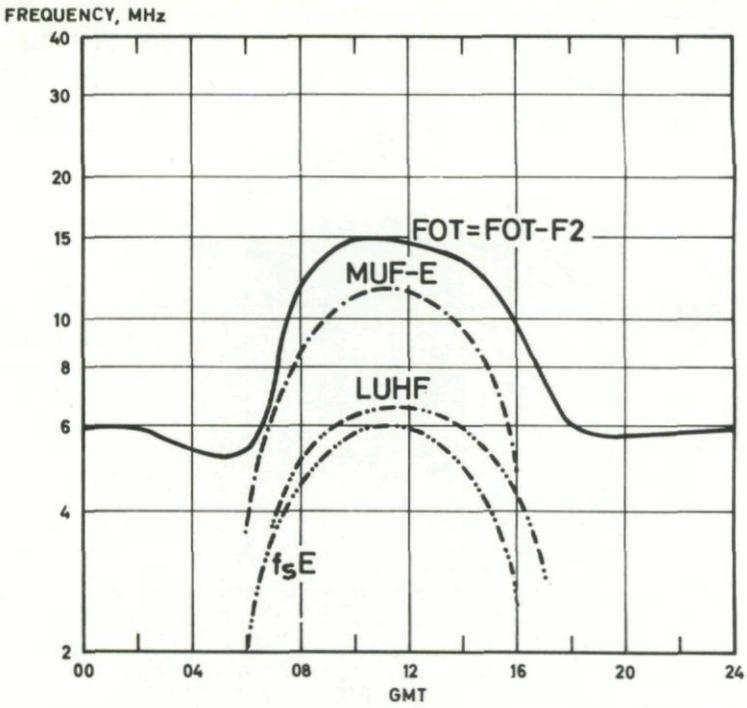


Figure 2.19 Predictions for the circuit Oslo - Rome December 1965

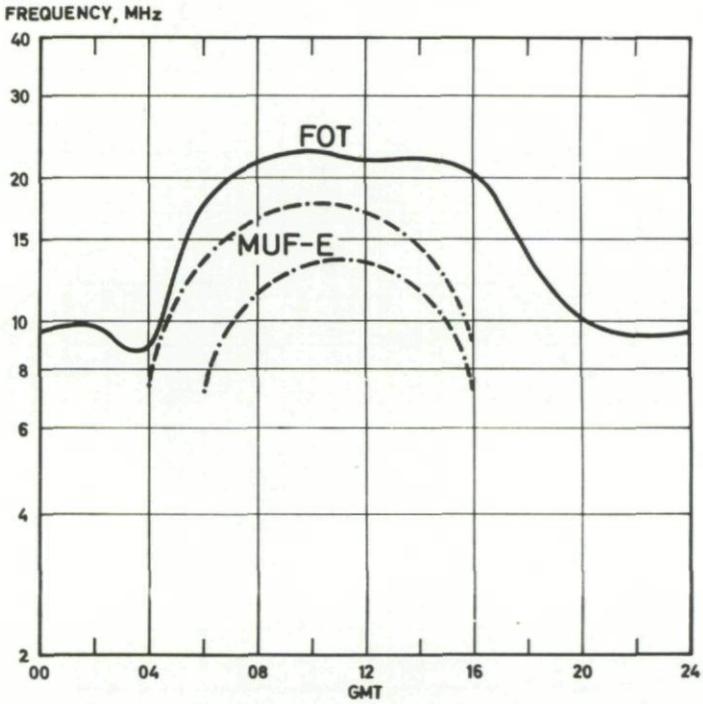


Figure 2.20 Predictions for the circuit Oslo - Johannesburg October 1965

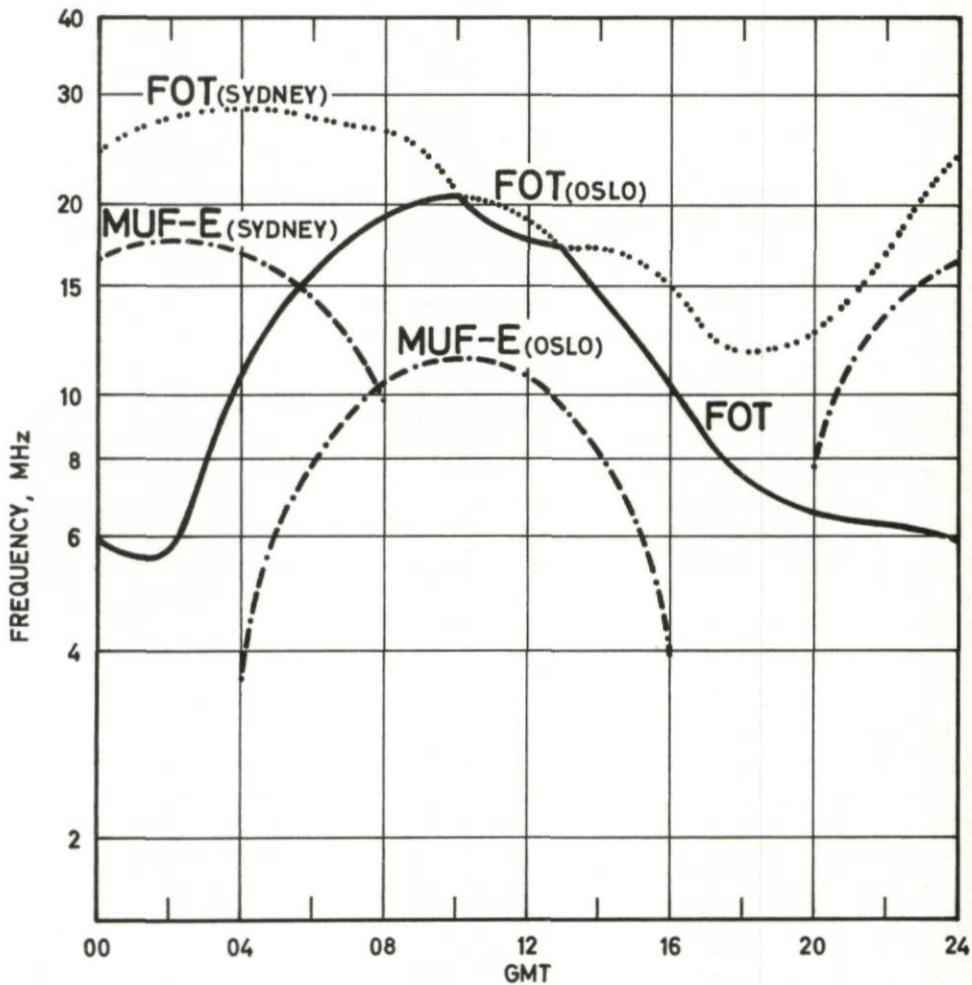


Figure 2.21 Predictions for the circuit Oslo - Sydney October 1965

In the morning hours the curves for FOT and MUF-E are very close together. A working frequency below FOT will also probably be below MUF-E, and give reflection from the E-layer. This will increase the number of hops compared to the F-layer mode, and therefore also increase the absorption. We conclude that communication is likely to be difficult at this time of the day.

The maximum frequencies that would be reflected from the E-layer at the two control points 1000 km from Oslo and from Johannesburg (MUF-E) are also shown. These frequencies will limit the possibilities of communication via F-layer modes, and they are approximately equal to the screening frequency  $f_s E$  discussed in the previous example. It appears that the one representing the MUF-E near Johannesburg, where the elevation of the sun is higher than near Oslo, must give a higher MUF. The shift between the time of occurrence of the maxima corresponds to the difference in longitude of the two control points.

**Oslo—Sydney (16 000 km)** — The predictions for the circuit Oslo—Sydney for October 1965 are shown in Figure 2. 21. Figure 2. 21 shows the variation of MUF-E and  $0.85 \cdot \text{MUF-F2}$  for the control points near Sydney and Oslo. Due to the difference in geographical location of the two stations (Sydney  $34^\circ\text{S}$ ,  $151^\circ\text{E}$ —Oslo  $60^\circ\text{N}$ ,  $11^\circ\text{E}$ ) the season and the solar zenith angle are different at the two control points. This explains why not only the magnitude of the MUF-E and MUF-F2, but also the shape of the curves change from one station to the other. The FOT for the circuit as a whole must be the lowest of the FOTs for Sydney and for Oslo. We see that the propagation conditions near Oslo determine the choice of frequency, except for the period 10-13 GMT where the FOT for Sydney is slightly below that for Oslo.

In the period 20-03 GMT we note that the FOT for Oslo is very low, we have, in fact, night-time conditions near Oslo, but day-time conditions near Sydney. This means that a frequency below 6 MHz must be used in order to get an F-layer reflection at the control point near Oslo. A frequency as low as this will probably be strongly absorbed near the other end of the path and communication may therefore be difficult.

### Reliability and use of predictions

It has already been stressed in preceding sections that, due to the variability in the ionospheric parameters, it is impossible to produce predictions that are reliable in all cases. It is the average behaviour of a circuit for a certain month that is normally predicted, and the prediction thus does not take into account the changes from one day to the other. Our knowledge of the ionospheric behaviour is not yet sufficiently complete for us to be able to predict the day to day changes.

Predictions of disturbances of the type occurring in the auroral zone are being made and will be discussed in later chapters of the book. Research work is also being done at a number of places in order to find methods and criteria that could be used in a day to day prediction service of the 'undisturbed' ionosphere, and it is quite possible that a service of this type will be in practical use within a few years. In this section we shall be concerned with the reliability of the present system of predictions, and discuss how the best use can be made of this.

### Testing of predictions

There are a number of methods that have been used in order to test the reliability and usefulness of radio frequency predictions. Some of the most important are listed below:

- (a) Wireless operators have been supplied with predictions for their circuits, and have been asked to send in report forms, which have been studied and analysed.
- (b) Operators have been asked to work on a number of frequencies and to report on the usability of these. The results from such reports are then compared with the predictions.
- (c) Continuous recordings of the field strength on a number of different frequencies have been made, and the results have been compared with those to be expected from predictions.
- (d) Results from oblique incidence and backscatter soundings have been compared with predictions. From these techniques, the MUF for the circuit may be determined directly, and the techniques are therefore well suited for the purpose of testing the reliability of the predictions.

The results from the different tests described above are not clear cut or internally consistent. The circuit operators (test a) normally state that they find the predictions of considerable use, but there are cases when the opposite statement is given. A detailed analysis of the data sheets from test (b), however, have shown that in many cases, when the predicted frequency does not give useful communication, one can find the reason for this. A better understanding of why the predictions fail may help the operator to benefit more from the predictions.

The methods (c) and (d) have been most extensively used in order to study *communication difficulties in polar regions*. In these regions the normal predictions are not very reliable. We shall discuss the polar problems in greater detail in Chapters 3 and 4. When the methods have been used at low and middle latitudes, one has generally found reasonably good agreement between the results and the predictions.

*Note: The oblique incidence ionospheric sounding technique is perhaps the most promising method upon which a short term prediction service can be based. If a large number of circuits are set up around the world, the results from these would provide one with an instantaneous picture of the ionosphere. Some centralized organization would be needed to receive the data from all the circuits, analyse this information, and transmit the predictions in a suitable form.*

Some reasons why the predictions may fail

Some of the most important reasons why the predictions may fail will be discussed in the following paragraphs.

(a) **The actual MUF is lower than the predicted FOT** — As was explained on page 58 the Optimum Working Frequency, FOT, is defined as 0.85 times the predicted MUF, for communication controlled by the F2-layer. This was in order to allow for the observed spread in the ionospheric parameters, mainly foF2. The factor of 0.85 is chosen as a compromise. On the one hand one wants the factor as high as possible in order to minimize the effect of ionospheric absorption, and on the other hand a low factor will satisfactorily take into account the great spread of the ionospheric parameters. About 10% of the observed foF2 values are smaller than 0.85 times the median value. This means, therefore, that in about 10% of the cases one will have the condition that the actual MUF is lower than the predicted FOT.

(b) **There is no possible F-region mode because of screening by the E-layer** — The effect of E-layer screening, see pages 45 and 58 is normally taken into account when preparing the predictions, and the E-layer is so stable that this effect can be predicted with fairly high accuracy. The screening effect of the sporadic E-layer can also be of importance. This effect is however difficult to predict.

(c) **The absorption is too high for a useful circuit to be obtainable** — It is very difficult to predict the LUHF, mainly because of lack of knowledge of the ionospheric absorption. The absorption is known to have considerable day to day variations.

During Sudden Ionospheric Disturbances, SID, see page 28 the LUHF may be greatly increased during shorter periods for all transmission paths on the sun-lit side of the earth.

A particular type of absorption is the winter-anomaly absorption, see page It is found generally that during one out of ten days on average, the ionospheric absorption is significantly higher than during the normal days. The winter-anomaly is observed at middle latitudes. Fortunately, the winter-anomaly is normally not serious for HF communication because the foF2 is very high in winter at middle latitudes. Thus, the increase in absorption on a fixed frequency can be compensated for by selecting a high working frequency.

How to select frequency when predictions fail

The general rule is that one wants to make use of as high a frequency as possible. By doing so one minimizes the effects of ionospheric absorption which decreases with the inverse square of the frequency. It is very difficult to formulate rules and advice about how one should tackle the problems when the predictions do not give useful communication.

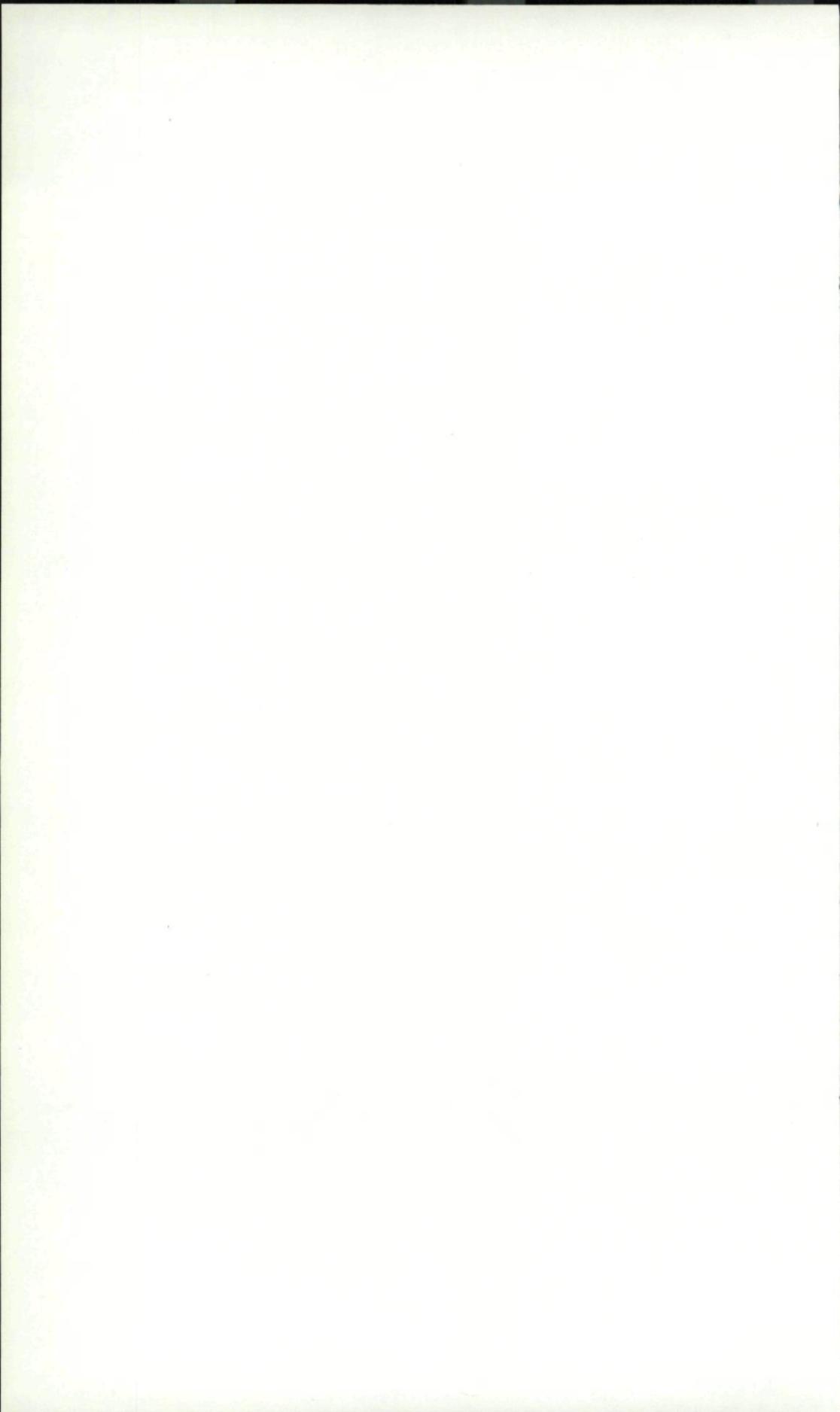
If a weak signal can be heard from the station one wants to communicate with, but this is too weak to be useful, the reason is most probably that the absorption is too high and use should be made of a higher frequency. If no signals can be observed on the next higher available frequency, communication is probably impossible. At mid-day and for fairly short distances there may be a possibility of sporadic E communication, and this may be tried by selecting an even higher frequency. If it is still impossible to obtain contact, and it is important to establish this as soon as possible, one should choose the frequency through which MUF is expected to pass first, and work with that. In the morning hour, when the MUF is increasing, one should therefore concentrate on the next available frequency above that at which the weak signal could be heard. In the afternoon (after about 2.00 p.m.) when the MUF is decreasing one should continue to work on the frequency at which the weak signal could be observed. By selecting a frequency as close to the MUF as possible, one does not only minimize the effects of ionospheric absorption, but one will also have an increase in the signal strength due to the focusing effect that occurs near the MUF, see page 29.

If no signals can be received on the predicted frequency, the reason may be any of the three given on page 75. If this occurs during night-time, neither screening by an E-layer nor strong absorption are probable reasons (except in polar regions), and it is quite probable that the actual MUF is lower than the predicted FOT. One should then try the next available lower frequency.

If no signals can be received on the predicted frequency during day-time, the same procedure should still be used as a first try. If this does not help, a higher frequency should be tried as both (b) and (c) may be possible reasons.

At fairly short distances the possibility of sporadic E propagation should also be tried.

The above rules can only be regarded as suggestions for intelligent use of the predictions. It is impossible to give general prescriptions that hold in all cases. We hope, however, that knowledge of the way in which the waves may propagate between the terminals, will help the operator to make the best possible use of his equipment.



## Communication Problems in Polar Regions

### Introduction

We shall turn now to the particular problems encountered in HF radio communications in polar regions. In the previous chapters we treated the problems as we meet them at low and middle latitudes.

We have seen that there is a considerable variation in the parameters of importance for HF radio communication from one day to the other even at low and middle latitudes. In particular, the variations of the maximum electron density in the F2-layer is important. During magnetic storms, a depression of the critical frequency for the F2-layer is normally observed, that affects HF communication possibilities. In polar regions the effects of the disturbances on the lower parts of the ionosphere are even more important.

On page 16 we discussed how free electrons are produced and how they are lost. The two sources of importance for ion production are: (i) electromagnetic radiation from the sun and (ii) energetic particles from the sun or galaxies. While most of the phenomena observed at low and middle latitudes can be properly explained in terms of ionization by electromagnetic radiation, this is not so in the polar ionosphere. In these regions, the ionization produced by the energetic particles is perhaps more important than that due to the electromagnetic radiation.

The stream of solar particles is highly irregular and varies considerably from day to day. While only major solar disturbances have appreciable effects on the ionosphere at low and middle latitudes, even small disturbances may affect the polar ionosphere. Here, disturbances of considerable importance for radio communication are quite frequent. The predictions discussed in Chapter 2 do not take into account these disturbances. These predictions are based upon our general knowledge about the behaviour of the ionospheric layers and their time variations, which has been gained mainly from a world wide network of ionosonde stations.

There has been an extensive research activity in the years after World War II in order to study the phenomena associated with the communication difficulties in the arctic, and in the present chapter we shall discuss some of the methods that have been used together with the most important results which have been derived.

