REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-R-744	ISBN 92-835-1539-0	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title
SPECIAL COURSE ON INTERACTION OF PROPAGATION AND DIGITAL TRANSMISSION TECHNIQUES

7. Presented at

8. Author(s)/Editor(s)	9. Date
Various	October 1986
10. Author's/Editor's Address	**11. Pages**
Various	318

12. Distribution Statement This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Electromagnetic wave transmission
Propagation
Data transmission
Digital techniques

14. Abstract

This Special Course has three main objectives: 1. To provide an opportunity for engineers and scientists to learn more of propagation mechanisms and communications system design in frequency bands other than those with which they may be closely associated, thus enabling a "cross-fertilisation" of ideas to take place; 2. To attempt to identify, for the various classes of systems, the propagation measurements/data still required to enable digital communication system design to be made more precise and effective; and 3. To identify trends in the requirements and design techniques for future digital communication systems.

This Special Course, sponsored by the Electromagnetic Wave Propagation Panel of AGARD, has been implemented by the Consultant and Exchange Programme of AGARD, and was presented at Jevnaker, Norway, 13—14 October 1986, at Copenhagen, Denmark, 16—17 October 1986 and in Lisbon, Portugal, 20—21 October 1986.

AGARD-R-744

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD REPORT No.744

Special Course on Interaction of Propagation and Digital Transmission Techniques.

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Report No.744

SPECIAL COURSE ON INTERACTION OF PROPAGATION AND DIGITAL TRANSMISSION TECHNIQUES

The material in this book has been prepared under the sponsorship of the Electromagnetic Wave Propagation Panel and the Consultant and Exchange Programme of AGARD and was presented as a Special Course at Jevnaker, Norway, 13—14 October 1986, at Copenhagen, Denmark, 16—17 October 1986 and in Lisbon, Portugal, 20—21 October 1986.

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Exchanging of scientific and technical information;

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

Published October 1986

ISBN 92-835-1539-0

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

THEME

This Special Course has three main objectives: 1. To provide an opportunity for engineers and scientists to learn more of propagation mechanisms and communications system design in frequency bands other than those with which they may be closely associated, thus enabling a "cross-fertilisation" of ideas to take place; 2. To attempt to identify, for the various classes of systems, the propagation measurements/data still required to enable digital communication system design to be made more precise and effective; and 3. To identify trends in the requirements and design techniques for future digital communication systems.

This Special Course, sponsored by the Electromagnetic Wave Propagation Panel of AGARD, has been implemented by the Consultant and Exchange Programme of AGARD.

SPECIAL COURSE STAFF

Special Course Director: Prof. M.Darnell
Dept. of Electronic Engineering
University of Hull
Hull, HU6 7RX
UK

LECTURERS

Dr F.J.Kelly
Consultant
Ionospheric Effects Branch
Space Science Division
Naval Research Laboratory
Washington D.C. 20375-5000
USA

Mr J.Østergaard
Senior Engineer
Dept. of Telecommunications Engineering
Electronik Centralen
Ven Lighedsvej 4
DK-2870 Horsholm
Denmark

Dr T.K.Fitzsimons
Head, EMC Section
NATO Headquarters
1110 Brussels
Belgium

Dr K-J.Hortenbach
Deutsche Welle
HF-Technik
P.O. Box 100444
D-Köln 1
Federal Republic of Germany

Mr R.Larsen
Propagation Group
Marconi Research Centre
West Hanningfield Road
Great Baddow, Chelmsford
Essex CM2 8HN
UK

Dr A.L.Johnson
Chief, Satellite Communications Group
AFWAL/AAAI
Wright Patterson AFB
Ohio 45433
USA

CONTENTS

*Not available at time of printing.

OVERVIEW OF NATO & NATIONAL REQUIREMENTS FOR DIGITAL TRANSMISSION

M Darnell
Department of Electronic Engineering
University of Hull
Hull HU6 7RX
UK

SUMMARY

The paper provides a general introduction to the scope, objectives and content of the Electromagnetic Wave Propagation Panel (EPP) Special Course on the "Interaction of Propagation and Digital Transmission Techniques".

1. INTRODUCTION

The impetus for this Special Course on the "Interaction of Propagation and Digital Transmission Techniques" arose from an AGARD EPP Specialists' Meeting on "Propagation Influences on Digital Transmission Systems - Problems and Solutions" held in Athens in 1984 (AGARD, 1984). From this meeting, it was apparent that there is still much to be done in relating the design of digital communication systems to the propagation mechanisms encountered in the various radio frequency bands, not least in characterising propagation more precisely and in a manner that can be more readily interpreted by communication engineers.

This course attempts to link, in a coherent manner, the disciplines of propagation measurement and analysis with that of digital communication system design. Six types of communication system are considered in the course:

(i) ELF/VLF/LF systems;
(ii) HF systems;
(iii) meteor-burst systems;
(iv) tropospheric scatter systems;
(v) VHF/UHF/microwave line-of-sight systems;
(vi) UHF/SHF satellite systems.

These systems employ propagation media which exhibit an extremely wide range of characteristics; also, the variety of transmission and control techniques used is considerable, with each band requiring a large measure of individual consideration. However, there are significant areas of commonality, and one of the purposes of this course is to identify such areas.

The course also examines the problems of spectrum management and system compatibility, both within and across the different frequency bands. Problems of electromagnetic compatibility (EMC) may arise, for example, when analogue and digital systems have to co-exist in the same portion of the radio frequency spectrum, or when wideband and narrowband transmission formats seek to employ the same frequency band simultaneously.

An introductory lecture presenting a review of modulation techniques, coding for error control and multiple-access, and speech digitisation procedures is also included for completeness.

The systems chosen for detailed discussion are, with the possible exception of meteor-burst, widely used by NATO and/or individual NATO nations.

2. SPECIAL COURSE OBJECTIVES

The specific objectives of the Special Course are as follows:

(a) To provide a forum for scientists and engineers from the various NATO nations to learn more of propagation effects and digital communication system design in frequency bands other than those with which they may be most familiar. By this means, it is anticipated that there will be a "cross-fertilisation" of ideas, theory and practice.

(b) To attempt to identify, for each class of communication system discussed, the propagation measurements and data that are still necessary to enable digital transmission systems to be designed more precisely and effectively. One important outcome of the availability of a more comprehensive propagation database would be to allow more realistic simulations of propagation and noise/interference effects during the design phase of a communication system.

(c) To demonstrate how propagation and noise/interference effects influence communication system design and the choice of system parameters.

(d) To identify trends in future digital communication requirements and system utilisation, both NATO-wide and nationally.

(e) To provide a forum for informal, and more formal round table, discussion of the

views and outstanding problems of the Special Course attendees.

(f) To provide, for more general distribution and use, a course manual which will be a useful reference document and design aid for both propagation specialists and communication engineers.