### The polar ionosphere

The purpose of the present section is to introduce briefly the main features of the polar ionosphere, of importance for HF radio communication. The methods of studying these, together with a more detailed description of the most important effects will then be given in the later sections.

### Origin of particles

The energetic particles of importance for the various effects on the polar ionosphere are of solar origin. They are associated with disturbances on the sun such as flares and sun spots. In Figure 3. 1 we have illustrated some of the most important effects a flare on the sun may have on the ionosphere. About 8 minutes after the occurrence of the solar flare an increase in the D-region ionization may be observed. This effect, which is called a Sudden Ionospheric Disturbance (SID), was discussed briefly on page 28. This ionization may cause a black-out of all HF circuits on the sun-lit part of the earth.

After a period that may vary from fifteen minutes to several hours, an increase in the particle flux is generally observed by cosmic ray observatories. The observations can be made either by ground based or balloon-borne instruments. The increase in the particle flux is due to high energy particles from the sun (protons and  $\alpha$ -particles).

From the communicator's point of view the increase in the D-region ionization associated with the particles is of particular importance. The radio wave absorption due to this ionization is called polar cap absorption. It will be discussed in some detail on page 93.

After a period that may vary from some hours to some days after the occurrence of the flare, various effects such as magnetic storms, visual aurorae, strong sporadic E-layers and increased D-region ionization may be observed.

While the first two types of effects are normally associated with distinct flares occurring on the solar disc, this is not always true for the magnetic storms and their associated phenomena. Magnetic storms are much more frequent and may be caused by less distinct disturbances on the sun.

#### Propagation of particles from sun to earth

The way in which the particles propagate from the sun to the earth, and how they are brought into the earth's atmosphere is being intensively studied by means of satellites, and still offers a number of unsolved and intriguing problems.

The irregular, continuous stream of charged particles leaving the sun's surface is called the solar wind, and the various effects observed in the polar ionosphere are due to an interaction of the solar wind with the boundary of the earth's magnetic field.

Note: The first rigorous approach to a solution of the problem of solar particle motion near the earth was made by the Norwegian physicist Carl Störmer. He showed theoretically how electrically charged particles of high energy traveling towards the earth are guided by the geomagnetic field lines towards the polar regions. He also showed that the particles can be trapped by the geomagnetic field. The existence of geomagnetically trapped particles was discovered by Van Allen in his rocket and satellite studies. The way in which the particles are brought into the earth's atmosphere is complex and not fully understood.

Recent satellite studies have shown that the geomagnetic field is distorted by the solar wind, Fig. 3. 2. The field lines on the side towards the sun are compressed, while on the side away from the sun they are extended to form a tail. This magnetospheric tail is of considerable importance.

#### Interaction of particles with the earth's atmosphere

Most of the charged particles interacting with the earth's magnetic field, will be guided into the atmosphere at geomagnetic latitudes corresponding to the auroral zones. When the energetic particles penetrate the atmosphere they may collide with the air atoms and molecules. In this way the aurora is produced, see page 16, and as we have already seen, the particles may produce layers of ionization of importance for radio communication.

Note: The interaction of high energy particles with the atmosphere can cause excitation or ionization of the gas atoms and molecules, see page 16. Most of the excited states of the atoms are unstable, and the electrons will rapidly fall back into their natural orbits. The free electrons will recombine with positive ions to form neutral molecules or atoms. In these two processes the electrons will lose their excess energy as electromagnetic radiation, and this radiation will often lie in the visible frequency band. An intense stream of particles impinging on the atmosphere can therefore cause the visual phenomenon known as the aurora. The different colours in the aurora are caused by excitation of different kinds of atoms and molecules. The more energetic the particle, the deeper it penetrates into the earth's atmosphere before it is stopped.

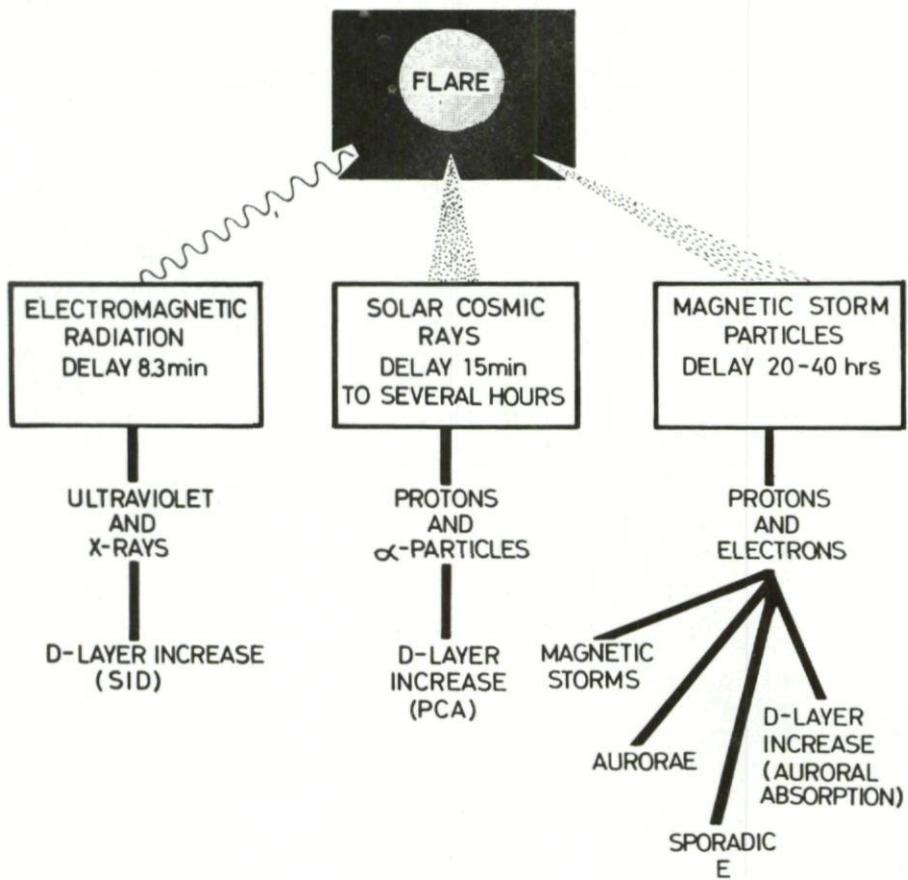


Figure 3.1 Effects on the polar ionosphere due to a flare on the sun

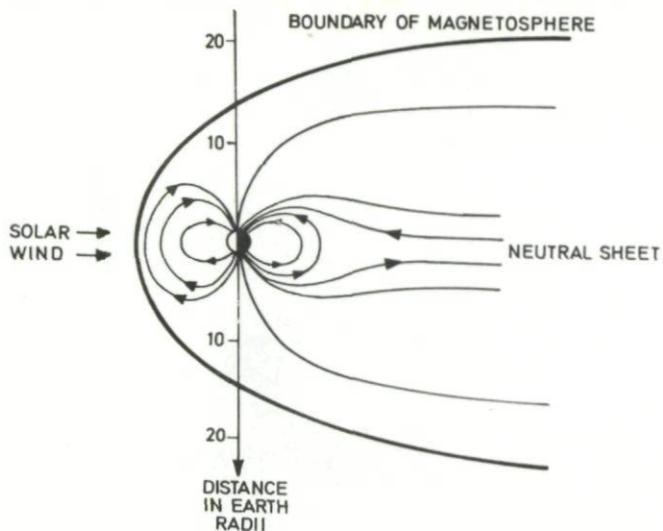


Figure 3.2 The geomagnetic field as measured by satellites (idealized)

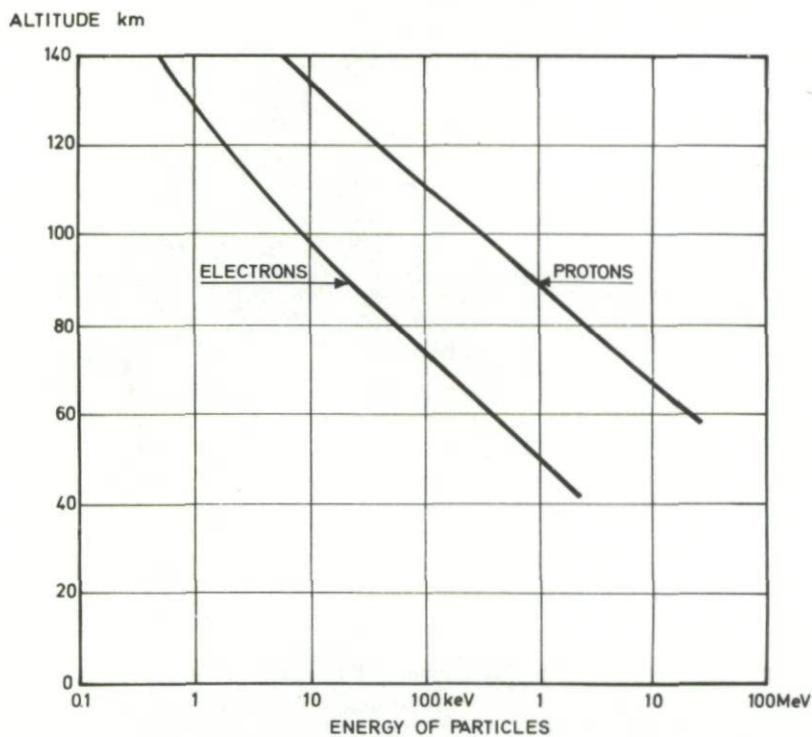


Figure 3.3 Penetration depth of electrons and protons in the earth's atmosphere

In Figure 3.3 we have shown the penetration depths of electrons and protons in the earth's atmosphere. The quantity plotted against altitude is the approximate height where the particles are stopped if their path is vertical. Until recently the effects of the particles on the F-region were not well known because often, during disturbances, the ionospheric absorption becomes so high that the F-region traces on ionograms cannot be observed. From results obtained in recent years from topside soundings, see page 19, we have gained new important information on the disturbed F-region. In fact the importance of ionization of the polar F-region by particles is now well established, and a trough in the total ionization of the F-region is normally observed in a transition region between the middle latitude F-region ionized mainly by ultraviolet radiation, and the polar F-region, ionized mainly by particles.

The ionization in the F-region is naturally of importance for HF radio communication, but for the particular problems one has to deal with in polar regions, the ionization at the lower altitudes is even more important.

The polar absorption is caused by protons in the energy range 1 to 20 MeV and possibly to some extent by  $\alpha$ -particles in the same energy range.

The absorption due to the increased D-region ionization is normally called auroral absorption. It is caused mainly by electrons in the energy range 50 to 500 keV.

The sporadic E-ionization occurs in the same height range as the visual aurorae. Both effects are probably caused by protons in the energy range 50 to 200 keV and electrons in the range 5 to 30 keV.

### Methods of study

Some of the most important experimental techniques used to study the normal ionospheric layers at low and middle latitudes were mentioned on page 19. Even if most of these techniques are very useful in the arctic, special methods are often needed to study the high latitude ionosphere, particularly during disturbances. Under such conditions the electron densities can be greatly enhanced in the lower part of the ionosphere which causes complete blackouts on the ionogram. Thus the ionosonde is of no use. Furthermore, due to the sudden rapid irregular variations there is also a need for techniques that provide continuous automatic recordings. The variations observed are irregular, not only in time but in space and a dense net of stations is important to map the geographical extent of the phenomena.

In following paragraphs some experimental methods, which are particularly useful for studies of the high latitude ionosphere, will be described.

### Vertical incidence and oblique soundings

The principle of vertical incidence sounding has been explained on page 19. The ionosonde is still the most useful tool designed to observe the ionosphere. A good quality ionogram, when properly interpreted can give a lot of detailed information on the structure of the ionization. The interpretation of a high latitude ionogram often presents special problems because of the complexity of the records, Figure 1.17. Not only is the vertical electron density distribution often more complex than at lower latitudes, but horizontal gradients in the electron density can occur, which cause irregular traces on the ionogram. Thus the sounding pulses can be reflected from irregularities in electron density not directly above the station. The movements and speeds of such irregularities may sometimes be derived from ionograms.

The principle of ionospheric sounding by means of high frequency radio pulses may also be applied to oblique propagation. An oblique incidence sounder may consist of a transmitter and a receiver at one place plus a receiver at a distant location. In Figure 3.4 a simple situation is sketched which illustrates this. Normally, antennas that have maximum gain at small angles of elevation are used. The frequency of the transmitter and the receivers will be changed simultaneously as on a conventional ionosonde, but the oblique sounder has a wider frequency range. For example an oblique incidence sounder may transmit 50 pulses per second, each pulse lasting for one millisecond, and the frequency of these pulses may be changed in steps from 4 MHz to 64 MHz in the course of about one minute.

The remote receiver records the arrival of pulses which have travelled from the transmitter T via the ionosphere to  $R_2$ . The travelling time for the pulses will vary with frequency, since different frequencies will penetrate to different depths in the ionosphere. Thus, if the frequency sweeps of transmitter and receiver are synchronized, it is possible to measure the relative delay times for the pulses. A typical oblique incidence ionogram for the simple case in Figure 3.4 is shown in Figure 3.5. The lowest trace in Figure 3.5 shows what is recorded in the simple case when a one hop F-mode is present.

The high and low angle rays, see page 29, suffer different time delays in the ionosphere. For the low angle ray this delay will increase with increasing frequency, while for the high angle ray it will decrease with frequency. The two traces are clearly evident in Figure 3.5 and it can be seen how they merge at a particular frequency. At this frequency therefore, the two rays are identical and the frequency must be the MUF as defined on page 29. If multiple hop modes are present, other traces will appear on the ionogram. In Figure 3.5 a 2F-mode and a 3F-mode trace can be seen.

Note: A special notation for oblique ionograms has been agreed on. Capital letters are always used. Some of the most important terms are:

MOF Maximum observable frequency (not always equal to the MUF as defined on page

LOF Lowest observable frequency

JF Junction frequency at which high and low angle rays merge.

The low and high angle rays are distinguished by the letters L and H. The number of hops is also indicated, for example 2FMOF, 2FJF.

The problem of synchronizing a receiver with a transmitter thousands of kilometers away is not an easy one. An accuracy of about one millisecond is normally necessary to synchronize the frequency sweeps of transmitter and receiver. If one wishes to measure the absolute travelling time from T to  $R_2$ , an accuracy of the order of one microsecond is required in the time measurements. This is not practical at present, since extremely expensive clocks are needed. However, valuable information can be obtained from the records even if only the relative time delays are measured. The fact that the MUF can be read directly off the records is of course of particular practical importance.

The receiver  $R_1$ , situated close to the transmitter, will record pulses that have been scattered back either from the ionosphere or from the ground as indicated in Figure 3.4. At low and medium latitudes, where the ionosphere is normally horizontally stratified, the most common type of backscatter is the 2F2 trace shown on Figure 3.6. These echoes have travelled from T to  $R_2$  via the F2-layer and back again along the same route. The MUF for a one hop mode over a given distance can be derived from this type of record. At high latitudes, irregularities in electron density often occur in the F2-layer. These irregularities can take the form of cylinders orientated along the magnetic field lines,

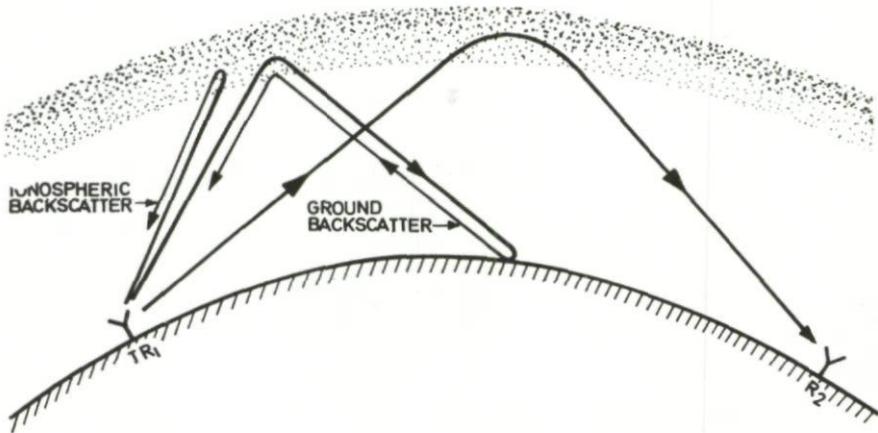


Figure 3.4 Oblique incidence sounding

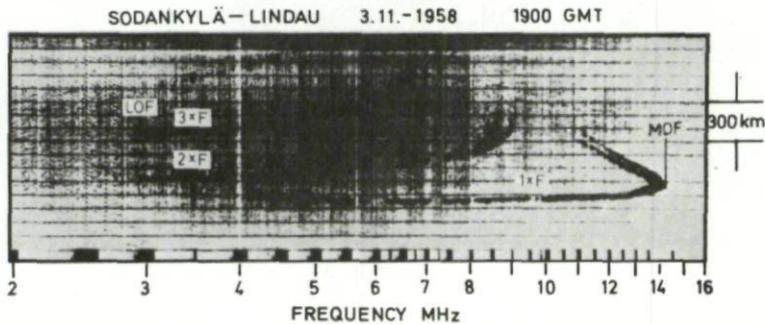


Figure 3.5 Oblique incidence ionogram  
 (Courtesy H.G. Moller, Max-Planck Institut für Ionosphären-  
 Physik, Lindau, West-Germany)

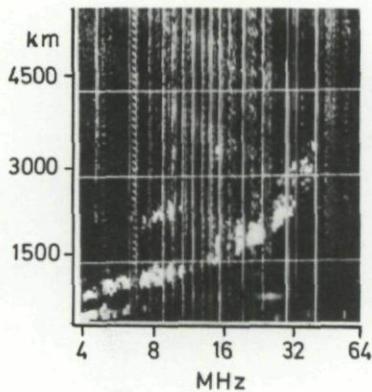


Figure 3.6 Backscatter record  
 (Courtesy K. Folkestad, Norwegian Defence Research Establishment)

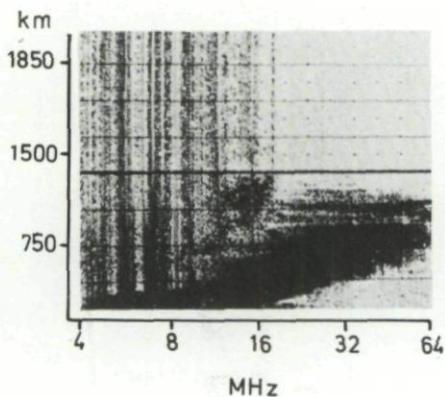


Figure 3.7 Backscatter record with auroral echoes  
 (Courtesy K. Folkestad, Norwegian Defence Research Establishment)

and they give rise to backscattered echoes. Irregularities can also be present in the E-layer, in particular in connection with auroral phenomena. Sometimes backscatter echoes from visual aurora are recorded. A backscatter record with reflections of the type described is shown in Figure 3. 7.

We understand that the oblique incidence sounder can be a useful tool to study the irregular phenomena occurring in the high latitude ionosphere. More details on the practical use of the method will be given on page 124.

#### Absorption measurements by means of reflected waves

So far we have mentioned techniques of importance for studies of the E- and F-layers, but which give little information about the lowest part of the ionosphere. On pages 28 and 38 it was pointed out that the D-region strongly absorbs high frequency radio waves, and that information about this region can be obtained from studies of the absorption of radio waves reflected from the E- and F-layers. By what is commonly called the A1 absorption method, the field strength of pulsed signals reflected vertically from the ionosphere is recorded on a number of frequencies in the range 1-7 MHz.

From the absorption versus frequency curve it is possible to obtain information about the intensity and position of the absorbing layer. Roughly speaking, the magnitude of the absorption gives the intensity of the ionization, while the shape of the absorption-frequency curve tells something about the height of the absorbing region. If most of the ionization is below 65 km, the measured absorption will be nearly independent of frequency, while ionization well above this level will produce absorption that varies inversely with the square of the frequency, see page 38.

A1 absorption measurements are often difficult to interpret because the non-deviative absorption occurring in the D-region must be distinguished from the deviative absorption suffered by the waves during the reflection process in the E- or F-layer, see page 38. Moreover, the method cannot be used during strong disturbances when the signals are completely absorbed.

An alternative method for studying the lower part of the ionosphere is to record the field strength of a continuous high frequency wave on a single frequency reflected vertically or obliquely from the E- or F-layers. This experiment is simple in that any high frequency broadcasting station can be used as a transmitter. On the other hand the interpretation of the data is difficult and direct information on the height of reflection cannot be obtained.

#### Measurements of cosmic noise absorption

As discussed on page 45 radio noise of extra-terrestrial origin is continually impinging on the earth, and for frequencies above the foF2 this noise will reach the earth's surface. By the A2 absorption method, or the riometer method (riometer = relative ionospheric opacity meter), the intensity of this type of noise is measured.

The cosmic radio noise comes from certain areas in the sky, mainly near the center of the galaxy. The riometer method is based on the assumption that the intensity of this radio noise is constant in time. Therefore, as the earth rotates relative to the noise sources, the intensity of radio noise falling on the earth's atmosphere at a fixed geographical position will vary in a regular manner and be a function of sidereal time only. The radio noise reaching the ground will have suffered absorption in the ionosphere, and variations in this absorption can be found by comparing measurements made at corresponding sidereal times.

The riometer equipment consists of a simple yagi antenna, a gain stable receiver, a noise diode and a pen-recorder. Calibration is done periodically by

injecting a known noise power from the noise diode into the receiver. A typical riometer operates on a frequency of about 30 MHz. A radio wave on 30 MHz travelling from space through the undisturbed ionosphere will suffer an absorption between 0.1 and 0.5 dB depending on the solar zenith angle. During disturbed conditions at high latitudes the absorption may be up to 15 dB. The riometer is mainly used to study such disturbances. The normal procedure is then to establish an average quiet day variation of radio noise intensity from measurements on days with no disturbances, and to measure the excessive absorption during disturbed conditions as the deviation from the smooth quiet day curve.

An example of a riometer record and a quiet day curve is shown in Figure 3.8. By using multifrequency riometers, information about the structure of the ionosphere can be obtained. The principle of this technique was described on page 53. Since riometers are inexpensive and simple to operate and maintain, they can be used to keep a continuous watch on the polar ionosphere, and are particularly well suited when a net of stations is needed to map the geographical extent of ionospheric phenomena.

#### VHF forward scatter

The maximum electron density in the ionosphere is rarely large enough to reflect waves in the VHF range (30-300 MHz). However, these waves can be scattered from irregularities in the electron density, and VHF forward scatter is used for communication purposes, Fig. 3.9. The mechanism is similar to that by which visible light is scattered when passing through a smoke-filled room. By studying the VHF signals scattered from the ionosphere during a polar radio blackout, information about the sizes and heights of the irregularities and about the absorption the waves suffer below the scattering region can be obtained.

#### Rockets and satellites

The rapid advances made in rocket and satellite techniques in recent years have made it possible and practical to carry instruments into the ionosphere to make in situ measurements. The rockets have proved particularly useful at high latitude because they can be used to study the lower part of the ionosphere in great detail during disturbances. The ionospheric conditions are then often changing so rapidly that more conventional ground-based techniques can give misleading results.

A number of different types of experiments have been flown on rockets for studying the arctic-ionosphere. Figure 3.10 shows a photograph of rocket payload launched in Northern Norway. This payload contains wave propagation experiments to measure the electron density and collision frequency, probes designed to measure ion densities along the rocket trajectory, particle counters and x-ray counters to study the influx of ionizing radiation, and photometers to measure the light intensity in the aurora. The rocket flight through the region of interest only lasts for a few minutes, and during this time the rocket must gather the data and telemeter the results down to earth.

The principle of a simple wave propagation experiment is sketched in Figure 3.11. The amplitudes and phases of HF-radio signals transmitted from the ground are measured in the rocket during its flight through the ionosphere and the information sent back to earth. From the measured changes in these parameters the electron density and collision frequency along the rocket trajectory can be deduced. Only by carrying out different types of measurements simultaneously, and comparing the results can one hope to map the very complicated processes taking place in the polar ionosphere during disturbed conditions. On page 19 we mentioned the satellite-borne ionosonde as an example of the use

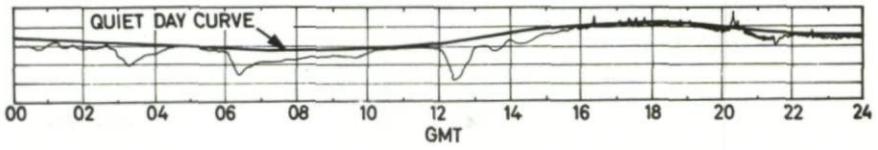


Figure 3.8 Riometer record during disturbance with quiet day curve superimposed (Tromso 7 November 1962)

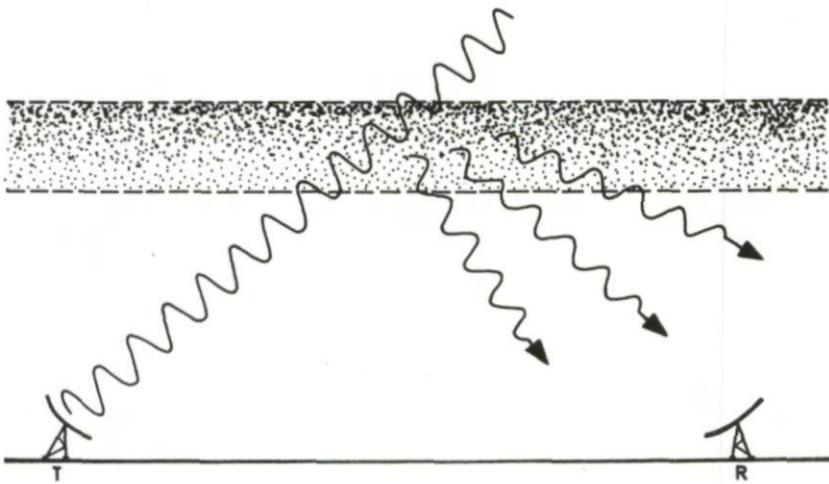


Figure 3.9 VHF forward scatter

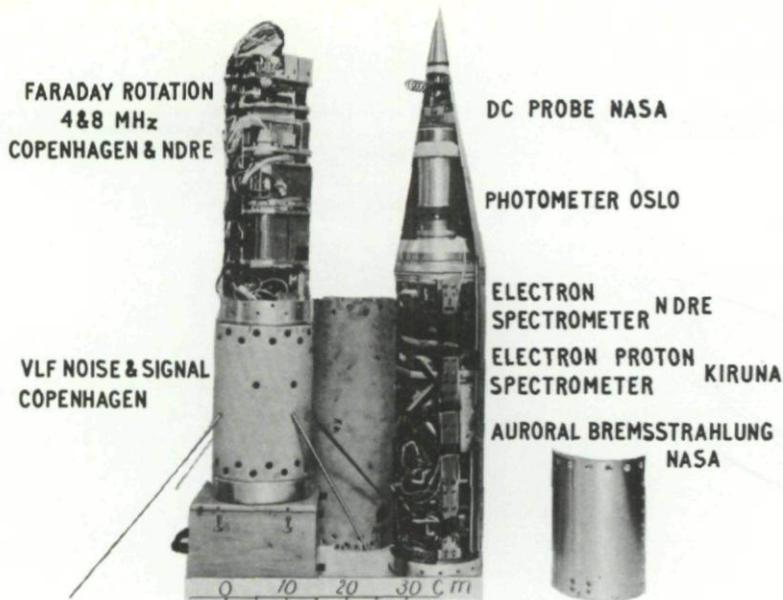


Figure 3. 10 A rocket payload launched from Andoya in Northern Norway

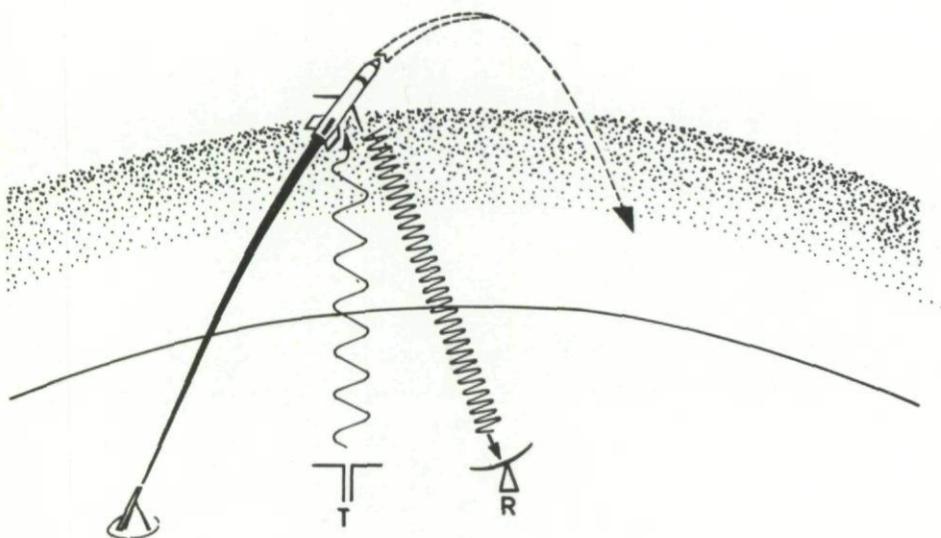


Figure 3. 11 Principle of a simple wave propagation rocket experiment

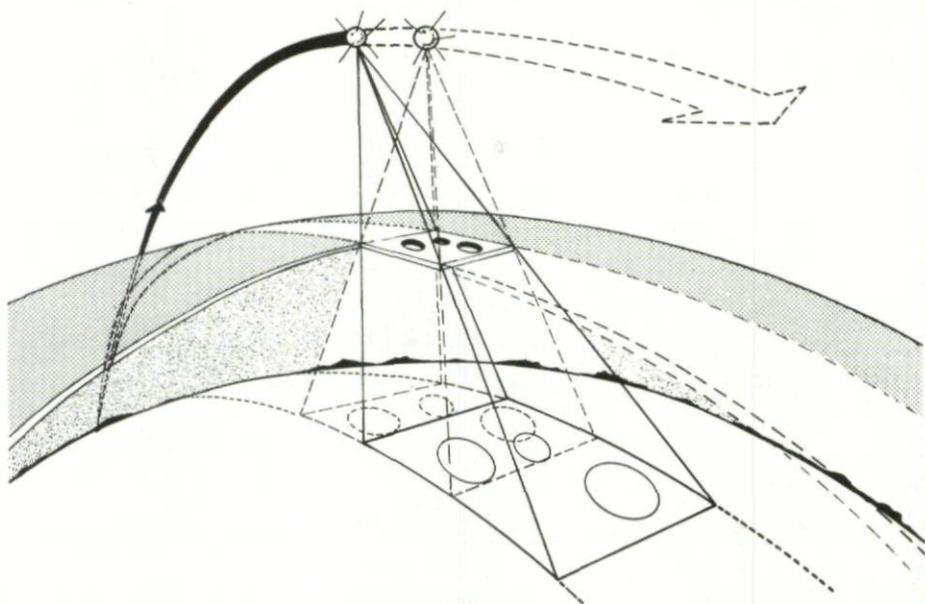


Figure 3. 12 Illustrating how satellite signals can be used to study irregularities in the ionosphere

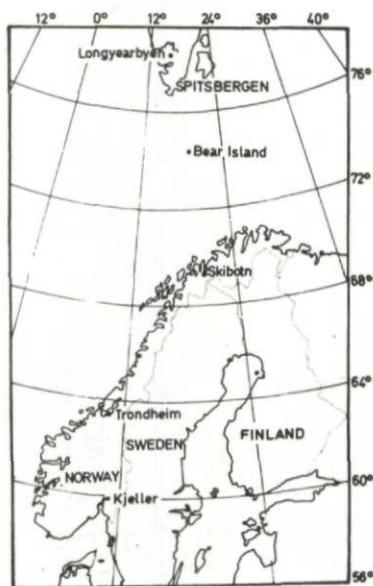


Figure 3. 13 Location of riometer stations

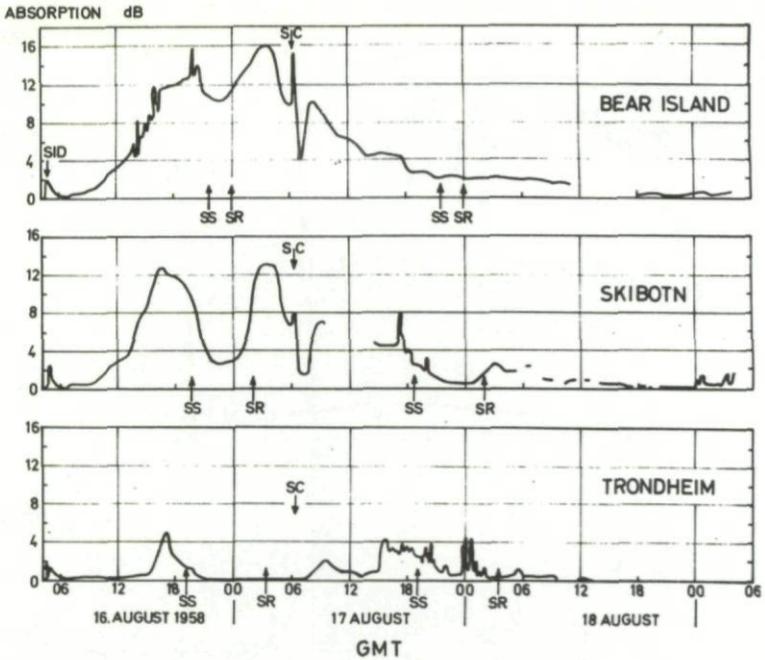


Figure 3.14 Riometer absorption observed at 27.6 MHz during the PCA event of 16 - 18 August 1958

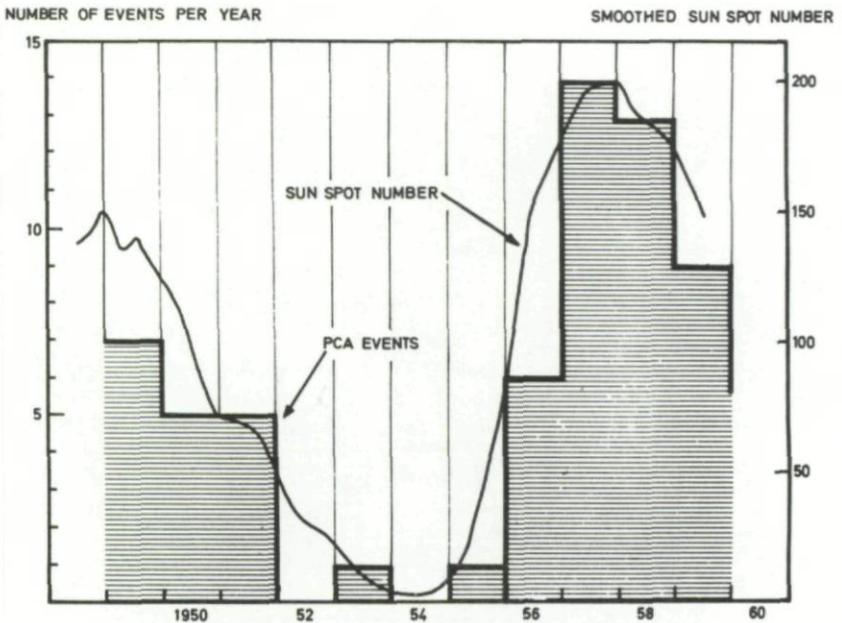


Figure 3.15 Occurrence of polar cap absorption events compared with smoothed sun spot numbers  
(After C. Collins et al, *Can J Phys*, 39, 35-52, 1961)

of satellites in ionospheric research. This instrument has greatly increased our knowledge of the polar ionosphere.

Nearly all artificial satellites transmit some kind of information down to earth, using frequencies well above the critical frequencies of the F2-layer, since the radio signal must at all times penetrate the ionosphere from above. Even these high or very high radio frequencies will be influenced by the ionosphere, and by studying the amplitude and phase of the satellite signal on the ground as the satellite passes overhead, useful information can be obtained. For example, it is possible to measure the total electron content of the ionosphere, that is the total number of electrons in a vertical column through the ionosphere with a cross section of  $1 \text{ cm}^2$ . Satellite signals can also be used to study irregularities in the ionosphere, Fig. 3. 12. The irregularities in ionization density will cause rapid fluctuations (scintillations) in amplitude and phase of the satellite signal as the satellite passes over the receiver. The satellite acts as a lamp that projects a moving picture of irregularities on the ground. The sizes and heights of the irregularities can be studied by receiving the signal simultaneously at several stations. On the whole rocket and satellite techniques are becoming increasingly important, but they will never replace ground-based experiments.

### Polar cap absorption (PCA)

The first PCA event to receive appreciable interest was the large event of 23 February 1956. Since that time a number of PCA-events have been studied, and information regarding, (i) the magnitude of the absorption, (ii) the frequency of occurrence of events, (iii) their average duration, and (iv) other effects of importance for HF communication, have been gained.

The most important ground-based observations of the PCA phenomenon have been made by means of riometers. Our present knowledge of the time and space variation of the absorption has mainly been obtained from such observations. In addition, information has been gained from analysis of forward scatter circuits. From rocket studies and sophisticated ground based techniques, information about electron density profiles has been obtained and finally satellite measurements of the ionizing particles have added to our knowledge of the physical processes. Many features of the polar cap absorption are still the subject of mere speculations, particularly the questions of the origin and interplanetary propagation of the ionizing particles. In the following we shall discuss some of the most important aspects of polar cap absorption.

### Description of a polar cap absorption period

We will introduce the subject by describing one particular PCA period that exhibits most of the characteristic features of such events. For this purpose the polar cap absorption that occurred from 16 to 18 August 1958 is chosen. The results from riometer observations at the three Norwegian stations at Bear Island, Skibotn, and Trondheim will be discussed. The riometers measured the galactic noise level at a frequency of 27. 6 MHz, and vertical three element yagi-antennas were used. The locations of the three stations are shown in Figure 3. 13 and the results obtained at the three stations are shown in Figure 3. 14.

A strong solar flare (importance 3+) occurred at 0433 GMT on 16 August. The flare was followed by an SID, which is clearly seen on the riometer records. The SID lasted for about 10 minutes, and the peak absorption was about 2 dB. About two hours later the absorption started to increase at the two northernmost stations, and from 1500 GMT also at Trondheim. A maximum was reached at about 1700 GMT at Trondheim and Skibotn, and some hours later at Bear Island. The maximum absorption was about 13 dB at Bear Island and Skibotn

and 5 dB at Trondheim. A minimum in the absorption was observed around midnight between the 16th and 17th, and in the morning hours of the 17th the absorption again increased at Bear Island and Skibotn reaching a maximum value of 16 dB and 13 dB respectively between 0300 and 0400 GMT. In this period there was no increased absorption at Trondheim.

At 0622 GMT on 17th August a magnetic storm started, and immediately following this, there was a brief but strong increase in the riometer absorption. After this increase, the absorption decreased considerably for a period of more than one hour. This decrease in absorption is often observed during PCA events. When the absorption again increases, this increase is also observed at the low latitude of Trondheim. A maximum absorption of 10 dB at Bear Island, 7 dB at Skibotn, and 2 dB at Trondheim, is reached two to four hours after the beginning of the magnetic storm. The absorption then decreases at all stations until about 1500 GMT, when a sudden strong increase is observed at Trondheim. There are no corresponding increases at the two northernmost stations and, in the period following this increase, the absorption is almost as strong at Trondheim as at the higher latitudes of the two other stations. In the following sections we will discuss briefly some of the interesting features of PCA events mentioned here.