3. RATIONALE FOR DIGITAL COMMUNICATION

Over the past two or three decades, there has been a consistent trend towards the introduction of a progressively greater degree of digitisation in communication systems. As part of the background to this Special Course, it is instructive to review the advantages and disadvantages accruing from the use of digital transmission schemes.

3.1 Advantages

The main advantages of digital transmission are:

(a) It allows the use of signal regenerators, rather than repeaters, to reconstitute the transmitted signal. Such reconstitution must be applied before the signal-to-noise (SNR) deteriorates too far so that any errors cannot be corrected via error control coding. In theory, there is no limit to the number of times regeneration can be applied, thus providing an effectively distance-independent SNR. In contrast, analogue repeaters amplify both signal and noise causing the SNR to degrade progressively.

(b) Data in digital form can be fully encrypted giving a theoretically perfect level of security; analogue signals, on the other hand, can at best be given limited security via time and frequency domain scrambling techniques.

(c) Error control coding techniques can be applied to the digital data.

(d) More flexible forms of multiple/random-access coding can be used.

(e) Advantage can be taken of digital signal processing technologies, particularly silicon-based, and also common technologies arising from developments in the computer industry.

3.2 Disadvantages

In addition to the advantages indicated above, digitisation brings a number of disadvantages in comparison with the more classical analogue techniques:

(a) Generally, more bandwidth is required to transmit a digitised version of an analogue signal than for the analogue signal itself. For example, compare the 3 kHz bandwidth required for a typical analogue speech signal with the bandwidth of several tens of kHz for say a single 64 kbits/s pulse-code modulated speech channel. If the digitised speech is to be constrained to the same 3 kHz bandwidth as the original analogue signal, this necessitates considerable complexity in the speech encoder/decoder.

(b) Transmission rates tend to be fixed in a given digital communication system, thus limiting the flexibility and potential for adaptation of the system. Analogue systems are normally more tolerant to relatively wide ranges of noise/interference.

(c) Specific techniques must be incorporated to allow the transmitter and receiver in a digital communication system to be synchronised both prior to, and during, transmission.

3.3 Channel Capacity

For any communication channel, the theoretical error-free capacity, C, is given by

$$C = B \log_2 [1 + (S/N)] \text{ bits/s} \qquad [1]$$

where B is the channel bandwidth in Hz, S and N being respectively the total signal and noise powers within the bandwidth B. Therefore, the channel capacity depends upon just two channel parameters, its bandwidth and SNR.

In practice, the measured capacity of a channel may be several times lower than the theoretical capacity calculated from expression [1]; this discrepancy is due to a number of factors, principally

(i) non-uniform distribution of signal and noise energy within the channel bandwidth B;

(ii) a non-"brickwall" channel bandwidth profile;

(iii) non-uniform amplitude probability density functions for signal and noise.

4. COMMUNICATION SYSTEMS DISCUSSED IN THE SPECIAL COURSE

As a brief introduction to the more detailed lectures of the course, the nature of the operational utilisation for the six types of communication systems discussed is now outlined.

4.1 ELF/VLF/LF Communication Systems

The primary operational importance of the ELF/VLF band below about 30 kHz is that it provides a means of transmitting information directly over very long distances to submersibles. The band is used by several nations for primary, low-rate, digital broadcasts to submarines and also for navigation purposes, eg OMEGA. Because of fundamental environmental limitations on antenna effectiveness, it is not possible to use the same band for direct transmission from a submersed platform; other techniques and frequency bands are normally employed for this purpose.

Since there are a relatively large number of operational requirements, and only a very limited total spectral width is available to fulfill these requirements, frequency assignments are rigidly controlled with little bandwidth being available for any given user. Such control is particularly vital because of the stable and world-wide nature of propagation in this frequency range. Bandwidth restrictions and the phase stability of the medium have tended to encourage the use of bandwidth-efficient modulation schemes and multiplexing for broadcast purposes.

For transmitting antenna structures to be reasonably efficient radiators of electromagnetic energy, their physical dimensions must be an appreciable fraction of the wavelength of the transmission. In the ELF/VLF bands, therefore, antennas tend to be very large and consequently physically vulnerable.

The LF band up to say 300kHz is used primarily for long-range broadcasts, ship-shore operations and navigational purposes. Propagation ranges are more restricted than at ELF/VLF, but antenna sizes are smaller and power radiation efficiencies somewhat greater.

4.2 HF Communication Systems

HF (2 - 30 MHz) systems represent a primary means of communication for some NATO nations, and a secondary means for others. NATO itself deploys significant HF assets in Allied Command Europe (ACE).

The chief advantage of the HF medium is that of flexibility. It is possibly the only viable means of long-range (in some cases world-wide) bi-directional communication, apart from satellite-based techniques; it can, however, also be employed at all ranges down to line-of-sight (LOS) by both static and mobile terminals. Both surface wave (often termed groundwave) propagation and propagation via ionospheric refraction (skywave) modes are used operationally.

HF communication systems are currently used in the following roles by NATO and the NATO nations:

(a) broadcast services;
(b) point-to-point and transportable land-based services;
(c) ship-ship services (primarily via surface wave);
(d) ship-shore services;
(e) air-ground services;
(f) combat net radio (CNR) services (primarily via surface wave).

The HF band is difficult to use efficiently and reliably because of the variable nature of propagation and the high and unpredictable levels of co-channel interference. One of the main problems associated with HF systems is a tendency to impose unrealistic operational requirements upon them, eg duplex secure voice with low-power mobile terminals. The HF medium has unique characteristics and a variable transmission capacity which must be fully appreciated if it is to be used effectively; it should not be viewed simplistically as an alternative or replacement for a 3 kHz-bandwidth line or satellite channel.

In the low probability of intercept/anti-jamming (LPI/AJ) context, HF communications has several potential advantages, not least the natural geographical screening and selectivity afforded by both surface wave and skywave propagation modes.

4.3 Meteor-Burst Communication Systems

Although meteor-burst systems have been the subject of considerable research and development activity for the past two to three decades, as yet they are not deployed in any defined military roles. Many working systems have been demonstrated, and there are a number of fully operational civil systems, see for example (Western Union, 1977).

Meteor-burst systems can provide beyond line-of-sight digital communications at distances ranging typically from a minimum of a few hundreds of kilometres to a maximum in excess of 2000 km. The propagation is such that the meteor trails will provide a virtually unique propagation link between two points on the earth's

surface, with only a very small "footprint" where interception is possible around the receiver terminal; the technique therefore has considerable LPI/AJ potential.

Operationally, it would appear that this means of propagation is best suited to the following roles:

(i) point-to-point links;
(ii) broadcast transmissions;
(iii) situations in which a master stations polls a number of outstations in time sequence.

The statistical occurrence of ionised meteor trails is relatively predictable; they provide a propagation mechanism which is intermittent in nature with the majority of trails being usable for less than a second. However, in that short interval, data can be transferred at relatively high rates. "Waiting times" between trails are of the order of a few tens of seconds.

Of late, meteor-burst communication in the low VHF range above 30 MHz has been seen as a potential adjunct to HF communication systems, eg by providing an additional broadcast component or possibly greater operational flexibility on a point-to-point circuit, say under disturbed ionospheric condidtions.

4.4 Tropospheric Scatter Communication Systems

Until relatively recently, tropospheric scatter has been regarded as a primary analogue communications medium for FDM voice transmission within NATO, as evidenced by its use as the backbone of the ACE HIGH system. However, it is now tending to become de-emphasised, with the trend now being towards the interconnection of national communication assets of various types rather than the use of a dedicated NATO network of the ACE HIGH type. These national assets may well incorporate a number of troposcatter links, but will also include VHF, UHF and microwave radio relay components.

It is foreseen that tropospheric scatter systems will continue to be employed in somewhat specialised and restricted roles, ie for

(i) medium-range links across water;
(ii) medium-range point-to-point links;
(iii) "tail" circuits to islands;
(iv) cross-border interconnection links.

It is also probable that more tactical, lower capacity, troposcatter equipments will continue to be developed and deployed. There is an increased interest in the more anomalous troposcatter propagation modes inasmuch as they may influence the performance, and extend the range, of nominally LOS systems, giving them a beyond-the-horizon capability.

In general, troposcatter systems can be engineered for high reliability, with the normal propagation mechanisms being described by comprehensive and reliable statistical models. An increased emphasis upon digital methods of transmission in modern troposcatter systems has stimulated the development of adaptive, high-rate modems using equaliser structures capable of responding dynamically to changes in the propagation environment. Such techniques, coupled with operational requirements for the same, or more, overall transmission capacity as provided by the previous analogue systems, impose increased bandwidth demands which are difficult to fulfill in the spectrally-congested European environment.

4.5 VHF/UHF/Microwave Line-of-Sight Communication Systems

VHF/UHF/microwave LOS systems fulfill a primary role for general voice and data communications in both static and mobile scenarios in NATO and national operations, as well as in more specialised requirements such as JTIDS. As was indicated in the previous section, these means of communication are also tending to replace troposcatter systems in some situations. Propagation modelling for such systems is relatively straightforward, although the effect of the ground terrain upon the nature of propagation requires precise characterisation for accurate system performance predictions.

Increased bandwidth requirements arising from the progressive introduction of digitised systems, together with greater numbers of requirements and systems deployed, have tended to add to the already high level of occupancy in this region of the spectrum; consequently, the existing spectrum allocated for these services has had to be more effectively controlled in order to maximise frequency re-use and minimise channel spacing. The suitability of higher frequencies at 30 GHz and above for LOS services is also being investigated, although there appear to be problems associated with the ease of power generation for jamming purposes at these frequencies. In addition, meteorological effects have a greater influence on system performance as the frequency of operation is increased.

The "mix" of LOS communication systems is becoming more diverse, with analogue systems having to co-exist with digital transmissions, and conventional narrowband links having to operate in conjunction with spread-spectrum formats. This diversity

of systems again adds to the problems of spectrum control and management.

There is obviously considerable activity associated with the development of LOS systems in the civil field, eg for cellular radio networks. It is possible that these civil developments may well dictate the technology to be used in future military systems to meet certain requirements.

4.6 UHF/SHF Satellite Communication Systems

For the past 20 years, satellite systems have provided the primary, beyond-LOS, high-capacity links required for many NATO and national purposes. Broadly speeaking, SHF systems have been used mainly for strategic operations and UHF systems in the more tactical environment, with the latter suffering more from propagation variability. Satellites provide a highly reliable, but potentially physically vulnerable, means of communication; the majority of the defined NATO operational region can be covered by means of a geostationary satellite, although there may be some coverage limitations in the high-latitude zones. At UHF, there are significant scintillation effects in both equatorial and polar regions.

The accuracy of propagation modelling for SHF satellite systems is high, although there are deficiencies in data relating to the effects of high-altitude nuclear bursts, dust-induced noise, scintillation mechanisms, etc. Again, the trend is towards the progressively greater use of digital transmission techniques to provide security, multiple-access capability and jamming-resistance. Higher frequency systems are also being investigated.

4.7 Spectrum Management & System Compatibility

From the preceding sections of this paper, it can be seen that the problems of spectrum management and of the electromagnetic compatibility (EMC) of communications systems are becoming increasingly severe in nearly all frequency bands to be covered in this Special Course. The problems stem from

(i) an increasing number of communications requirements;
(ii) a greater number of systems deployed in all bands;
(iii) a greater variety of systems types within a band, eg digital and analogue, narrowband and wideband, etc;
(iv) more precise propagation and system modelling leading to reduced channel spacing, greater frequency re-use, polarisation discrimination, etc.

The discipline of communications EMC in a multi-user environment is, as yet, in its infancy. Propagation models do exist which allow the potential interactions between systems to be predicted with an accuracy sufficient for current frequency assignment purposes. However, more subtle system interactions, jamming effects and the effects of abnormal propagation mechanisms cannot yet be characterised with the same degree of confidence.

Spectrum management and EMC considerations are likely to become progressively more important in the future, and will require greater EMC expertise to be introduced in the system design and operational planning phases. Also, additional data will have to be gathered to allow more realistic modelling of interactions.

The main body of the Special Course material commences with the review of modulation, coding and speech digitisation techniques.

5. REFERENCES

1. AGARD, 1984, "Propagation influences on digital transmission systems - problems and solutions", AGARD CP No.363.

2. Western Union Telegraph Company, 1977, "Western Union's meteor burst communications", Western Union, McLean, VA.

REVIEW OF MODULATION, CODING & SPEECH DIGITISATION TECHNIQUES

M. Darnell
Department of Electronic Engineering
University of Hull
Hull HU6 7RX
UK

SUMMARY

As a basis for the more detailed discussion of specific forms of digital communication system in this Special Course, this lecture introduces the topics of

(a) modulation
(b) coding for error control
(c) coding for multiple-access
(d) coding for encryption
(e) speech digitisation

in general terms.

1. INTRODUCTION

In the previous lecture, the major advantages and disadvantages of digitisation in the general communications system context were outlined. In this lecture, the specific techniques involved in the implementation of many types of digital communication systems are reviewed briefly. It will be shown in this Special Course that the range of propagation and noise/interference effects in the various frequency bands is very large. To a great extent, it is these effects which influence the nature of the transmission techniques employed in any given digital communication system. Therefore, not all systems will use all the techniques introduced here, but all are required for the whole range of digital communication systems discussed in this course.

The main topics to be reviewed are as follows:

(a) modulation
(b) coding for error control
(c) coding for multiple-access
(d) coding for encryption
(e) speech digitisation.