#### Large scale features of PCA events

The PCA event of 16 to 18 August 1958 discussed in the preceding section illustrates some of the large scale features of PCA events. The time variations of absorption measured at the two northernmost stations are quite similar, at least for the daylight hours. This is in agreement with the theory that there is a uniform particle influx over the entire polar region. During the dark hours a decrease of the absorption to about one fifth of the day-time value is observed at Skibotn. This is in agreement with the general result that the day-time absorption is about four to five times stronger than the night-time absorptions. At Bear Island a much smaller decrease in the absorption is observed for the dark hours. This is because, at this station, part of the absorbing layer was still sunlit.

Note: The difference in the magnitude of the absorption observed for day and night-time is due to a photo-detachment process. During the night-time most of the free electrons will attach to neutral atoms or molecules and form stable negative ions. During day-time, however, the solar radiation will detach the electrons from the negative ions. The negative ions, although they carry the same electrical charge as the electron, will not seriously influence the radio wave propagation. This is because they are so much heavier than the electrons.

#### The occurrence frequency of PCA events

The occurrence probability of polar cap events is strongly dependent on the solar activity. As already stated this type of event was discovered in 1956. It has been possible, however, by analysing older data, to study the frequency of occurrence back to 1949. Figure 3.15 shows the distribution of PCA events together with smoothed sun spot numbers for the period 1949 to 1959. It is doubtful whether there is a seasonal variation in the frequency of occurrence of PCA events.

#### Duration of PCA events

From the communicator's point of view the percentage of the total time with polar cap absorption is more important than the number of events per year as discussed in the previous section. Though the events are relatively few, they may be of rather long duration. During the International Geophysical Year (1957-1958), for instance, one day out of eight was a day with polar cap absorption. In Figure 3.16 the distribution of the duration of the PCA events is given.

It follows that the average duration is about 1.5 days, but that particular events may last for several days.

#### Importance of PCA for radio communication

In estimating the importance of the polar cap absorption on HF communication, not only the frequency of occurrence and duration of the events are of importance. It is also important to know the magnitude of the absorption. In order to make use of results from riometers to estimate the equivalent absorption at frequencies in the HF band it is important to know the height at which the absorption takes place. We have seen that at the lowest altitudes, where the collision frequency  $\nu$  is large compared with wave frequency  $f$ , the absorption will be independent of the frequency, while at the greater heights where the collision frequency  $\nu$  is much smaller than  $f$ , the absorption will vary as  $1/f^2$ .

Only meagre information is available on the electron density profiles during PCA events. In Figure 3.17 we have shown some profiles measured by means of rockets and ground based observations. It is seen that although the ionization extends down to low altitudes, a considerable electron density is observed at the greater heights (where the absorption is proportional to  $1/f^2$ ), and this shows that the absorption at HF frequencies must be quite strong. This result is also found from analysis of ionospheric soundings, and of field strength measurements on actual HF circuits.

#### Auroral absorption

We shall now turn to a description of the absorption associated with solar particles of lower energies, Fig. 3.1. The effects on the ionosphere of these particles are normally observed after a period varying from 20 to 40 hours after the occurrence of a disturbance on the sun. However, as explained on page 79, these particles can be associated with minor as well as major disturbances. Minor solar disturbances occur quite frequently, and they are therefore of considerable practical interest for radio communication in the polar regions. This type of absorption is closely related to geomagnetic disturbances and auroral activity. A particular event normally occurs over a limited geographical area. From rocket studies information about the electron density distribution in the D-region has been obtained. Measurements of the ionizing particles carried out in rockets and satellites have also been of importance for our understanding of the physical processes. Most of our knowledge regarding the large scale temporal and geographical characteristics of the disturbances has been obtained from riometer observations. Quite extensive observations have been carried out in Alaska, Canada and Scandinavia during recent years. We shall make use of some results that have been obtained in Norway, in order to illustrate the main features of this type of absorption.

#### Direct measurements of primary particles and ionization profiles

As we have seen, the higher the energy of the particle the further down into the ionosphere will it penetrate and produce ionization. Particles of high energy may also produce x-rays (bremsstrahlung), capable of ionizing the very low part of the ionosphere (50-70 km height range). An energy spectrum in which there are relatively many particles with high energy is called a hard spectrum, contrary to a soft spectrum in which the low energy particles dominate. Results from measurements of particles in satellites and balloons, indicate that primary electrons of energy greater than 100 keV cannot be very important. This indicates that the particle spectrum is fairly soft and that ionization by bremsstrahlung is of little significance.

From the communicator's point of view it is of course the shape of the electron density distribution  $N(h)$  that is of importance. Only meagre information

based on actual measurements is available. In Figure 3. 18 we have shown some profiles that have been measured by means of research rockets from Andöya in Northern Norway. In the illustration we have given the magnitude of the absorption as measured by a riometer at a frequency of 27. 6 MHz in each case. We have distinguished between results obtained during night and day-time conditions. The results of Figure 3. 18 are in fair agreement with the previously mentioned result of a fairly soft spectrum of the primary particles.

The values of the collision frequency  $\nu$  as a function of height, can also be deduced from measurements in rockets. Using the best available data of this quantity and the results from Figure 3. 18, one can calculate the absorption as a function of height for different frequencies. The conclusion is that, most of the ionization is produced at that high levels that the frequency dependence of the absorption to a reasonably good approximation is given by the '1/f<sup>2</sup>-law'.

### Temporal and spatial characteristics of absorption

Results from riometer observations are of particular importance in deducing the temporal and spatial characteristics of the auroral type absorption.

We have seen in the preceding section that the absorption is roughly proportional to 1/f<sup>2</sup>. An absorption of say 1 dB at 27. 6 MHz would then correspond to the absorption L(f) at a frequency f, gives as:

$$L(f) = 1 \left( \frac{27.6}{f} \right)^2 \text{ dB}$$

For a frequency of 7 MHz the corresponding absorption would then be about 15 dB. This is the absorption for vertical transmission through the D-region. For a normal circuit one would have to calculate the total absorption for transmissions two times through the D-region, and also take into account that time transmission would be oblique, see page 43. If, for example, the angle of elevation  $\alpha = 30^\circ$ , a rough estimate of the excess absorption will be

$$\frac{2 \cdot L(f)}{\sin \alpha} = \frac{2 \cdot 15}{0.5} \text{ dB} = 60 \text{ dB.}$$

This is a considerable excess absorption, and it would seem reasonable to assume that most HF circuits would be blacked out during such conditions. We shall therefore regard 1 dB of riometer absorption as a threshold value in the further discussions.

In Figure 3. 19 we have shown a curve giving the diurnal variation of the percentage of the total time the absorption exceeds 1 dB at a typical auroral latitude station. We see that there is a clear maximum of occurrence in the early morning hours, and that in 16% of the total time the absorption exceeds 1 dB.

From a comparison of Norwegian and Canadian riometer results it has been found that the time of maximum probability of auroral absorption is governed by local geomagnetic time rather than local solar time. In Figure 3. 20 we have therefore plotted contours of percentage of time with absorption greater than 1 dB on a map in geomagnetic coordinates, (geomagnetic latitude and time). We see that the phenomenon occurs most frequently at a geomagnetic latitude of 66° and at a local geomagnetic time of about 08. 00 hours. The geomagnetic latitude where the absorption is most frequent, corresponds roughly to the zone of maximum auroral activity.

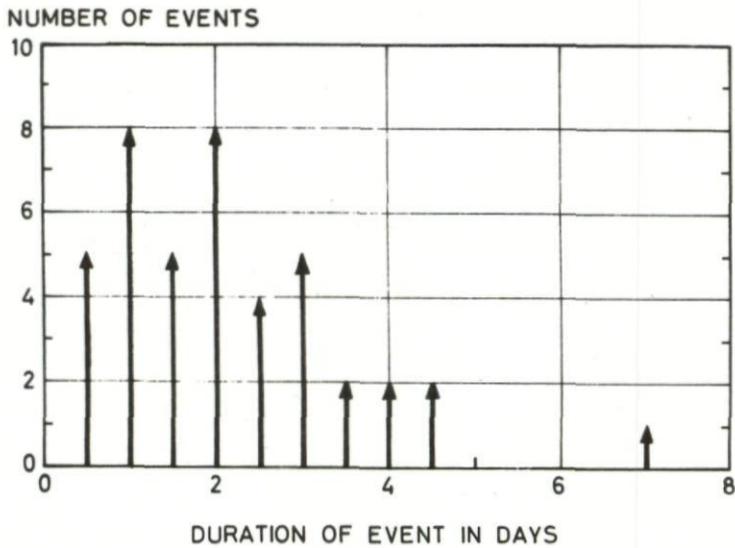


Figure 3.16 Distribution of duration of PCA events  
 (Prepared after table by C.S. Warwick and M.W. Haurwitz,  
 J Geophys Res, 67, 1317-1332, 1962)

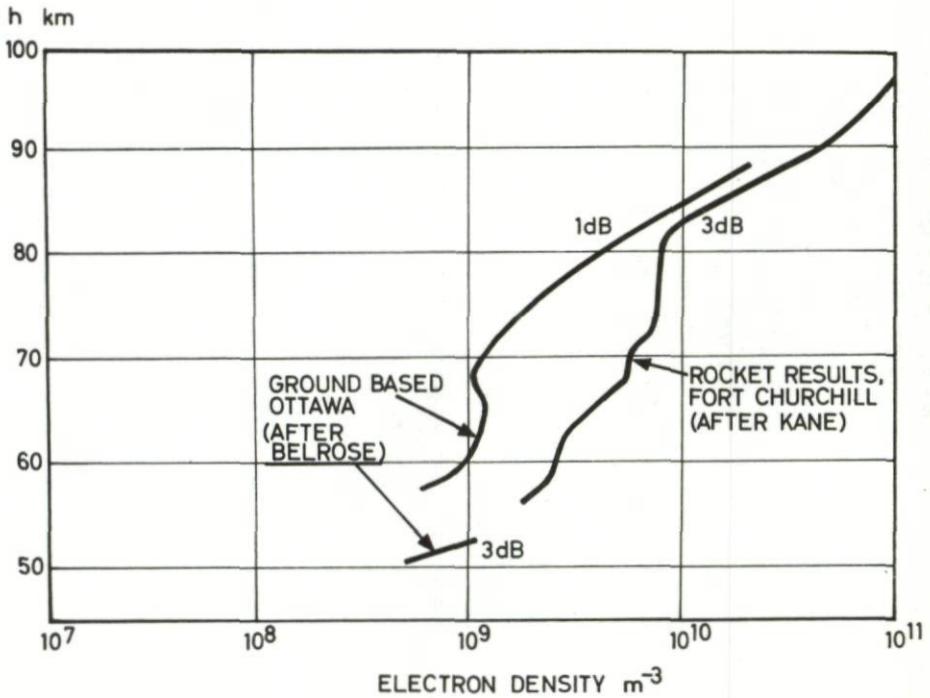


Figure 3.17 Electron density measurements during polar cap absorption

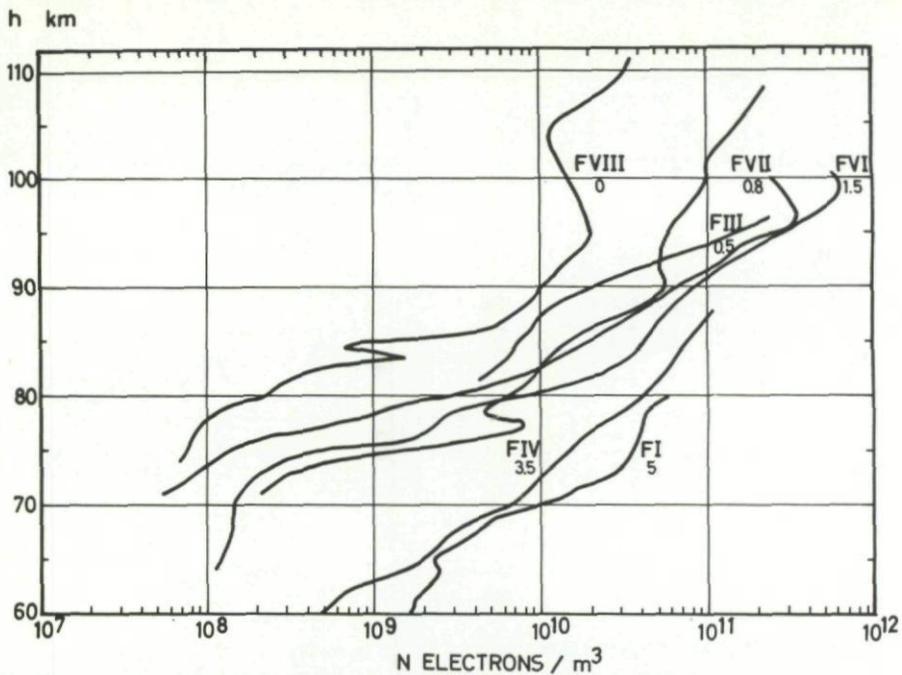


Figure 3.18 Electron density profiles measured by research rockets

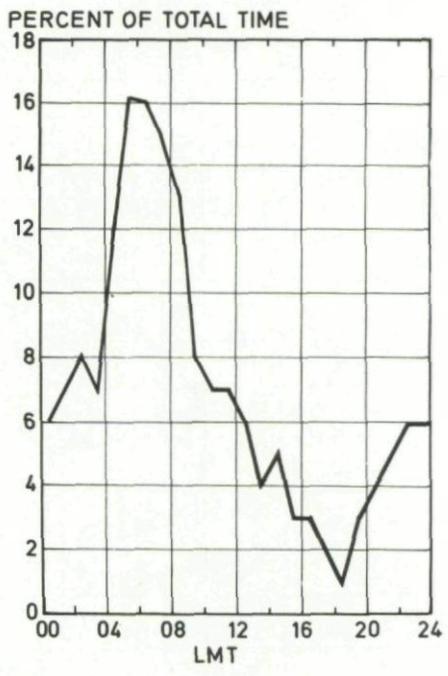


Figure 3.19 Diurnal variation of percentage of time with absorption greater than 1 dB. Observations from Skibotn, Norway

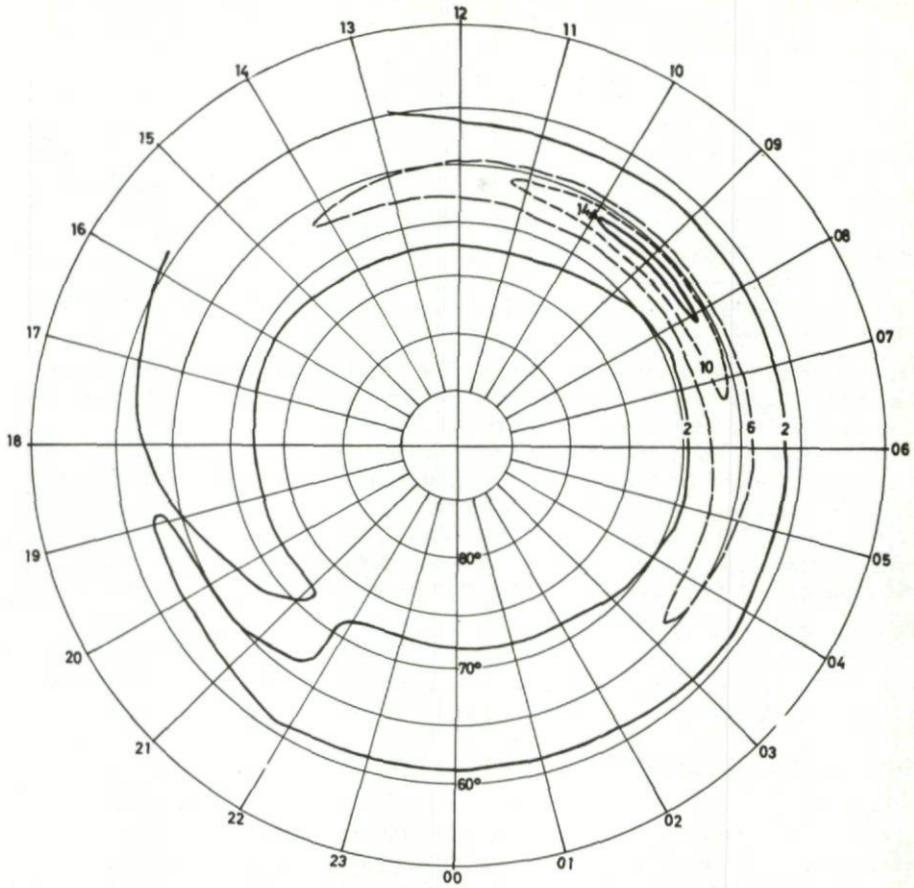


Figure 3.20 Percentage of time with absorption greater than 1 dB plotted in geomagnetic coordinates

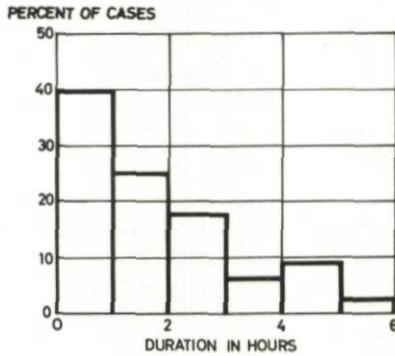


Figure 3.21 Observed probability that an interval with absorption greater than 0.5 dB is of a certain duration

### Duration and structure of absorbing regions

As already stated, the auroral absorption varies rapidly in time and space. In order to deduce the average duration of the events, we have measured the duration of all periods of continuous absorption greater than 0.5 dB for one year of observations (1958-1959) at five stations Longyearbyen, Bear Island, Skibotn, Trondheim and Kjeller, Fig. 3. 12.

We have grouped these observations into different classes according to the duration of the periods and shown the results as percentage occurrence frequencies in Figure 3. 21. It is seen that the duration of an event is normally shorter than about 2 hours. In order to study the structure of individual events, five riometer stations have been operated during a one year period (1959-1960) in a fairly closely spaced system, Fig. 3. 22.

The riometers were set up at Hammerfest, Alta, Kautokeino, Skibotn and Harstad. In Figure 3. 23 results of riometer recordings as obtained at the five stations during a disturbed period are shown. It may be seen from this illustration that the time variations of absorption measured at the five stations are very similar, except for some details. This result indicates that in this particular case the absorbing clouds are fairly large (of the order of some hundred kilometres) in horizontal extent. A great number of time periods when absorption was observed at the stations have been examined and it is found that this result is quite general.

There is a close relationship between the increased D-region ionization and geomagnetic and auroral activity. This is however, not a one to one relation in that one may have strong absorption without related strong aurora or magnetic activity and vice versa.

### High latitude sporadic E

In addition to the auroral absorption D-layer discussed in the previous section there is another sporadic layer of ionization in the auroral ionosphere which affects radio wave propagation at auroral latitudes. This ionization is situated in the height range 100-150 km, and is regularly referred to as the sporadic E-layer, the Es-layer or the night E-layer.

Contrary to the excessive ionization which is produced in the D-region during high latitude disturbances, the sporadic ionization at E-region heights may sometimes provide help in arctic communication. This possibility is not much exploited because the actual reflection properties of this layer for radiowaves have not been fully investigated. The sporadic E ionization in the auroral zone is mainly observed during the dark hours of the day, and it is generally believed that this layer, is caused by fast, electrically charged particles from space which, during their retardation in the earth's upper atmosphere, produce free electrons.

Ionosonde records show two types of Es traces, and these two types are referred to as the retardation and the non-retardation types of sporadic E ionization, respectively. The layer which gives rise to the retardation type of Es traces on the ionogram is fairly thick, whereas the non-retardation type of Es traces may be referred to as a thin, blobby, ionospheric layer. The time and spatial variations in the sporadic E ionization in the auroral zone are treated on page 104 and the importance of this layer in radio wave communication at high geographic latitudes is discussed on page 105.