Figs.1(a), (b) and (c) show respectively schematic diagrams of three important classes of practical communication system architecture, ie

(i) point-to-point
(ii) broadcast
(iii) multiple-access.

Fig.2 depicts what might be termed a "generalised communication system"; it represents functionally the processes which occur in passing information from a given source to a specified destination in any of the forms of communication system shown in Fig.1. each of the elements of this generalised system will now be discussed briefly.

1.1 Source

The information source may be analogue or digital and will generally comprise two components, "information" which is unpredicatable and "redundancy" which is largely predictable. For example, in speech many of the characteristics of the signal are to some extent predictable and not strictly essential for the comprehension of the information content; variations in loudness, pitch, etc., can be substantially eliminated and yet the speech remain intelligible. Similarly, the higher frequency components of speech above say 3 kHz can be removed by filtering without compromising the speech intelligibility.

1.2 Source Encoder

The function of the source encoder is to modify the characteristics of the source signal in such a way that it is in a better format for transmission over the available communication medium, or can better meet the defined operational requirement. For example, the process of analogue-to-digital (A/D) conversion is a form of source encoding which can be applied to an analogue speech signal to enable it to be passed over a digital transmission system. Similarly, the processes of encryption of digital data or "scrambling" of analogue speech signals are again forms of source encoding which enable the data to be made more secure and hence better able to meet an operational requirement for source information protection.

A source encoder can also perform the function of data compression, ie reduction or

removal of the redundancy present in the original source signal - as is the case with a vocoder applied to a speech signal (see Section 6 of this paper).

1.3 Channel Encoder

The function of the channel encoder is to condition the source encoder output in a manner that will enable it to better withstand the types of noise and perturbation which it is likely to experience during transmission over the communication channel. In order to do this, the channel encoder typically adds redundancy to the source encoder output, but now in a form specifically designed to counteract the channel noise and distortion by allowing their effects to be partially or wholly removed at the receiver.

1.4 Channel

The channel itself will normally impose various forms of noise and perturbation upon the transmitted signal, such as

(a) multipath
(b) time dispersion
(c) frequency dispersion (Doppler spreading)
(d) additive Gaussian or non-Gaussian noise.

Consequently, the signal arriving at the receiving terminal will be a corrupted version of the transmitted signal.

1.5 Channel Decoder

The channel decoder takes the corrupted received signal and operates upon it to remove the effects of the imposed noise and distortion by making use of a knowledge of the channel encoding procedures employed at the transmitter site.

1.6 Source Decoder

The source decoder carries out the complememtary processes to those of the source encoder, eg it may simply take the form of a digital-to-analogue (D/A) converter to return a digital representation of speech back to an analogue form; alternatively, it may re-introduce redundancy eliminated by a data compression algorithm.

Fig.3 summarises, in diagramatic form, the functions of the source encoder and channel encoder. It is seen that a source, initially comprising information I and redundancy R1, is operated upon by the source encoder to produce a compressed/modified source signal in which the information I has been retained but the redundancy has been reduced to a value R2. Subsequently, the channel encoding process adds redundancy to the source encoder output to bring it up to a value R3, with the additional redundancy now being in a form to combat the channel noise and distortion.

In general, it can be said that the more that is known about the characteristics of the source, the more efficient can the source encoding procedures be made; similarly, the more that is known about the characteristics of the channel, the more effective can the channel encoding procedures be made.

In this lecture, the topics of modulation, error control coding and coding for multiple-access can be viewed as channel encoding functions; coding for encryption and speech digitisation on the other hand represent source encoding functions although, in some instances, the distinction between the two forms of encoding is somewhat imprecise.

2. MODULATION

In the following, it will be assumed that the basic digital modulation schemes, ie amplitude-shift keying (ASK) - sometimes referred to as on-off keying (OOK), frequency-shift keying (FSK), phase-shift keying (PSK) and differential phase-shift keying (DPSK) are reasonably familiar concepts. For convenience, Figs.4(a) and (b) summarise the error probability characteristics of these basic modulation types for binary transmission, with coherent and non-coherent detection as appropriate. Fig.4(a) describes steady signal conditions whereas Fig.4(b) gives the performance under conditions of Rayleigh fading. It is seen that the performance of all schemes depends essentially upon one system parameter, ie the received signal-to-noise ratio (SNR) (see for example (Stein & Jones, 1967)).

The factors affecting the choice of modulation technique are:

(a) available bandwidth
(b) electromagnetic compatibility (EMC) constraints
(c) time and frequency dispersion
(d) noise and interfering signal characteristics
(e) rates of change of channel parameters
(f) requirements for interaction with other digital signal processing

techniques.

The significance of each of these factors will now be examined individually.

2.1 Available Bandwidth

Generally, increases in the carrier frequency value will be accompanied by increases in the bandwidth available for any given service. Thus, in the ELF/VLF/LF bands, bandwidths of a few tens of Hz may be assigned; satellite or line-of-sight (LOS) microwave systems would typically have bandwidths of several MHz, whilst optical fibre transmission systems may well operate with bandwidths approaching a GHz.

Clearly, it is important to use bandwidth-efficient modulation schemes, ie those which maximise the number of bits/s/Hz, in situations where bandwidth is limited. An example of this type of modulation is minimum-shift keying (MSK) (Pasupathy, 1979) which is now used extensively for communication purposes over phase-stable channels.

Fig.5 illustrates the principle of MSK modulation; two, time-offset, quadrature sub-channels, I and Q, at the same frequency w_c, are independently phase-reversal modulated by binary data streams $a_i(t)$ and $a_q(t)$ (either +1 or -1) derived from the original single data source. The transitions between the different phase states for both I and Q sub-channel signals $m_i(t)$ and $m_q(t)$ are sinusoidally weighted so that, after combination, they form a constant amplitude MSK signal, m(t). Therefore

$$m(t) = m_i(t) + m_q(t) \qquad [1]$$

$$= a_i(t) \cos [(\pi t)/(2T)] \cos w_c t + a_q(t) \sin [(\pi t)/(2T)] \sin w_c t \qquad [2]$$

which can be simplified to give

$$m(t) = \cos [w_c t + b(t)[(\pi t)/(2T)] + \phi(t)] \qquad [3]$$

where

$$b(t) = -a_i(t)\, a_q(t) \qquad [4]$$

and

$$\phi(t) = \begin{cases} 0; & \text{if } a_i(t) = +1 \\ \pi; & \text{if } a_i(t) = -1 \end{cases} \qquad [5]$$

MSK may be considered as a special case of continuous-phase FSK, with the carrier deviation equal to half the bit rate of the modulating signal - the minimum theoretically achievable.

It should be noted that

(a) the composite MSK signal has a constant amplitude envelope;

(b) the RF carrier is phase-continuous at the symbol transitions, and the phase changes linearly over the symbol intervals;

(c) MSK may be viewed alternatively as shaped offset quadrature PSK (QPSK).

The bandwidth efficiency of MSK, as compared with conventional QPSK, may be evaluated by comparing the bandwidths within which say 99% of the total transmitted power is contained. For MSK, this bandwidth

$$B \doteq (1.2)/T \text{ Hz} \qquad [6]$$

whilst for QPSK

$$B \doteq 8/T \text{ Hz} \qquad [7]$$

However, MSK has a substantially wider main spectral lobe than has QPSK.

A number of variants of the MSK principle have now been developed, some of which appear to offer comparable, or better, bandwidth efficiency.

In situations where bandwidth is relatively plentiful, eg with satellite and other LOS systems, there is an increasing trend towards the use of modulation schemes that deliberately expand the transmission bandwidth. Such schemes, known as spread-spectrum techniques (Dixon, 1975), can be employed to overcome non-linearities in the communications system elements, to overcome high levels of interference or jamming, to make interception difficult, or to provide multiple-access facilities. Spread-spectrum systems fall into two basic categories:

(i) direct-sequence spread-spectrum (DSSS) - where the carrier signal is subjected to digital phase modulation at a high rate;

(ii) frequency-hopped spread-spectrum (FHSS) - where the instantaneous signal frequency is changed rapidly over a wide spectral range.

Figs.6(a) and (b) show diagramatically the principles of DSSS processing. At the input to the spread-spectrum (correlation) receiver, the ratio of the total wanted signal power, S, to the total unwanted noise/interference power, N, is

$$[S/N]_i \doteqdot S/(N_0 \Delta f) \qquad [8]$$

where N is the average noise power spectral density. At the correlation receiver output, the corresponding SNR is

$$[S/N]_0 \doteqdot S/(N_0 \delta f) \qquad [9]$$

Thus, the processing gain is

$$G_p = [S/N]_0/[S/N]_i \qquad [10]$$

$$\doteqdot \Delta f/\delta f \qquad [11]$$

Hence, in principle, large processing gains can be achieved if the ratio of spread-spectrum transmission bandwidth to information bandwidth is high. In practice, receiving system noise and implementation constraints will limit the gain obtainable.

In a FHSS system, assuming a uniform noise power spectral density, the processing gain can be shown to be (Utlaut, 1978)

$$G_p = n \qquad [12]$$

where n is the total number of hopping sub-channels employed.

2.2 EMC Constraints

By EMC is meant the ability of a given electromagnetic (EM) system to operate in conjunction with other EM systems without being adversely affected by emissions from those systems, and without itself generating emissions which will adversely affect the operation of those other systems. In the context of digital communication systems, the important characteristics of the modulation scheme which affect its EMC properties are

(i) its effective occupied bandwidth

(ii) the level of sideband energy outside the nominal assigned bandwidth

(iii) its tolerance to noise/interference.

Property (ii) will form a source of co-channel interference for adjacent channel signals.

In certain regions of the radio frequency spectrum, notably the VHF and UHF bands, assigned channel separations have been progressively reduced thus placing more stringent constraints upon the allowable energy "overspill" from one channel to the next. Here again, spectrally-efficient modulation schemes such as MSK offer advantages - providing the channel characteristics are stable enough to support that form of modulation. Spread-spectrum encoding represents a form of modulation which is particularly resistant to narrowband noise/interference sources (property (iii) above).

2.3 Time and Frequency Dispersion

The degree of time and frequency dispersion imposed by the propagation medium on the transmitted signal is possibly the most important factor affecting the choice of modulation technique. For simple transmission schemes, the inter-symbol interference due to multipath and time dispersion dictates the minimum symbol interval. It is for this reason that medium-speed (1.2 -2.4 kbits/s) HF data modems have tended to employ multiple, low-rate, parallel sub-channels in which the maximum expected multipath spread is small compared with the symbol interval.

In the frequency domain, the phase instability, or Doppler spreading of the spectrum, of a signal must be taken into account when selecting a modulation format. Here, it should be emphasised that the rate of change of signal phase ϕ with respect to time, $d\phi/dt$, has the dimensions of frequency but does not necessarily represent a steady frequency translation in either a positive or negative sense; it is quite possible for a received signal to have a large value of $d\phi/dt$, but to have zero mean frequency offset from its true value.

As an example, a PSK transmission would be inappropriate over a path exhibiting large random phase fluctuations, since it normally requires an accurate long-term phase reference to be derived from the incoming signal; DPSK, in which the phase reference for the current symbol is derived from the previous symbol phase, may be appropriate if the short-term phase stability of the medium is reasonable.

Frequency/phase instability and offsets may also arise as a result of the characteristics of the communications equipment comprising the system and must be taken into account in the overall design.

Therefore, the basic choice which has to be made in a given situation is whether a coherent (synchronous) or non-coherent (asynchronous) modulation and detection scheme should be employed; this choice will be influenced primarily by the phase and/or frequency stability of the propagation medium. Evidently, phase-stable paths such as ELF/VLF/LF terrestrial and SHF satellite links can support phase-coherent modulation methods, eg MSK, PSK and QPSK. In contrast, a relatively unstable medium such as HF must normally utilise non-coherent ASK or FSK, with DPSK also being a possibility.

In a situation where time and/or frequency dispersion is significant, it is normally possible to reduce the detrimental effects at the receiver by making use of one or more forms of diversity processing, eg

(a) space diversity
(b) polarisation diversity
(c) frequency diversity
(d) time diversity
(e) multipath diversity
(f) geographical diversity.

2.4 Noise & Interfering Signal Characteristics

The performance of a given modulation scheme will be very dependent upon the nature of the noise and interfering signal corruption introduced during transmission over the channel. Classical comparisons of the performance of different digital modulation types are typically computed for say steady signals and Rayleigh fading signals in a background of Gaussian white noise - simply because such calculations are relatively simple. However, these idealised propagation and noise/interference models are rarely completely representative, certainly in respect of the type of noise present in many communication systems. For example, in the ELF/VLF bands the atmospheric noise may often be modelled more accurately as log-normal; at HF, manmade interference tends to be the dominant source of degradation; in the VHF band with urban mobiles, noise from electrical machinery and appliances together with car ignition noise is important. In all three of these examples, the noise processes are fundamentally non-Gaussian and may well affect specific modulation types in a somewhat unpredictable manner. There is no guarantee that a performance ranking order in GWN will be valid in a non-Gaussian environment.

If possible, a reasonable model of the noise/interference environment should be derived from practical measurement and used in any system simulations carried out at the design stage.