### World wide distribution of the auroral Es-layer

The sporadic E-layer ionization at high latitudes is restricted to a certain latitude interval, which is centered at a latitude close to the maximum of visual auroral activity. This zone is at an average geographical latitude of

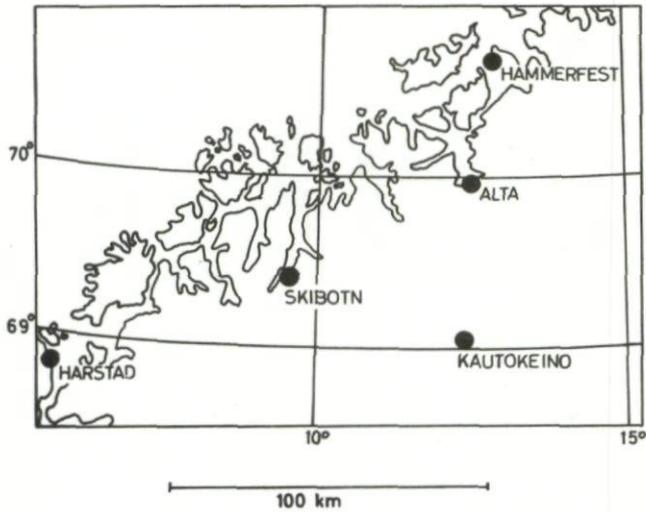


Figure 3.22 Location of riometer stations

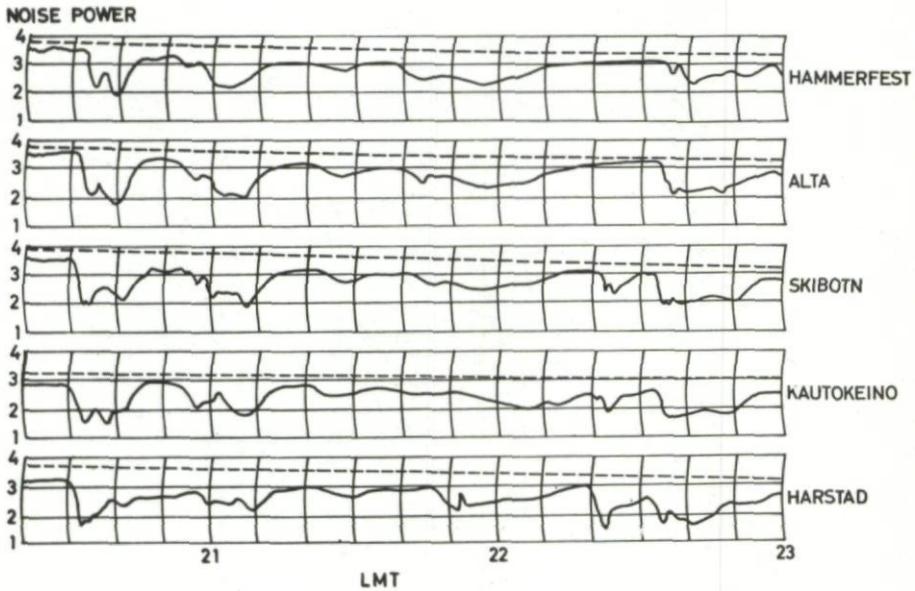


Figure 3.23 Example of riometer results obtained during a disturbed period (The observations were made on 23 January 1960)

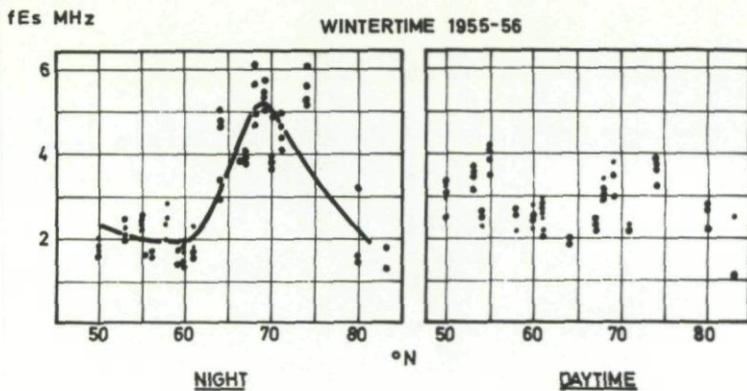


Figure 3.24 Average night-time values of  $fE_s$  observed over North America during the winter 1955 - 56  
 (After S. Matsushita, *J Atmosf Terr Phys*, 15, 68, 1959)

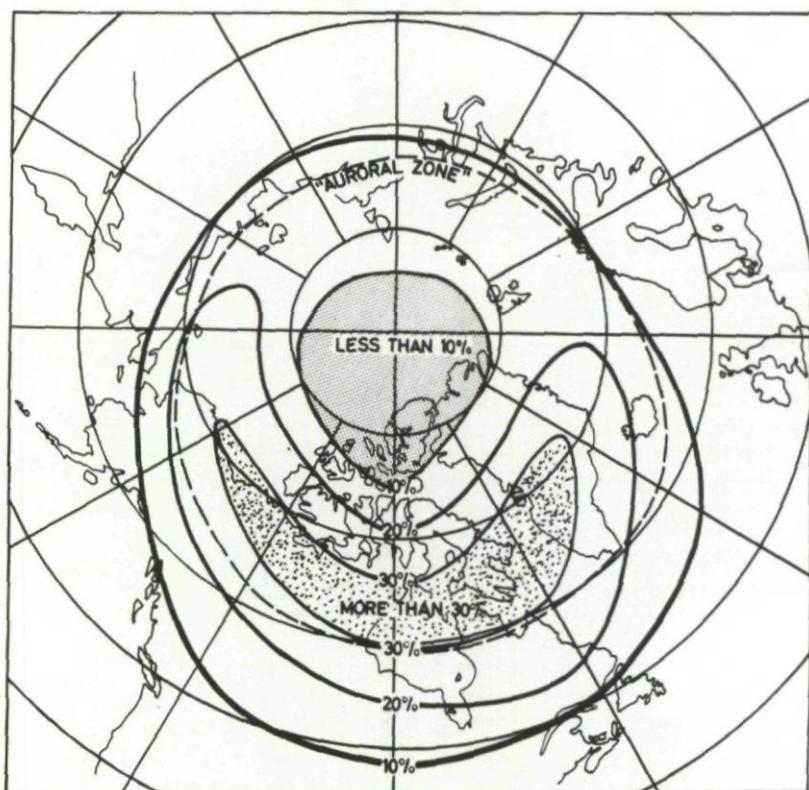


Figure 3.25 Percentage number of hours when  $fE_s$  is greater than 5 MHz at various geographical locations  
 (After E.K. Smith, National Bureau of Standards Circular 582, 1957)

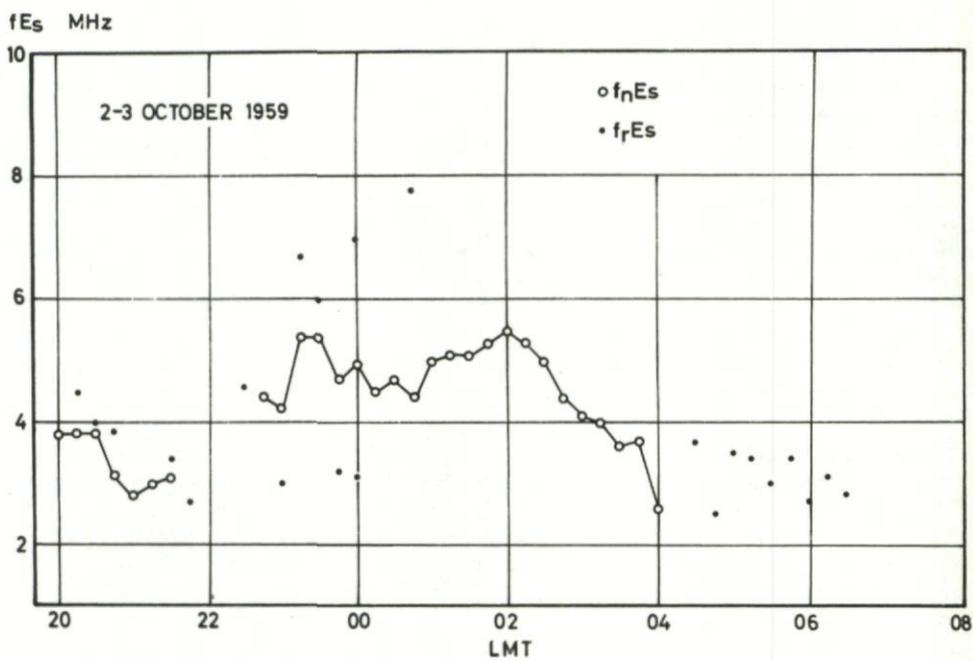


Figure 3.26 Typical time variation of  $f_{r}E_s$  (retardation type) and  $f_{n}E_s$  (non-retardation type)

about 70° over Northern Europe and at about 60° geographic latitude over North America. A similar zone is found in the Southern Hemisphere close to the zone of maximum occurrence of Aurora Australis. The average latitude variation of the highest HF radio wave reflected from the sporadic E-layer is shown in Figure 3. 24. This illustration shows that the region of high ionization is restricted to a latitude interval about 10° wide. During day-time the excess ionization is hardly visible in these regions. As shown, the peak plasma frequency of the layer seldom exceeds 6 MHz, which corresponds to an electron density of about  $0.5 \times 10^{12} \text{ m}^{-3}$ . The average maximum night-time fEs during the winter 1955-1956 is about 5 MHz.

For practical purposes it has proven useful to give the statistics of the Es-layer in terms of percentage number of hours when fEs exceeds a given value. The mean annual occurrence in the auroral zone of fEs greater than 5 MHz is given in Figure 3. 25. The most important results that can be derived from this illustration will be discussed here. Firstly, we note that during the period covered by Figure 3. 25, the fEs exceeded 5 MHz in 10-30% of all hours at high latitudes, except very close to the North Pole where the frequency of occurrence of sporadic E ionization is small. Secondly, it may be seen that the sporadic E ionization is dependent on geographic longitude. Thus, the phenomenon occurs about three times as frequently over North America as over Siberia. This has been assumed to be due to the geometry of the earth's magnetic field relative to the geographic coordinates, but the longitudinal variation of the Es ionization is far from being fully understood. Thirdly, Figure 3. 25 shows that the zone of maximum occurrence of fEs 5 MHz is close to the latitude of the visual auroral zone. The latter zone, which is indicated by the inner, broken line is situated a few degrees south of the maximum Es ionization.

An ionospheric layer does not necessarily provide a good medium for radio communication, even if the density of free electrons within the layer is large. The presence of small irregular clouds in the electron density may cause a high degree of fading, and this may cause serious problems in radio wave communication. It is therefore, important to know how the ionization varies in space before we can determine the usefulness of the layer for communication purposes.

The retardation type of sporadic E ionization is fairly uniform in a horizontal direction. Thus, ionosonde recordings at points separated by 100 km, are almost identical. From that point of view the layer is perfect for radio wave communication. The non-retardation Es, on the other hand, shows a very irregular structure. Probably this layer is made up of small clouds with high electron density. The actual size of the irregularities is not known.

#### Time variations

The sporadic E-layer in the auroral zone is mainly observed during night hours in contrast to the corresponding type of ionization at middle latitudes which is a typical day-time phenomenon. During 70-100% of all nights the fEs exceeds 5 MHz in the auroral zone, whereas such high values are seldom observed during day-time.

At stations, situated at the northern border of the auroral zone, the strongest Es ionization is found at about 2100 local time. Further south this maximum in the Es ionization occurs later in the night, and at the lower latitudes of the auroral Es-zone the maximum occurrence of this ionization is found in the early morning hours. The latitudinal variation of the maximum of Es ionization has never been properly explained. It is believed, however, that it is in some way related to the motion of the fast, charged particles which produce the ionization. The short term variations of the ionization in the two types

of Es are quite different. As mentioned earlier, the non-retardation type of Es-layer shows large and rapid time variations during the night. For that reason this layer is not very reliable for radio communication. Furthermore, the radio waves reflected from this layer normally fade rapidly, which is also inconvenient when information is to be transmitted via the ionosphere.

The retardation Es, on the other hand, shows more regular, slow variations during the night. In fact, this layer behaves more like a regular ionospheric layer, and it is, by some research workers, referred to as the night E-layer. Typical variations through the night of the critical frequencies of the two types of Es-layers are shown in Figure 3. 26. The difference in the time variations of the two types is clearly seen from this illustration.

In addition to the 'diurnal variation', the sporadic E-layer undergoes a typical 'storm-time variation'. During periods of strong auroral activity the earth's magnetic field lines are distorted. Normally, during geomagnetic storms, the sporadic E ionization near the auroral zone is particularly well developed. Furthermore, during such periods the maximum Es ionization is observed south of the visual auroral zone rather than in the middle of the zone. The maximum ionization may be shifted  $3^{\circ}$ - $6^{\circ}$  south of its normal geographical latitude during heavily disturbed periods.

#### Importance of sporadic E ionization for HF propagation

It follows from the preceding sections, that the retardation types of sporadic E are quite common phenomena and are therefore of considerable interest for radio communications in the polar regions. This layer occurs during conditions when no distinct disturbances are present, and could therefore be taken into account when planning an HF circuit. Both the sporadic E ionization and the auroral type ionization in the D-region are considerably enhanced during geomagnetic storms. As a result, sporadic E propagation will be of particular importance during periods when communication becomes difficult because of the auroral type absorption.

We have seen that the auroral absorption is occurring over somewhat limited geographical areas, and that it varies rapidly with time. Thus, we found from Figure 3. 21 that the duration of a period with continuous riometer absorption greater than 0.5 dB is not likely to be more than at most 2 hours. In the periods between the high absorption peaks, the sporadic E-layer often provides possibilities of communication. Such propagation may be possible at quite high frequencies, because of the large values of the factor M for this low layer. The highest possible value of the factor was discussed on page 33, see Figure 1. 30.

The most promising methods used to study the importance of sporadic E propagation during disturbed conditions in the polar regions, are perhaps the backscatter and oblique sounding techniques. We shall therefore, come back to the practical importance of sporadic E propagation in Chapter 4 where we shall discuss in some detail the usefulness of these techniques.

#### Communication in other frequency bands

The main aim of this book is to discuss problems of HF radio communications. It may be appropriate however, to include some of the most important results that have been obtained in connection with radio wave communications in the polar regions in other frequency bands.

#### VLF waves

The VLF band is normally defined to cover the frequency range from 3 to 30 kHz. These very long waves are not greatly influenced by the D-region

disturbances that occur in polar regions, as far as their usefulness for communication purposes is concerned. Communication on VLF will therefore be reliable even during very strong disturbances. Recordings of amplitude and phase on VLF circuits do, however, show a number of interesting features that are caused by the enhanced D-region ionization, and these may be of importance in connection, for example, with problems encountered in VLF navigation systems.

#### LF waves

The LF band is defined to cover the frequency range from 30-300 kHz. Although these waves are not quite as reliable for communications in the polar regions as the VLF waves, they may at times provide communication possibilities even during strong disturbances.

#### MF waves

The MF waves, covering the frequency band from 300 kHz to 3 MHz cannot be regarded as reliable for communications in the disturbed polar regions. They do often, however, provide communication possibilities during conditions when the HF waves fail. This is especially so for fairly weak disturbances. MF circuits are therefore in practical use, for example for communication between Spitsbergen and Norway.

The main difficulty with VLF, LF and MF waves is of course that fairly large antennas are needed to radiate sufficient power. This is why these waves are not well suited for transmissions from ships and aircraft.

#### VHF waves

Two communication systems are of importance at VHF waves covering the frequency band from 30 to 300 MHz, namely: ionosphere forward scatter, and meteor scatter circuits. It has been found that both systems may fail during polar cap absorption events, especially when frequencies in the lower part of the frequency band are used.

During auroral type absorption one normally observes an increase in the signal level on ionospheric forward scatter circuits. However, the fading may become very rapid during such conditions, introducing signal distortion that makes the circuits unsuitable for services needing a large bandwidth. The effects of auroral absorption on meteor scatter circuits is not very important, except for very strong disturbances and at the lower frequencies.

## Prediction and Reduction of Polar Communication Difficulties

### Introduction

The increasing need for a reliable network of radio communication in arctic and antarctic regions makes a prediction service for these regions imperative. The regular radio propagation prediction service surveyed in Chapter 2 is not intended for these parts of the world, partly because the 'regular' ionospheric layers do not behave regularly at high latitudes, partly because a number of 'irregular' layers occur in these regions. These irregularities in the high latitude ionosphere were discussed at some length in Chapter 3 of this book.

The prediction service for radio communication in the polar regions is only rudimentary, and only a small fraction of the predictions is available to the wireless operator in his work. Frequently, the ionospheric conditions prohibit a regular traffic between two terminals in polar regions. In these cases the radio operator has to transmit his information via other channels. A method which may augment the reliability of radio communication at high geographical latitudes has been developed recently. All of these points are discussed in this chapter.

### Predictions of disturbances

Long term and short term predictions of anomalous conditions in the upper atmosphere which affect HF radio wave communication, are submitted to the wireless operator from prediction centers located in various countries.

Most of the disturbances are directly related to irregularities in the solar emission. If these irregularities were predicted and we had a complete knowledge of the interaction between the solar phenomena and irregular variations in the earth's ionosphere, precautions could possibly be taken to prevent radio communication 'break down' during disturbed periods.

### Irregular solar emission

During periods when no disturbances occur on the sun, a constant flux of energy is transmitted from the sun into space. This energy is carried by energetic, charged particles and by electromagnetic radiation. The emission of solar energy interacts with the earth's upper atmosphere in various ways. The electromagnetic radiation at wavelengths greater than  $1 \text{ \AA}$  ( $1 \text{ \AA} = 10^{-10}$  meters) will be absorbed above 50 km, and increase the ionization at these altitudes.

The electrically charged particles from the sun, which are mainly electrons, hydrogen nuclei (protons) and helium nuclei ( $\alpha$ -particles) are directly and indirectly of importance for time variations in the upper atmosphere. Some of the particles may penetrate into the upper atmosphere at high latitudes and cause ionization in the atmosphere. The main bulk of the solar particles expelled from the undisturbed sun have, however, too low energy to contribute anything to the ionospheric layers. They may, however, contribute indirectly to time variations in the regular ionosphere, as they cause deformations in the earth's magnetic field. If we consider the stream of particles as

equivalent to a solar wind, we realize that the pressure on the day-side of the earth is higher than the pressure on the night-side. This has led some people to believe that the outer boundary of the earth's magnetic field is drop-shaped, similar to the shape of a water drop falling through air, Fig. 3. 2.

At times violent disturbances occur on the sun (solar storms), and during such periods the solar emission is normally enhanced. These enhancements are accompanied by various irregular upper atmosphere phenomena, which may seriously affect radio wave communication. The nature of the irregular ionization depends on what part of the solar radiation is enhanced. Enhancements in the electromagnetic radiation will cause disturbances in the ionosphere on the entire day-side of the earth, simultaneously at all latitudes. Enhancements in the emission of solar particles, on the other hand, will be accompanied by ionospheric effects at high latitudes only, both on the day and the night-side of the earth. The most important of these ionospheric disturbances are: (i) polar cap absorption ionization, (ii) auroral D-layer, (iii) auroral zone Es-layer, and (iv) sudden ionospheric disturbance (SID). In addition, irregular variations occur in the F2-layer after solar eruptions which are not considered in regular prediction bulletins. Certain aspects of the problems of predicting these disturbances are discussed in this section in the light of observed variations on the sun and in the earth's magnetic field.

#### Solar disturbances and terrestrial effects

Most of the irregular, sporadic ionizations in the upper atmosphere occur in conjunction with eruptions on the sun. All types of flares do not, however, have the same effect on the earth's upper atmosphere. During a large solar flare, which normally has a duration of a few minutes, soft x-rays are emitted from the sun and reach the earth after about eight minutes. In addition, during (and possibly after) a flare, fast charged particles are emitted from the solar corona. Some of these particles reach the earth after a few hours, whereas the main bulk of the particles need one to four days to travel from the sun to the earth. The solar flares occur most frequently in certain active regions on the sun, but no methods have yet been developed for predicting the occurrence of flares. We have, at present, no way of forecasting warnings for the ionospheric effect (SID) associated with the x-ray emission from flares. The particle-induced ionospheric phenomena may, however, be predicted, due to their long travel time from the sun. The effects of these particles have been discussed earlier, and only a few additional comments will be given.