2.5 Rates of Change of Channel Parameters

The rates of change of channel parameters are related to the dispersion and noise/interference characteristics introduced previously. If the frequency-selective fading properties or the noise profile of a channel change significantly with time, it may well be that a modulation scheme, which initially provides a satisfactory level of performance, will subsequently yield an inadequate grade of service. Hence, it will be necessary to adapt the modulation format or, alternatively, the processing procedures employed at the receiver in order to compensate for the changes. Examples of this include the adjustment of the tone frequencies of an FSK transmission (Darnell, 1979) or modems utilising equalisers which adapt to the prevailing multipath structure (Monsen, 1983). In the latter case, the rates of change of channel parameters will determine whether or not the adaptation algorithms can perform correctly and track parameter changes.

Again, as a design aid, the need for simulation models which accurately reproduce rates of change of propagation and noise/interference is evident.

2.6 Interaction with Other Digital Signal Processing Techniques

It is often necessary to interface modems with other digital signal processing techniques. Possibly the most important interaction of this type is between a demodulator and an error control decoder of the form to be outlined in Section 3 of this lecture. Here, information from the demodulator is employed to enhance the error detection and/or correction capability of the decoder via the vehicle of "soft-decision" data.

Fig.7 illustrates the principle of soft-decision decoding. A normal binary demodulator makes a decision as to whether the received signal is a +1 or -1 according to the position of the analogue demodulator output with respect to a threshold level; this process is termed "hard-decision" decoding. A soft-decision associates a confidence level with each hard decision; in Fig.7, the soft-decision levels (shown in brackets) represent the average received signal level duing the hard-decision intervals, ie the distance from the threshold. This soft-decision data can be incorporated into the error control decoder algorithm to increase its capability relative to a hard-decision decoder. Other parameters of the received signal, apart from its amplitude, can be used as soft-decision parameters, eg phase

margins with a PSK or DPSK signal.

Modulation techniques may also interact with multiple-access schemes (see Section 4) in say satellite or terrestrial LOS communication systems. For example, in a non-linear satellite transponder, intermodulation can limit the performance of a conventional frequency-division multiple access (FDMA) procedure. Spread-spectrum multiple-access (SSMA), often known as code-division multiple-access (CDMA), involving DSSS modulation, can alleviate the effects of non-linearity by effectively randomising the intermodulation product energy across the transponder bandwidth.

3. CODING FOR ERROR CONTROL

As stated in Section 1, error control coding, alternatively known as "error protection" coding or "error detection and correction (EDC)" coding, is a form of channel encoding and involves the addition of deterministic redundancy to the data to be transmitted. At the receiver, a knowledge of the structure of this imposed redundancy will enable defined classes of transmission errors to be detected and possibly corrected. Error control codes fall into two basic categories:

(i) block codes
(ii) convolutional codes.

These will now be discussed individually.

3.1 Block Coding

In a block coder, input digits are taken k at a time; c parity check digits are then computed from combinations of these k digits, giving an overall codeword length of n digits where

$$n = k + c \qquad [13]$$

The resulting set of codewords is termed an (n, k) code and has a code rate, or efficiency, R, given by

$$R = k / n \qquad [14]$$

An important parameter of a block code is its minimum "Hamming distance", d, which is defined as the minimum number of digit positions by which any two codewords, taken from a given code, differ; it is thus a measure of the uniqueness or confusability of codewords in a code, and is clearly related to the error detection and correction potential of a code. The greater the Hamming distance, the lower the probability that a given codeword will be transformed by transmission errors into another valid codeword. In general, if a code is to detect e errors, its minimum Hamming distance must be at least

$$d = (e + 1) \qquad [15]$$

Also, if the code is to correct t errors, the Hamming distance must be at least

$$d = (2e + 1) \qquad [16]$$

Block code designs exist for the detection and correction of many different types of error patterns, from isolated random errors to long error bursts (Lin & Costello, 1983). It is vital that the statistical nature of the errors which are likely to occur on a given propagation path is characterised prior to the application of an error control scheme; otherwise, a coding procedure that is badly matched to the channel error characteristics can actually make a bad situation even worse and cause a further increase in overall error rate. As mentioned in Section 2.6, soft-decision data can be incorporated to enhance the performance of the error control decoder.

It is possible to use block codes in an open-loop, or forward error correction (FEC) mode; alternatively, a closed-loop, or automatic repeat request (ARQ) mode may be employed. The latter is particularly valuable when the error characteristics of the channel vary considerably with time, eg as with some HF paths; the data to be transmitted is encoded with a relatively high-rate block code, principally designed for efficient error detection. If any block is found to be in error on reception, a repeat transmission of that block is requested via a feedback link from receiver to transmitter. The technique can be given greater potential for adaptation if the power of the block code can be varied in accordance with observed channel conditions; ie, if the number of repeats requested in a specified time interval exceeds a certain threshold value, a more powerful lower-rate block code is selected (Goodman & Farrell, 1975).

ARQ techniques have also been applied to intermittent meteor-burst channels, where the system continuously requests repeats, except during those relatively short periods when the propagation path exists and data can be transmitted at high rate (Bartholomé & Vogt, 1968).

Other error control techniques, based upon a data block structure, include

(a) Time-diversity modems where effectively a simple repetition code is used with majority voting, with the same information being transmitted on a number of frequency-spaced sub-channels, each offset by different delays with respect to the others (McCarthy, 1975). An input data buffer is required to produce the desired time offsets, thus introducing a delay into the transmission process.

(b) The CODEM modem, where a 2-dimensional array code is employed in conjunction with soft-decision data in a multiple sub-channel transmission format (Chase, 1973).

(c) Interleaving in which the input data is buffered and re-ordered prior to transmission, thus allowing error bursts to be randomised at the receiver by a complementary de-interleaving process. A simple random error block code will then suffice, rather than a more complex burst-error correcting code. Again. an additional transmission delay will be introduced by the need to buffer data at both the transmitter and receiver (Douglas & Hercus, 1971).

3.2 Convolutional Coding

Whereas block coding derives its c parity checks from a single block of k input digits, convolutional coding derives parity checks from several blocks of input digits.

Fig.8(a) shows a schematic diagram of a convolutional encoder; the term "convolutional" is used because the output data stream may be viewed as resulting from the convolution of the input data stream and the response function of the encoder, as is the situation with a linear time-invariant system (Lee, 1960). In the case of the convolutional encoder, the system may also be described as "linear" if the encoding process only involves modulo-2 arithmetic operations in a binary system.

When the input is applied to the encoder, it stores h blocks of k_o digits; as each new block of k_o digits is accepted, the oldest stored block of k_o digits is discarded. For each unique set of hk digits, n encoded digits are computed, where

$$n_o > k_o \qquad [17]$$

Thus, the encoding process can be considered to have an input constraint length

$$L_i = h\, k_o \qquad [18]$$

and an output constraint length of

$$L_o = h\, n_o \qquad [19]$$

Hence, the code rate , R, or its efficiency, can be specified by

$$R = (h\, k_o)/(h\, n_o) \qquad [20]$$

$$= (h\, k_o)/n \qquad [21]$$

$$= k_o/n_o \qquad [22]$$

where

$$n = h\, n_o \qquad [23]$$

and may be viewed as being equivalent to the codeword length of a block code.

Fig.8(b) shows an example of a convolutional encoder where

$$\left.\begin{array}{l} k = 1 \\ h = 2 \\ n = 2 \end{array}\right\} \qquad [24]$$

For this implementation

$$R = k_o/n_o = 0.5 \qquad [25]$$

and

$$n = h\, n_o = 4 \qquad [26]$$

Ideally, in order to decode a convolutional code, all the received digits should be processed to find the most probable sequence of transmitted code digits. However, certain procedures such as sequential decoding (Wozencraft & Jacobs, 1965) and Viterbi decoding (Viterbi, 1967) have been developed in which a subset of the received digits is analysed at any time, thus making them computationally efficient in comparison with the exhaustive decoding methods. The loss of performance using these sub-optimum decoding algorithms can be made practically insignificant.

Convolutional codes have been applied to many different types of communication channels since they can be designed to counteract a variety of error characteristics.

Their primary use would seem to be in the area of satellite systems.

4. CODING FOR MULTIPLE-ACCESS

With the progressive increase in the number of users in all bands of the radio frequency spectrum, it is essential that the available frequency resources are used as efficiently as possible. Multiple-access coding techniques can contribute to increased efficiency by enabling a number of independent communications links to be established simultaneously over a single propagation channel. Multiple-access techniques are particularly appropriate when a number of "tail" circuits from different users concentrate at a given physical location where they can be combined and subsequently transmitted via composite modulation of a single carrier or carriers. VLF/LF/MF/HF broadcast systems are examples of this unidirectional mode of operation. Alternatively, several distinct duplex circuits may wish to make use of the same portion of the RF spectrum without interaction, as is the case with a satellite system. The choice of multiple-access techniques is influenced largely by the nature of the propagation path and the characteristics of the communications system equipment.

Three basic forms of multiple-access procedure are commonly used:

(a) time-division multiple-access (TDMA)

(b) frequency-division multiple-access (FDMA)

(c) code-division multiple-access (CDMA) - also known as spread-spectrum multiple-access (SSMA).

TDMA, in which different groups of time slots are assigned to different system users, may be difficult to use in situations where the propagation path exhibits severe time dispersion, variable propagation delays, etc., because of the problem of maintaining synchronisation. It has the advantage that only one user is actually accessing the communication medium at any instant in time.

FDMA is widely used in broadcast-type systems, with different frequency sub-channels being assigned to different users. As mentioned previously, FDMA in conjunction with a non-linear transmission medium such as a satellite transponder can give rise to high levels of intermodulation, which ultimately will limit the overall throughput of the system. Another problem which may arise with FDMA is associated with the random combination in amplitude and phase of the access carriers; this can result in a composite time domain waveform having a high peak-to-mean ratio. If the transmission system is peak power limited, operation with such a signal results in a low average output power: note that it average SNR, rather than peak SNR, that determines the level of performance obtainable at the receiver.

In CDMA, the different system users are assigned uncorrelated access codes. As shown in Fig.9, at any receiver a matched filter or correlation detector can be used to extract the appropriate code for that receiver without interference from the other system users. Information is transmitted typically via phase-reversal modulation of the uncorrelated carrier codes. The main practical problem with this technique is to synthesise sufficiently large sets of uncorrelated access codes. However, the technique does provide each user with a high and deterministic level of protection against interfering signals, since each access is essentially an independent spread-spectrum system.

5. CODING FOR ENCRYPTION

Full cryptographic security can only be achieved if the data to be encrypted is in digital form. If this requirement is met, various forms of encryption are available, depending upon the physical nature of the communication system, the characteristics of the propagation medium, and the operational requirements. For analogue transmissions, limited security can be obtained from time domain and/or frequency domain "scrambling" procedures.

Fig.10 shows a block diagram of a simplified secure communication system. Any interceptor is assumed only to have access to the encrypted data as it is being passed over the channel. Prior to a secure message being transmitted, both originator and recipient must agree on a key k, with the data defining k being previously distributed by a completely secure means. At the transmitter, the encryption algorithm is defined by

$$c = t(m) \qquad [27]$$

where $t(\ )$ indicates a cryptographic transformation of the message m to yield a cryptogram c. At the receiver, the procedure must be reversible such that

$$t^{-1}(c) = m \qquad [28]$$

where t^{-1} is the "inverse" of t.

Therefore, the intended recipient will know k, c and t, whilst a potential interceptor only has access to c and may possibly be able to make an intelligent guess about the nature of t. For an encryption system to be effective, this level of knowledge on the part of the interceptor should not allow the system to be compromised.

In general, cryptographic procedures can be grouped into three main classes:

(a) Those using "transposition", where the characters of the original message retain their identity but are shifted in position.

(b) Those using substitution, where the characters of the original message are replaced by other characters derived from a separate alphabet by direct mapping.

(c) Those involving algebraic manipulation.

A simple example of (c) above is where the digits of a binary message m are added modulo-2 to a binary key stream k to give the encrypted stream c, ie

$$m \oplus_2 k = c \qquad [29]$$

At the receiver, a correctly synchronised version of the key stream k is available and is added modulo-2 to the incoming encrypted signal c giving

$$c \oplus_2 k = (m \oplus_2 k) \oplus_2 k \qquad [30]$$

using the result of [29]; this can be re-ordered to give

$$m \oplus_2 (k \oplus_2 k) = m \qquad [31]$$

Therefore, to decrypt the incoming message, a correctly synchronised version of k must be used. In practice, the derivation of this correctly phased key stream is one of the major problems associated with the transmission of encrypted data over radio paths. It is normally achieved by preceding the sending of the actual encrypted message by a robust synchronisation preamble.

6. SPEECH DIGITISATION

One of the most important forms of communication traffic is speech, since it provides a direct user-system interface and eliminates the need for message preparation. In the military context, considerable emphasis is placed upon the protection of speech transmissions from unauthorised interception. Therefore, as indicated in the previous section, it is necessary for the analogue speech signal to be converted to digital form if it is to be rendered fully secure via encryption. The main factors influencing the choice of speech digitisation procedure are as follows (Darnell, 1984):

(a) The nature of the available transmission medium, ie whether a wideband or narrowband channel is available.

(b) The speech quality required, eg whether it simply necessary for the speaker to be intelligible, or whether the voice quality must be such as to allow speaker recognition.

(c) Requirements for tandeming say a wideband speech digitisation algorithm with a narrowband algorithm, eg when an HF link forms a tail to a microwave LOS system.

(d) Whether speakers of different nationalities are required to use the system , possibly employing non-native languages.

(e) The user environment, eg are speakers situated in benign office-type surroundings, or in say a mobile with a high ambient acoustic noise level.

(f) The requirements for signalling, control and adaptation, which are themselves dependent upon the nature of the propagation path and its variability.