The solar particles are emitted in a cone from the flare region and most of the particles travel many hours through the interplanetary medium before they eventually reach the earth's atmosphere. The estimated travel time is one to four days. The particles have greatest probability of hitting the earth when the flare occurs near the center of the solar disc. Hence, the most intense PCA events and auroral absorption events have been observed when the flare occurred near the central meridian of the sun. In this way we have two possibilities of predicting such events: (i) from visual observations of solar flares, short-term predictions of high-latitude ionospheric irregular events may be derived, (ii) as the largest disturbances occur when an active region is near the central meridian of the sun, there is a recurrency of twenty-seven days in the disturbances. This period refers to the period of solar rotation. The short-term and long-term predictions are discussed in some detail in the following two sections.

#### Short term predictions

The short-term predictions of ionospheric disturbances which affect radio wave communication, are mainly based on (i) optical observations of the sun, and (ii) knowledge of statistical relationships between solar phenomena and

the earth's ionosphere. The sun is continuously watched by a number of solar observatories on the earth. If certain anomalous phenomena occur on the sun, reports are sent to the radio wave propagation prediction centers, which immediately analyse the data and, if they find the situation serious, provide radio propagation and geomagnetic storm warnings for the nearest future. The following solar phenomena are of particular importance for the fore-caster:

(a) **Large solar flares** — The solar flares are characterized by an 'importance parameter', ranging from 0 to 3+. Solar flares of importance 3 and 3+ are almost always accompanied by disturbances in the earth's atmosphere.

(b) **Position of flares and active regions on the sun** — When a region of activity appears near the central part of the solar disc, the probability of a disturbance in the ionosphere is somewhat larger than when the active region is situated near the limb of the sun.

(c) **Characteristics of the solar disturbance** — During periods of activity the electromagnetic emission from the sun is enhanced. The characteristics of this emission are of great value in the predictions. In the early stages of a solar flare, emissions of very high frequency radio waves occur on the sun. Observation of the radio emission on a frequency of say 200 MHz, can give indication of an active region in progress.

From a selection of the data from various solar observatories, short-term predictions are prepared for radio communication conditions and geomagnetic activity. In this section some general comments only will be given on the predictions.

All predictions are based on statistical experience gained from a large number of earlier, similar events. These statistics tell the forecaster that, when a certain type of disturbance occurs on the sun, there is a given probability that this will be accompanied by ionospheric effects. In most cases the prediction will be given by a code number which refers to phrases like 'there is a good chance', 'there is a fair chance' etc. depending on whether the experience shows that this occurs always or less frequently in conjunction with the corresponding solar phenomena. Generally speaking, the vagueness of the forecast indicates whether or not there is a well established relationship between a certain solar disturbance and the corresponding ionospheric phenomenon.

The predicted geomagnetic storm parameter A can vary from 0-400, the higher the value the greater the disturbance. Experience tells us that the quality of a radio circuit via the ionosphere is related to the value of the A-index. The geomagnetic predictions may be used with confidence as an indicator for radio communication near the auroral zone. An average picture of A versus radio communication conditions in the auroral zone is given in Figure 4. 1.

#### Long term predictions

If we look at the average degree of disturbance in the ionosphere and the geomagnetic field, we realize that there is a clear recurrence tendency within a period of 27 days. If, as an example, the geomagnetic disturbance index is arranged in a 27 day sequence, we see that the disturbance is repeated every 27 days for some months, Fig. 4. 2.

The explanation for this recurrence tendency is as follows:

Disturbances in the earth's upper atmosphere are closely linked to disturbances on the sun. Disturbances in the ionosphere are observed, particularly during periods when active solar flare regions cross the central meridian line of

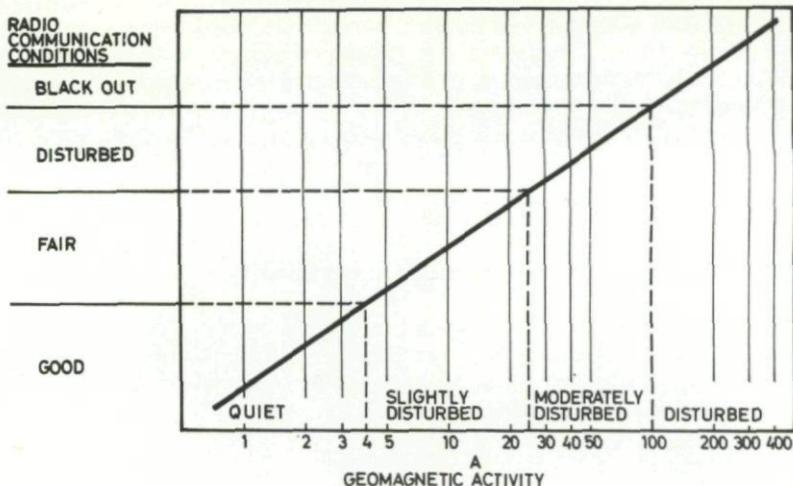


Figure 4.1. Empirical relationship between the radio propagation index and the geomagnetic storm index A  
(Courtesy CRPL Forecast Center, For Belvoir Va, U.S.A.)

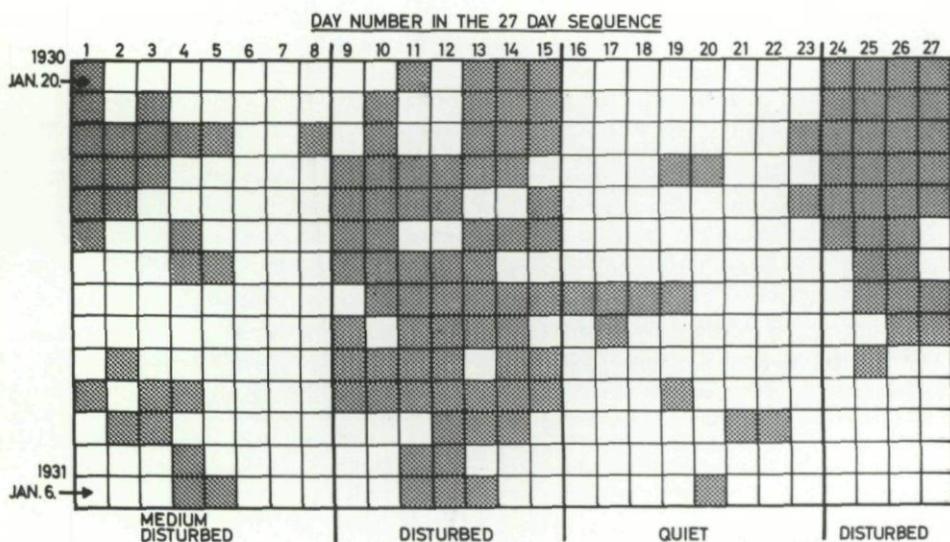


Figure 4.2 The year 1930 - 31 arranged in 27 day sequences to show the recurrency of geomagnetically disturbed days (indicated by black squares)  
(After S. Chapman and J. Bartels, *Geomagnetism*, Clarendon Press, 1940)

the sun. The active solar centres frequently have a lifetime of several rotations. Hence, one single group of irregular emissions may cross the central meridian of the sun more than one time during its lifetime. The period of the solar rotation is about 27 days which explains the long term periodicity in the ionospheric and geomagnetic activity.

During the years of minimum sun spot activity the 27 day recurrency is more pronounced than near maximum periods. Therefore, during periods of low solar activity one might predict some of the ionospheric disturbances about a month ahead. Such long term predictions are available from the prediction centers in the world.

#### Available predictions of disturbance

Short-term and long-term HF radio propagation forecasts are submitted from various prediction centres in the world. In this section the forms of the forecasts distributed by the ITSA in the United States of America are reviewed as examples. The ITSA forecast centres provide four types of radio propagation predictions:

- (a) **Advanced Forecasts** — which cover periods from 7 to 25 days.
- (b) **Medium Term Forecasts** — for high latitude circuits, covering a period of 24 hours.
- (c) **Short Term Forecasts** — for high latitude circuits covering a period of 6 hours.
- (d) **Fade Out Warning** — for the following day.

Most of the predictions are distributed by air mail, telephone or telegraph on special request only. For that reason most of these are of little or no use to the wireless operator. In this section we will therefore concentrate on the short term forecast which is the only radio propagation forecast submitted directly to the user.

The National Bureau of Standards station WWV transmits the short term predictions for the North Atlantic region in international morse code every five minutes. Short term forecasts are issued four times a day at nominal six-hour intervals: 0500, 1200 (1100 in summer), 1700, and 2300 GMT. Each forecast refers primarily to the six-hour period beginning one hour after the time of issue, for example, the 0500 forecast refers to the 0600-1200 GMT period.

The forecast statement consists of a letter and a number; the number is the forecast, while the letter identifies the quality of radio propagation conditions prevailing at the time the forecast is issued.

**Number** — The radio quality which is forecast for the coming period is expressed in the following quality grades:

1 - useless	4 - poor-to-fair	7 - good
2 - very poor	5 - fair	8 - very good
3 - poor	6 - fair-to-good	9 - excellent

**Letter** — The radio quality at the time of issue of the forecast is identified as follows:

- W - disturbed (quality 1, 2, 3 or 4)
- U - unsettled (quality 5)
- N - normal (quality 6, 7, 8 or 9)

The letter describes the average quality in the two hours preceding the issue time — 0300 to 0500, 1000 to 1200, 1500 to 1700, or 2100 to 2300 GMT.

A typical forecast statement is U-6 which means, 'propagation conditions are now unsettled but radio quality is expected to improve to fair-to-good during the hours covered by the forecast'.

#### Interpretation of predictions

When interpreting the short term HF radio propagation warnings the wireless operator should realize the following limitations of the predictions:

- (a) The predictions are meant to cover certain typical, standard high latitude radio circuits, such as Washington-Paris, Tokyo-Anchorage etc. This implies that shorter circuits may not necessarily experience the same radio disturbances, as the region of disturbances is frequently very localized. Furthermore, the short-term predictions will generally be of no use for radio circuits at lower geographical latitudes.
- (b) The predictions are intended to cover the entire high frequency band, from about 3 to 30 MHz.
- (c) The radio quality scale  $Q_a$  which runs from 1 to 9 is essentially based on subjective evaluation of conditions. The wireless operator has to use his own experience in order to translate the quality scale into a form which is usable for himself. In order to clarify this point, the procedure used by the ITSA for deducing the radio quality scale for North Atlantic Radio Paths is briefly reviewed.

The scale of the North Atlantic Radio Quality Figures is based on the average experience reported by about ten communication and monitoring agencies. These are asked to report the performance of their circuits on the 1 to 9 scale, having regard for the subjective equivalents of the quality grades. The distribution of all ratings by all these reporters in a period of, say, twelve months is regarded as the master distribution. In order to calibrate the quality scale of individual reporters, regardless of whether or not the reports are on the 1 to 9 scale, the distribution of quality grades reported by the individual is related to the master distribution for the same interval of time. This results in a conversion scale for reducing individual reports to the North Atlantic  $Q_a$ -scale. Because there is a significant diurnal variation in quality as reported to ITSA, separate distributions are derived for each six-hour interval of the day.

A typical master distribution is shown in Figure 4.3, and with it is the distribution of quality grades by one of the individual reporters which used the 1 to 9 scale and therefore helped determine the master distribution except at the extreme quiet end of the scale. This reporter consistently graded conditions less disturbed than average. Thus, while his reports are on the 1 to 9 scale, they must be adjusted downward for best agreement with the average of several such reporters. The individual distribution is thus forced to agree with the master distribution, and current indices for a given portion of the day are reduced by the corresponding conversion table to the  $Q_a$ -figure scale.

#### De-routing of traffic

In some cases, when it is impossible to obtain radio communication because of excessive radio wave attenuation in the ionosphere, communication may be re-established by making use of a relay station or by selecting another station for forwarding the traffic. If, for example, a ship tries to contact a coast station, and no contact can be obtained because of ionospheric absorption, it might

PERCENTAGE OF TIME  
A GIVEN QUALITY  
WAS NOT EXCEEDED

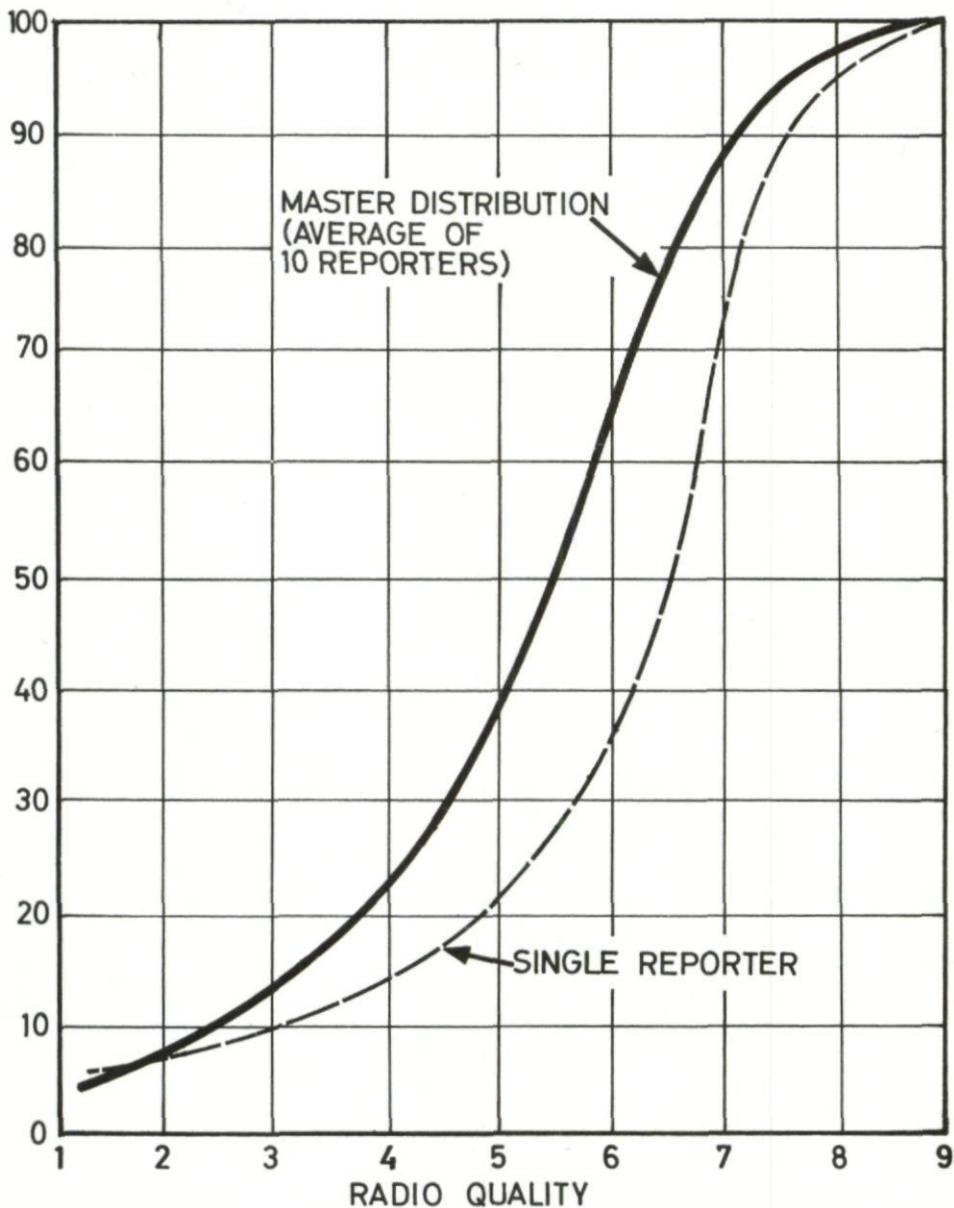


Figure 4.3 Average reported HF radio communication conditions versus the (Q<sub>a</sub>) radio quality scale compared with the report from one single observer  
(Courtesy CRPL Forecast Center, Fort Belvoir, Va, U.S.A.)

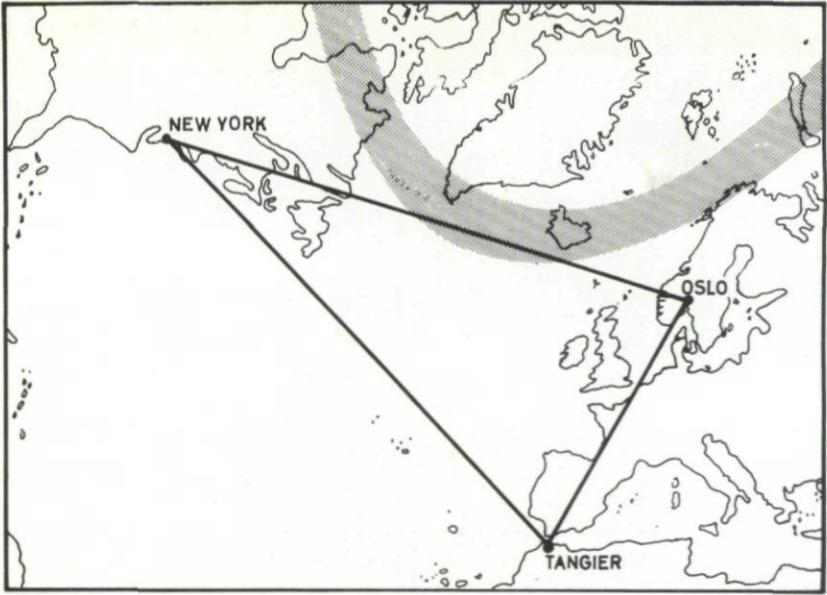


Figure 4.4 De-routing of traffic Oslo - New York via Tangier

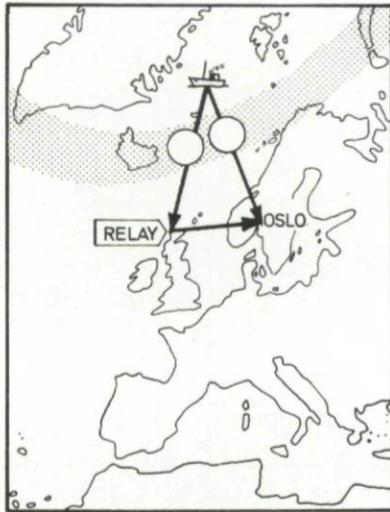


Figure 4.5 Suggested situation

be possible to contact another coast station for forwarding the traffic. In these cases it may not be necessary to rely entirely upon HF radio waves, as cables or LF waves are often used between stations on land. Transmission of LF waves from ships and aircraft is normally very difficult because of the low gain of the antennas.

In this chapter we shall discuss the possibilities of avoiding radio communication break down in polar regions.

#### General considerations

There are two main types of irregular ionospheric radio wave absorption in polar regions, (i) the polar absorption and (ii) the auroral absorption. During a polar cap absorption event, the area of excessive absorption normally covers the entire polar cap, and it is not easy to find means for maintaining communication on HF waves during such events.

When auroral absorption occurs, the problem is less difficult. We have seen that disturbed regions are generally limited in horizontal extent, and that the phenomenon varies rapidly in time and space. Even during disturbed conditions there are quite often short time intervals when propagation via the sporadic E-layer becomes possible. If a relay station is used in addition to the ordinary terminal, the time intervals when sporadic E propagation is possible may not coincide for the two stations, and one will therefore increase the total time with a possible Es-mode.

We have seen that the auroral absorption is occurring most frequently near the auroral zone. For radio circuits having their great circle path along this zone, the radio communication is frequently disturbed by auroral absorption. As seen from Figure 4.4 the great circle path for a circuit between Oslo, Norway and New York, USA, is partly within the auroral zone. This circuit is very often disturbed by auroral absorption. To avoid the absorption on this circuit and most of the other circuits between Northern Europe and New York, one has chosen to use a relaying station situated in Tangier in North Africa. We can see from the illustration that the two great circle paths necessary for this relayed circuit are well outside the auroral zone and not subject to auroral absorption. By using modern relaying techniques the time delay is negligible. Another example where de-routing of traffic has proved important was experienced during the war by Allied convoys. They often found that traffic to England could be maintained through a relay station in Alexandria, Egypt, under conditions when the direct radio circuit could not be used.

#### Deductions based upon results from riometer observations

The duration and size of the absorbing regions were discussed on page 100. The discussion on that page was based mainly upon results from riometer observations, where five riometers were set up in Northern Norway in a fairly closely spaced system, Fig. 3.22. The results from these measurements can be used in order to estimate the possible reduction of the periods of outages using a relay technique.

Consider for example, a situation as shown in Figure 4.5. A ship, the position of which is shown in the illustration, wishes to communicate with Oslo, Norway. Assume that this can be done either; (i) by a direct circuit, or (ii) by de-routing the traffic through a station in Northern Scotland. In this example, we assume that most of the radio wave attenuation in the two ship-to-shore circuits would take place in the shadowed area. The two circles show the part of the D-region of importance for the absorption. The absorption as measured by riometers in these circles would give a qualitative estimate of the absorption to be expected for the two suggested radio circuits. The distance between the two circles in Figure 4.5 is about the same as the distance between the riometers at Harstad

and Kautokeino, Norway, Fig. 3. 22. It seems reasonable therefore, that one can make use of the results obtained from these stations in order to estimate the possible gain introduced by our relaying system.

In Figure 4. 6 we have shown the percentage of the total time the riometer absorption is below selected values. We see for example, that in about 90% of the total time, the riometer absorption was less than one decibel at the two stations, and that the results for the two stations are similar. In the illustration we have also shown the results for the two stations combined, that is, the percentage of the total time the absorption is below the indicated value at at least one of the stations. It follows from the illustration that this is the case for absorption less than one decibel in about 96% of the total time.

We have seen earlier, that a riometer absorption of about one decibel can be regarded as a threshold level, such that higher absorption would give a circuit output. The net gain by introducing the relay circuit would therefore be a reduction of the output time from about 10% to about 4% of the total time. The example discussed, is one of rather short distance de-routing. Consider again the ship in the same position, Fig. 4. 7. In this illustration an alternative relay station in Tangier, North Africa is suggested in order to communicate with Oslo, Norway. The idea is that, for the circuit from the ship to Tangier, the waves will be transmitted from the ship at very low elevation angles, in order to propagate underneath the D-region in the auroral zone (shadowed area) where the excessive absorption is most frequent. Results from riometers situated in the centres of the two circles in Figure 4. 7 could be used to estimate the auroral type absorption for the circuits from the ship to Tangier and to Oslo. The relative positions of the two circles versus the auroral zone correspond roughly to those of Kautokeino and Kjeller, Figs. 3. 13 and 3. 22, and results from these stations can be used in order to estimate the output time for the circuits.

In Figure 4. 8, we have shown the percentage of total time the riometer absorption is below different values at the two stations. This illustration shows that the auroral type absorption is less frequent at Kjeller than at Kautokeino. The radio circuit from the ship to Tangier would therefore be considerably less affected by the auroral absorption than the direct circuit from the ship to Oslo. If we again regard one decibel of riometer absorption as a threshold level, the output time due to auroral absorption would be reduced from about 10% of the total time for the direct circuit to about 2% of the total time for the circuit to Tangier. This figure is somewhat optimistic because of the higher normal absorption to be expected for the circuit from the ship to Tangier as compared with the direct circuit to Oslo.

### **Oblique incidence and backscatter soundings**

In Chapter 3 it was shown that the polar ionosphere is extremely variable in time and space, and that reliable predictions of the useful frequency range for a given circuit at a given time are difficult to produce.

The radio operator would like to see a technique yielding an instantaneous picture of the state of the ionosphere in a wide area around his station. He would like to know the location, extent and intensity of the disturbed regions and how the disturbances developed from minute to minute. He would then be able to judge the situation and at any time choose the best propagation path and frequency. The oblique incidence sounding technique is at present the most promising method for achieving this goal, and in this section we shall describe how the technique can be used in order to improve the reliability of polar communication. Unfortunately, oblique incidence sounders are still quite expensive, but in the future they may become standard equipment at the main wireless stations.

PERCENTAGE OF TOTAL TIME

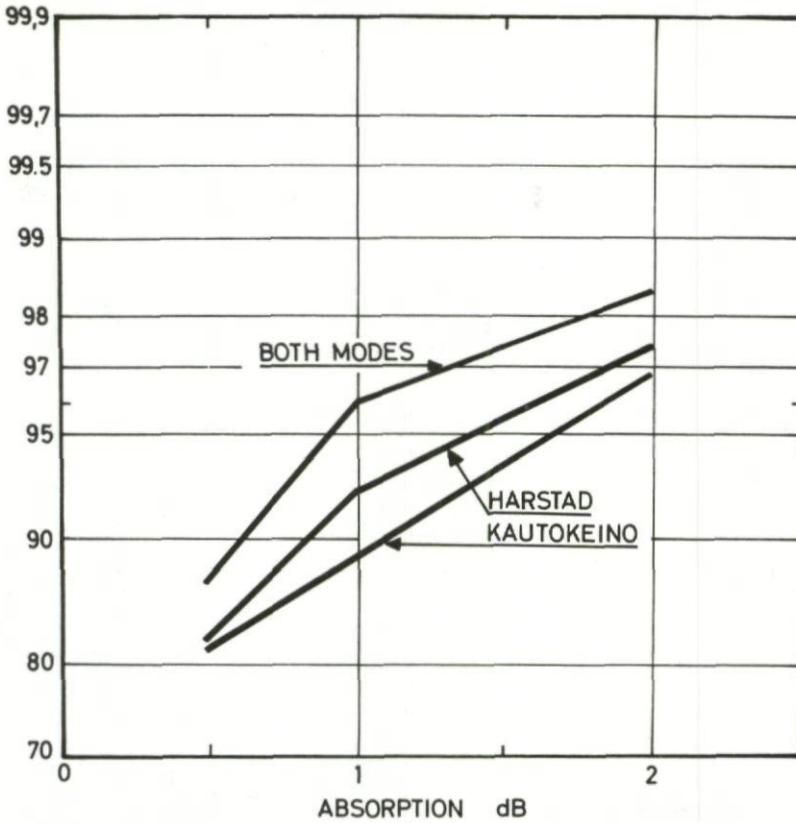


Figure 4.6 Percentage of total time the riometer absorption is below different values (Harstad and Kautokeino)

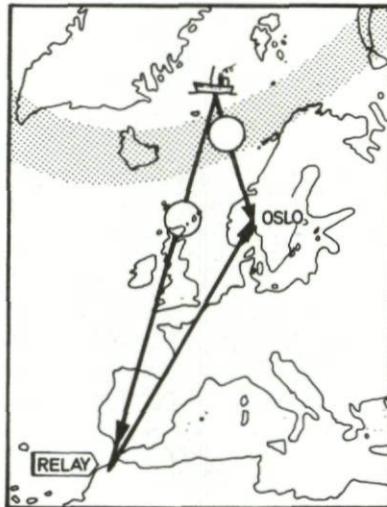


Figure 4.7 Suggested situation for an alternative de-routing

PERCENTAGE OF TOTAL TIME

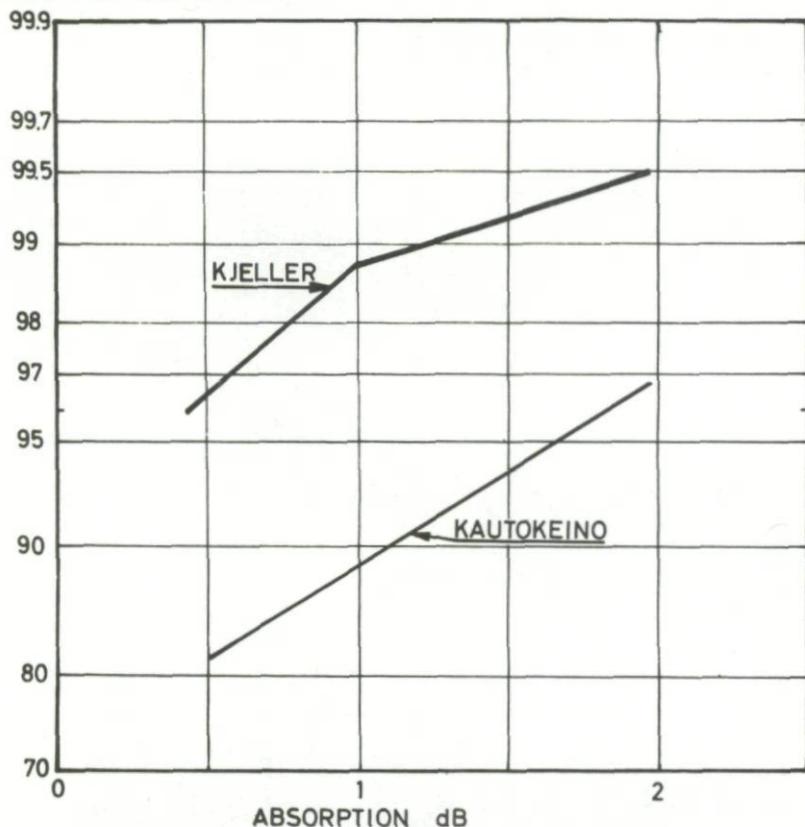


Figure 4.8 Percentage of total time with riometer absorption below different values (Kautokeino and Kjeller)

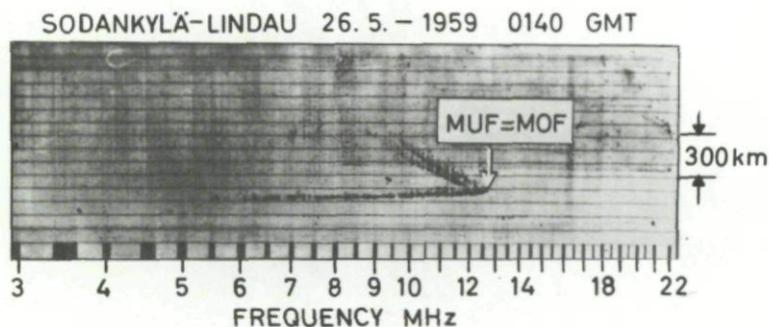
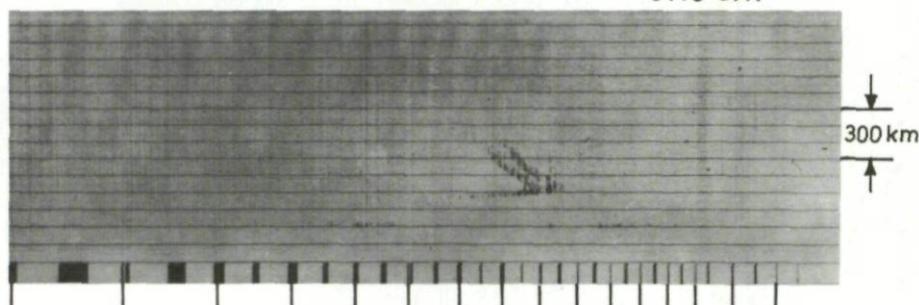
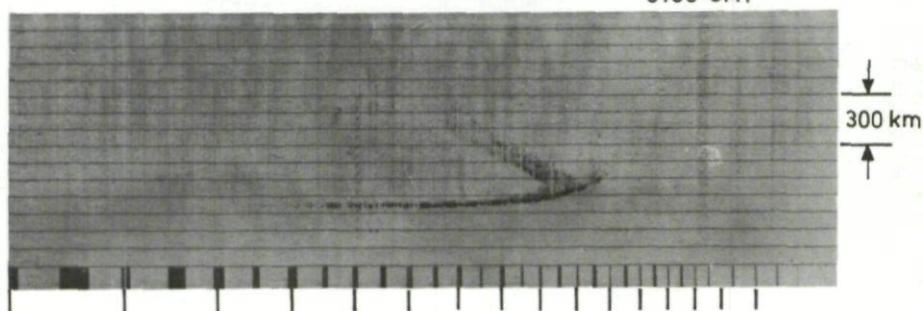


Figure 4.9 Typical oblique incidence ionogram (Courtesy H. G. Moller, Max-Planck Institut für Ionosphären-Physik, Lindau, West-Germany)

0110 GMT



0130 GMT



0140 GMT

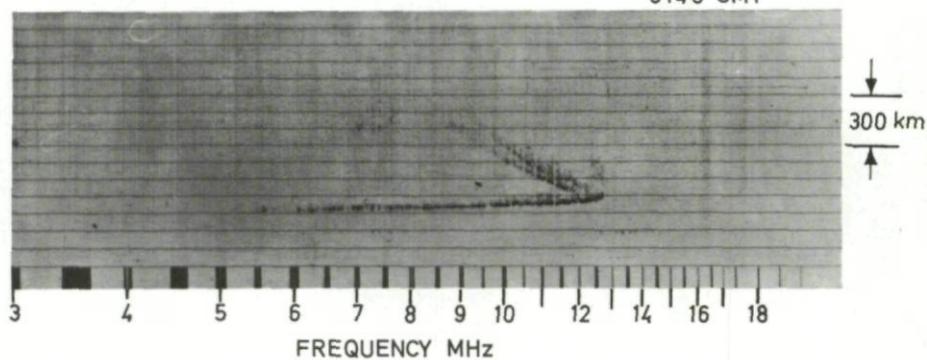


Figure 4.10 Oblique incidence ionograms (strong absorption)  
(Courtesy H.G. Möller, Max-Planck Institut für Ionosphären-Physik, Lindau, West-Germany)

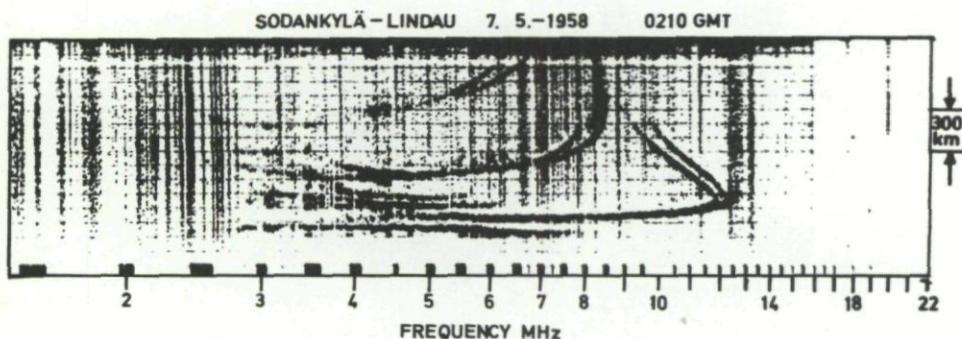


Figure 4.11 Oblique incidence ionogram obtained during a period with weak absorption  
 (Courtesy H. G. Moller, Max-Planck Institut fur Ionosphären-Physik, Lindau, West-Germany)

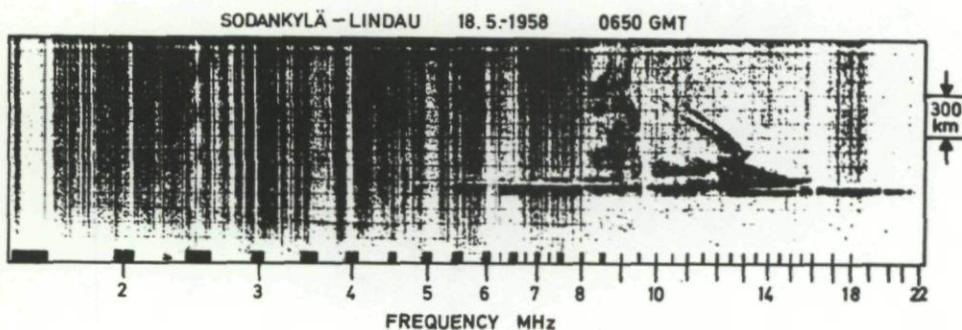


Figure 4.12 Oblique incidence ionograms with sporadic E trace  
 (Courtesy H. G. Moller, Max-Planck Institut fur Ionosphären-Physik, Lindau, West-Germany)

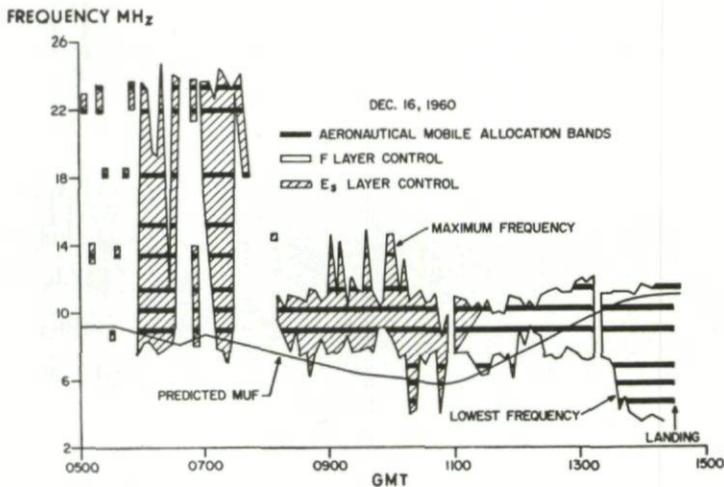
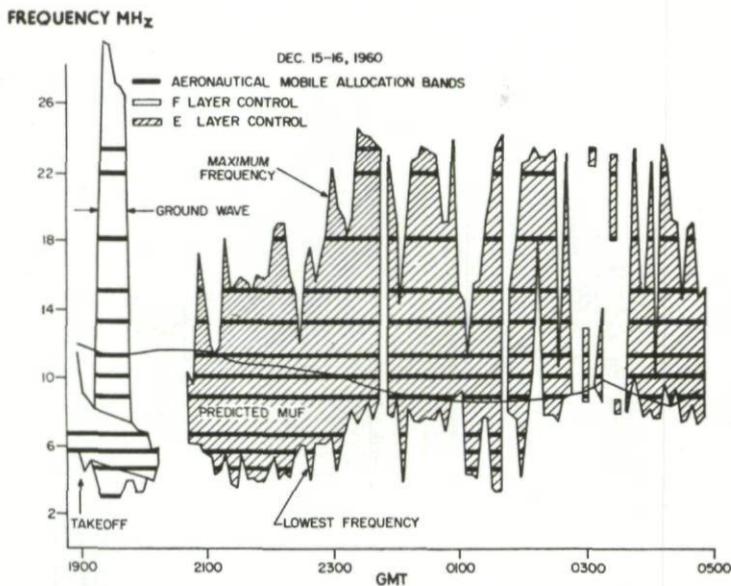


Figure 4.13 The maximum and lowest frequencies observed with the airborne sounding equipment, 15 - 16 December 1960

(a) Outgoing leg of flight

(b) Incoming leg of flight

(After G.W. Jull, D.J. Doyle, G.W. Irvine and J.P. Murray, Proc IRE 50 1676 1962)

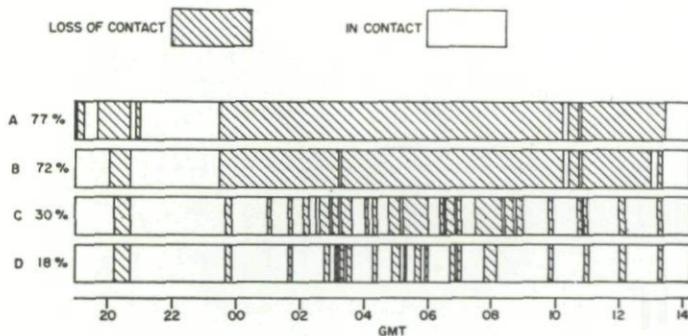


Figure 4.14 Durations of loss of radio contact, 15 - 16 December 1960

(a) Predictions - using a frequency in the band closest to the FOT

(b) Predictions - using frequencies in the two bands closest to the FOT

(c) Frequency sounding - using frequencies in two bands selected each hour

(d) Frequency sounding - using frequencies in all aeronautical bands as required

(After G.W. Jull, D.J. Doyle, G.W. Irvine and J.P. Murray, Proc IRE 50 1676 1962)

## Possibilities of using oblique incidence sounding for prediction purposes

The principles of oblique incidence sounding were briefly described earlier. Two different types of records may be obtained by this method. The first is the oblique incidence ionogram where the travel time of high frequency radio pulses from the transmitter to a remote receiver is recorded as a function of frequency. An example is shown in Figure 4.9. A record of this type can be obtained in a few minutes, and the picture can be displayed on a cathode ray oscilloscope. The most important feature of the record is that the MUF can be read off immediately.

Unfortunately, the LUHF cannot be obtained as easily as the MUF. However, the LOF, that is the lowest frequency at which an oblique echo is observed, will give some information on the total loss over the path and how this loss varies with time. If therefore, an oblique incidence sounder is in operation over a particular propagation path, the radio operator will have good information about the ionospheric conditions along this communication circuit and will be able to choose his working frequency accordingly. The most reliable information will be obtained when the oblique incidence sounder and the communication system use the same antennas.

A few typical oblique incidence ionograms will serve to demonstrate the usefulness of the technique. Figure 4.10 shows three oblique ionograms obtained during a period with strong radio wave absorption. The lowest frequency at which a trace is observed is quite high and varies rapidly with time, and communication in the lowest frequency bands will probably be difficult. However, if the operator can take advantage of the skip distance focusing and select his working frequency close to the MUF he may still get through.

Figure 4.11 shows a record obtained during a period when there was very little ionospheric absorption. The LOF is low, and clear two and three hop traces are present. Near the MUF there will probably be interference between the high and low angle rays, therefore, the operator may be better off by choosing a lower frequency, say about 9 MHz. If he uses an even lower frequency, he may have interference between different modes.

Figure 4.12 shows a rather complex record where the normal one hop F-trace is partly obscured by other traces. On the other hand there is a clear Es-trace which gives echoes on frequencies well above the FJF. In this case, therefore, the best solution would be to use a frequency above the junction frequency, FJF, for example 18 MHz.

In the polar regions conditions can change very rapidly and the operator may have to face situations like the ones sketched above, in rapid succession. It seems obvious that an oblique incidence sounder at both terminals of an HF communication circuit should reduce the time the circuit is out of operation due to ionospheric disturbances. That the method may really be useful was proved by an experiment made in Canada in which an aircraft carrying a small oblique incidence sounder was flown through the auroral zone. The flight was made during a moderate geomagnetic storm accompanied by black out of the entire HF band. During this test the radio operator in the aircraft carried out air-ground-air communications on frequency bands selected from the sounding information. The propagation conditions were extremely variable and it was necessary to run the sounder every five minutes and to change working frequencies as often as once every fifteen minutes.

The test proved that radio contact could be obtained during a large percentage of the time, even during the difficult conditions prevailing at the time. This was due to the fact that the sounder made it possible to take advantage of reflection from clouds of sporadic E along the propagation path. This is illustrated in Figure 4.13 where the maximum and minimum frequencies observed

with the sounder are plotted as a function of time during the flight. The predicted MUF is also shown for comparison. Most of the time the maximum frequency observed is very much higher than the predicted MUF, and the waves are reflected from an extremely variable E<sub>s</sub>-layer.

A comparison of the durations of contact obtained using four different methods was made. The results are shown in Figure 4.14. In cases A and B the available predictions were used, while in cases C and D the information supplied by the sounder was taken into account. As will be seen the aircraft would have been out of contact with the ground station for 77% of the time if only one frequency band based on the predictions was selected. The time with loss of contact decreased to only 18% of the total flight time when the frequencies were selected using frequent oblique incidence sounding.

#### Backscatter sounding

In backscatter measurements the radio signals, reflected back to the transmitter from the ground via the ionosphere and reflected directly from irregularities in the ionosphere, are recorded. A typical example obtained during quiet conditions is shown in Figure 3.6. From the 2F<sub>2</sub>-trace in this record it is, in principle, possible to deduce the MUF for a given distance. In the polar regions, however, the usefulness of backscatter measurements as a means of improving communication reliability is limited. This is because the backscatter records only give information on the ionosphere as far as the first hop, and because the 2F<sub>2</sub>-trace is often obscured by echoes from irregularities in the E- and F-regions. On the other hand, the backscatter experiment is simpler because it does not depend on a receiver at the other end of the propagation path. Also, by employing a rotating antenna, it is possible to obtain information in a wide area around the station.

#### The poor man's oblique incidence sounder

Since oblique incidence sounders are still rather expensive, most radio operators have to get along without this tool. However, by using their receivers in the right way they can to some extent obtain the information supplied by an oblique incidence sounder. An example will serve to illustrate this. Suppose the radio operator on board a ship in the Pacific wishes to communicate with some station in Europe. To select the best working frequency he could listen to one of the large European coast radio stations. These are on the air almost continuously in the frequency bands 4, 6, 8, 12, 16, 22 and 25 MHz, thus covering a large part of the HF band. By selecting a frequency in the band where the radio operator receives the strongest signal, he has a good chance of getting through. If contact is lost, the operator must run his receiver through the frequency range again to determine a new working frequency.

## Appendix 1

### The Earth's Magnetic Field and its Time Variations

The earth's magnetic field plays an important role in radio wave communication, particularly at high latitudes, for three reasons:

- (a) The motion of the ionizing particles is governed by the field
- (b) the irregularities in the ionization causing spread echoes are frequently aligned along the field lines
- (c) the difference in the ordinary and extraordinary modes of propagation is a function of the field.

Some of the main parameters related to the field are reviewed in this appendix.

#### Average field

The earth's magnetic field can be approximated to that of a magnetic dipole with magnetic moment  $M = 8.1 \cdot 10^{25}$  Gauss  $\text{cm}^3$ , centered in the center of the earth. The geomagnetic axis is defined as a line which intersects the earth's surface at the geomagnetic poles ( $78.5^\circ\text{N}$ ,  $69^\circ\text{W}$  and  $78.5^\circ\text{S}$ ,  $111^\circ\text{E}$ ). A plane, normal to the geomagnetic axis, midway between the geomagnetic poles is called the geomagnetic equatorial plane, and the intersection between this plane and the earth's surface is the geomagnetic equator. A plane defined by the geomagnetic axis and a given point on the earth's surface is defined as the geomagnetic meridian plane through the point.

The geomagnetic longitude of the point corresponds to the angle between this meridian plane and the meridian plane through the geographical poles. Thus, the geographic longitudes  $69^\circ\text{W}$ , and  $111^\circ\text{E}$  correspond to geomagnetic longitudes of  $0^\circ$ , and  $180^\circ$ , respectively. The angle between the zenith direction for a given point and the geomagnetic equatorial plane is defined as the geomagnetic latitude of the point.

#### Geomagnetic time

We recall that one solar hour is defined as the time it takes the earth to rotate  $15^\circ$  relative to the sun. In analogy with this definition we define one geomagnetic hour as the time between the solar passage of two geomagnetic meridian planes separated by  $15^\circ$ . 1200 geomagnetic local time for a given point, corresponds to the time when the sun is observed in zenith somewhere along the geomagnetic meridian through the point. Due to the skew orientation of the geomagnetic axis relative to the geographic axis of the earth, the duration of one geomagnetic hour varies over the day and over the year. In spite of this strange property of the geomagnetic clock, it is still a useful concept for studies of high latitude ionospheric phenomena.

#### Variations in the earth's field

The earth's magnetic field undergoes a series of regular and irregular variations. The amplitudes of the regular variations are very small, whereas the irregular variations in the geomagnetic field (geomagnetic storms) are much larger. The time variations in the field intensity during severe storms may be 1-2% of the total field. The two main types of geomagnetic storms are (i) world-wide storm, and (ii) high latitude storms (or auroral magnetic storm). An average world-wide storm has the following, typical trend. The storm normally starts with a rapid, shortlived increase in the field intensity. This is called the sudden commencement (SC). After the SC the field stays above the normal value for some hours during the initial phase of the storm. After this

phase the field drops below its normal value, and the field may stay depressed for days. This decrease in the field is called the main phase of the storm. The sudden commencement is normally explained as due to the shock when the cloud of particles from the disturbed sun collides with the outer boundary of the earth's magnetic field. During the initial phase of the storm the particle cloud compresses the earth's magnetic field lines. Finally, due to the solar storm, an East-West electric current is generated 4-10 earth radii from the centre of the earth in the equatorial plane. This ring current is responsible for the depression in the magnetic field observed at the earth's surface.

The high-latitude geomagnetic storm does not follow the same scheme as the world-wide storm. The high-latitude storm, which shows a very irregular trend, is thought to be due to an electrical East-West current in the 100-150 km altitude region, close to the visual auroral zone. The current is set up during periods when the electrical conductivity at this altitude is high, due to enhanced ionization. The relationship between the geomagnetic activity and auroral phenomena was discussed in chapter 3. The degree of geomagnetic disturbances is normally described in terms of 3-hour planetary K-index ( $K_p$ ), which ranges from 0 (quiet) to 9 (strong disturbances). An alternative measure of geomagnetic activity is the 24-hour A-index, which ranges from 0 to 400, the higher the value the stronger the storm.

# Subject Index

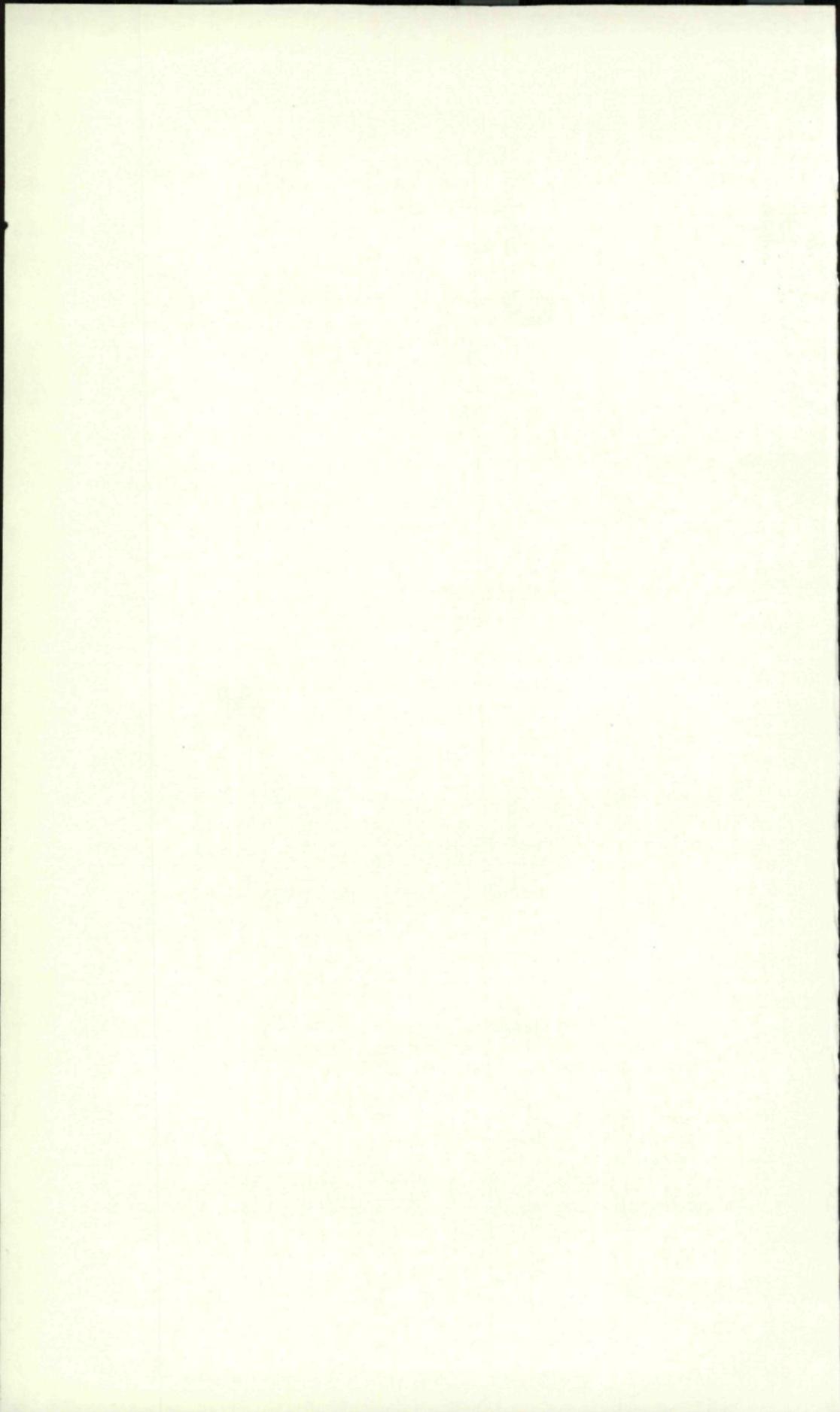
	Page
Absorption	34
Auroral	95
absorbing regions, maximum probability of	96
characteristics, temporal and spatial duration	96
and structure	100
ionization profiles, direct measurements of	95
primary particles, direct measurements of	95
Deviative	14, 41
Non-deviative	14, 41
Frequency dependance of	41
Irregular	
cosmic noise, measurement of	115
main types of	87
Polar Cap	
description of	93
events, duration of	94
importance to radio communication	95
large scale features of	94
occurrence frequency of	94
Radio waves	14
Attenuation	
Due to fading	44
Free space	37
Breit and Tuve	
Theorem	9
Communication	
In other frequency bands	105
VLF waves	105
LF waves	106
MF waves	106
VHF waves	106
Polar difficulties, prediction of	107
long term	108
short term	109
Disturbances	
Available predictions of	111
Characteristics of	109
Effects, solar and terrestrial	108
Flares, solar	109
Position of	109
	127

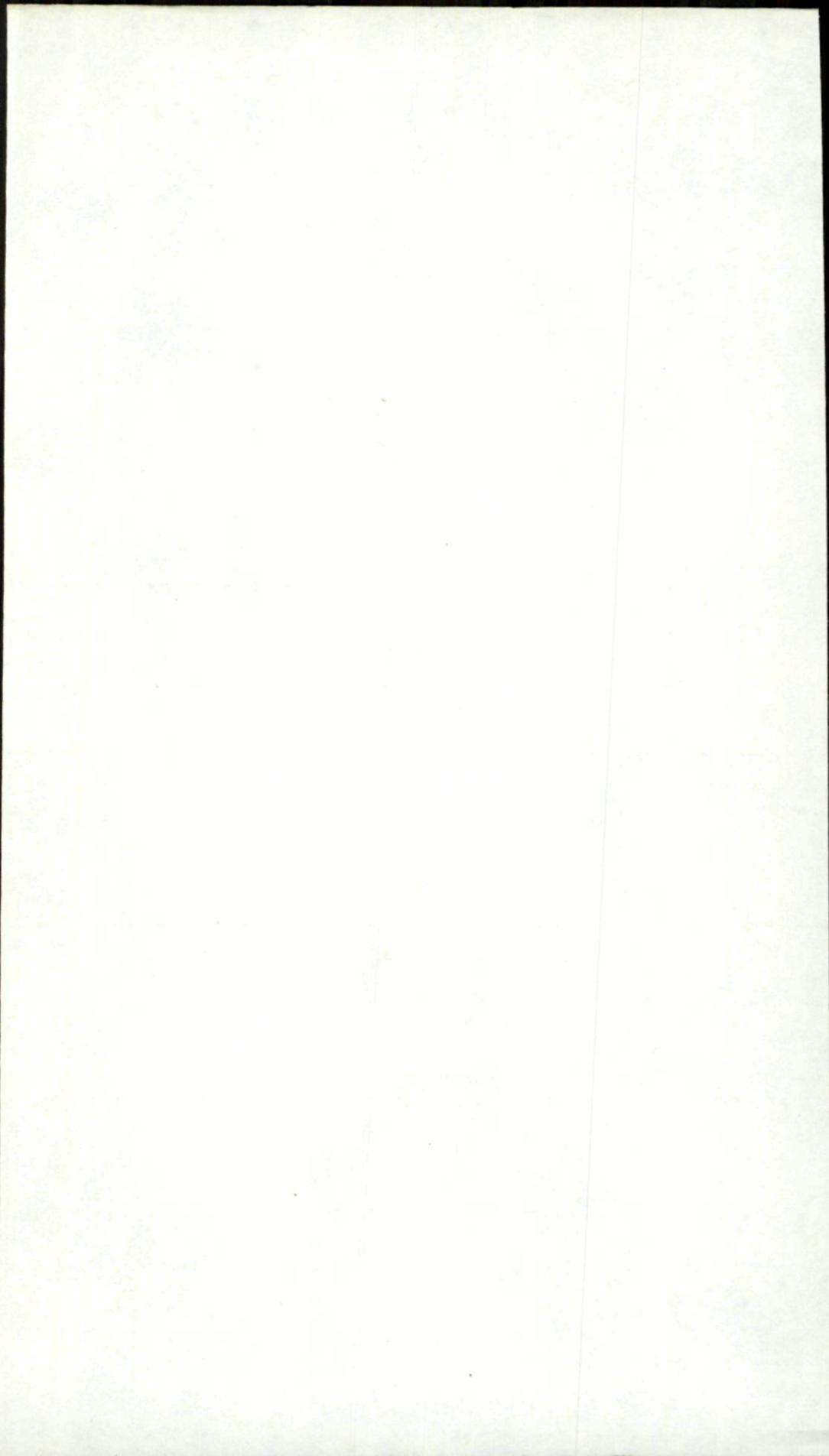
Earth, the	
Field, magnetic	
Average	125
influence on radiowave propagation	15
variations	125
Time, geomagnetic	125
variations	125
Electrons, free	
Production in the ionosphere	16
Loss in the ionosphere	16
Emission	
Irregular solar	107
Fading	
causes of	44
Field strength	
Received	44
Total loss of	44
Frequency limitations	
Method of prediction	53
High Frequency curves	
Vertical incidence of	6
Ionograms, oblique	
notation of	84
Ionosphere	
Absorption, total	43
Arctic, study of	88
D-layer, irregularities	26
region, time variations of	26
E-layer; irregularities	28
right	100
region, time variations of	25
screening by	52
Layers, regular	
methods of observation	19
regular three	17
sustention of	17
time variations of	20
variations of	16
Polar	79
methods of study	83
particles, origin of interaction with earth's atmosphere	80
soundings, oblique	83
vertical incidence	83
Predictions	
conditions of	53
form and presentation of	58
Refracting medium	2
Sporadic E	
auroral, world-wide distribution of	100
Es-layer	100
H.F. propagation, importance to	105
high latitude	100
right E-layer	100
time variations of	104
V.H.F. forward scatter	88

Ionospheric Sounder	
Basic principles of	9
<b>Lowest Useful High Frequency (LUHF)</b>	<b>34, 67</b>
Absorption, ionospheric	38
Ground losses	38
Resultant	48
<b>Martyn</b>	
Theorem	11
<b>Maximum Useable Frequency (MUF)</b>	<b>29, 53</b>
Distances greater than 400 Km	60
Earth's curvature, effect of	33
Factor M, determination of	30
Skip distance	29
Transmission curves	30
<b>Observations</b>	
Riometer, deductions based on results from	115
Optimum traffic frequency	60
<b>Parameters, Physical</b>	
Predictions	53
basic principles of	54
Related, determination of	59
Plasma frequency	6
Predictions	
Failure, selection of frequency when	76
some reasons for	75
Interpretation of	112
Mode considerations	58
Reliability and use of	73
Testing of	73
<b>Radio</b>	
Quality classification	111
Traffic, de-routing of	112
Communication, principle of	1
Noise	34
Atmospheric	46
Cosmic	47
Level	45
Man-made	47
Measurement of	47
Power of	45
Precipitation of	45
Thermal	45
Receiver	
Synchronization with transmitter	84
Refraction	
Law of	1
Refractive index, frequency dependance of	5
Rockets and Satellites	88
	<b>129</b>

<b>Selected circuits</b>	
Examples of	68
Snells Law	1
Soundings	
Oblique incidence and backscatter	116, 124
Possible use for predictions	123
<b>Transmitter</b>	
Power radiated from	34
<b>Wave propagation</b>	
Reflected, use of	87
Refractive medium in	1
Simple experiment, principle of	88









TECHNIVISION