Speech digitisation procedures can be broadly classified under two main headings:

(i) Wideband (waveform encoding) systems

(ii) Narrowband (parameter encoding) systems.

Each of these two categories will now be considered in more detail.

6.1 Wideband Systems

Wideband speech digitisation systems are used in situations where transmission bandwidth is plentiful, typically of the order of a few tens of kHz per speech channel. Thus, VHF/UHF/microwave LOS and SHF satellite systems would normally employ this form of digitisation.

Wideband systems make use of a technique which is termed "waveform" encoding. Fig.11 shows a basic waveform encoding process in which the speech signal, after band-limiting, is sampled and the samples subjected to direct A/D conversion. No account is taken of the fundamental nature and properties of the input speech signal: in principle, the input to the waveform encoder could be any arbitrary signal. If the speech is bandlimited to about 3 kHz, as is the case with a normal telephone channel, then typical values of f and n might be 8 kHz and 8, giving an overal transmission rate

$$R = 8000.\ 8 = 64 \text{ kbits/s} \qquad [32]$$

The most important practical classes of waveform encoding are

(i) Pulse-code modulation (PCM)

(ii) Delta modulation (DM).

Block digrams of a PCM and DM system are shown in Figs.12 and 13 respectively. The PCM system is simply an eleboration of the basic digitisation process illustrated in Fig.11. DM involves a feedback loop which forces a quantised version of the input signal to track the actual input signal as closely as possible by injecting positive- or negative-going pulses into an integrator. It is these binary pulses which are actually transmitted over the channel.

There are many practical variants of these basic techniques, but generally the range of digitisation rates obtainable for a single speech channel is from about 16 to 64 kbits/s (Jayant, 1974).

6.2 Narrowband Systems

Narrowband speech digitisation techniques are applicable in situations where the transmission bandwidth is restricted to a maximum of a few kHz. They utilise a principle known as "parameter" encoding which is illustrated in Fig.14.

Based upon a knowledge of the mechanisms of speech production, an appropriate model of the human vocal tract and its excitation function is derived. Typically, this model would describe the major vocal tract resonances, and whether the excitation corresponds to a voiced or unvoiced sound. For voiced sounds, the excitation is assumed to comprise a regular train of pitch pulses at the pitch frequency; for an unvoiced sound, the excitation is assumed to be a wideband, noise-like, signal. If h(t) is the unit impulse response function of the vocal tract, then the output speech signal is generated by the convolution

$$V_o(t) = h(t) \otimes V_i(t) \qquad [33]$$

It is found that the vocal tract descriptive parameters and excitation type tend to vary relatively slowly for the average speaker, and only need updating at about 20 ms intervals. Thus a typical narrowband speech encoder, frequently known as a "vocoder", would transmit sample values of

- vocal tract model parameters (h(t))
- voiced/unvoiced decision
- pitch frequency (if voiced)
- amplitude scaling

as digital data frames every 20 ms or so. With parameter encoding schemes of this type, transmission rates in the range 0.8 to 2.4 kbits/s are obtainable, thus representing a considerable reduction in comparison with waveform encoding techniques. However, the quality of speech reproduction is also reduced considerably.

The most important practical forms of narrowband coders are linear-predictive coding (LPC) devices (Markel & Gray, 1976), channel vocoders (Kelly, 1970) and formant vocoders (Holmes, 1978). Interest is also developing in mid-band encoders operating in the transmission rate range from 6 to 16 kbits/s (Sambur, 1982), and in speech recognition/synthesis message transmission systems (Darnell & Chesmore, 1984). Mid-band encoding would be especially appropriate for speech communication in the VHF/UHF bands where wideband algorithms are used currently, but where there is considerable pressure to further reduce channel separations. Speech recognition/synthesis transmission systems, in which the data rate can be reduced to a few tens of bits/s, is potentially applicable to communication in the ELF/VLF/LF/MF/HF bands.

7. REFERENCES

1. Stein, S. and Jones, J.J., 1967: "Modern communication principles", McGraw-Hill.

2. Pasupathy, S., 1979: "Minimum shift keying: a spectrally efficient modulation", IEEE Comms. Soc. Mag., Vol. 17, No. 4.

3. Dixon, R.C., 1975: "Spread spectrum systems", Wiley (New York).

4. Utlaut, W.F., 1978: "Spread spectrum: principles and possible application to spectrum utilization and allocation", IEEE Comms. Soc. Mag., Vol. 16, No. 5.

6. Darnell, M., 1979: "An HF data modem with in-band frequency agility", Proc. IEE Int. Conf. on "Recent advances in HF communication systems and techniques", London.

7. Monsen, P., 1983: "Modern HF communications, modulation and coding", AGARD Lecture Series No.127 on "Modern HF communications".

8. Lin, S. and Costello, D.J. (Jr), 1983: "Error control coding: fundamentals and applications", Prentice-Hall.

9. Goodman, R.M.F. and Farrell, P.G., 1975: "Data transmission with variable redundancy error control over a high-frequency channel", Proc. IEE, Vol. 122, No. 2.

10. Bartholomé, P.J. and Vogt, I.M., 1968: "COMET - a new meteor-burst system incorporating ARQ and diversity reception", Trans. IEEE, Vol. COM-16, No. 2.

11. McCarthy, R.F., 1975: "Error control with time diversity techniques", Signal, May/June.

12. Chase, D., 1973: "A combined coding and modulation approach for communication over dispersive channels", Trans. IEEE, Vol. COM-21.

13. Douglas, F.W. and Hercus, P.T., 1971: "Development of AUTOSPEC Mark II", Point-to-Point Telecommunications, May.

14. Lee, Y.W., 1960: "Statistical theory of communication", Wiley.

15. Wozencraft, J.M. and Jacobs, I.M., 1965: "Communication engineering", Wiley.

16. Viterbi, A.J., 1967: "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", Trans. IEEE, Vol. IT-13.

17. Darnell, M., 1984: "Speech digitisation techniques", Proc. IEE Int. Conf. on "Secure communication systems", London.

18. Jayant, N.S., 1974: "Digital coding of speech waveforms: PCM, DPCM and DM quantizers", Proc. IEEE, Vol. 62.

19. Markel, J.D. and Gray, A.H., 1976: "Linear prediction of speech", Springer-Verlag.

20. Kelly, L.C., 1970: "Speech and vocoders", Radio & Electronic Engineer, Vol. 40, No. 2.

21. Holmes, J.N., 1978: "Parallel formant vocoders", IEEE Eascon 78.

22. Sambur, M.R., 1982: "Speech algorithm advances promise toll-quality medium-band digitized speech", Speech Technology, Vol. 1, No. 3.

23. Darnell, M. and Chesmore, E.D., 1984: "Aspects of digitised speech communication over limited capacity channels", Proc. IEE Int. Conf. "Communications 84", Birmingham, UK.

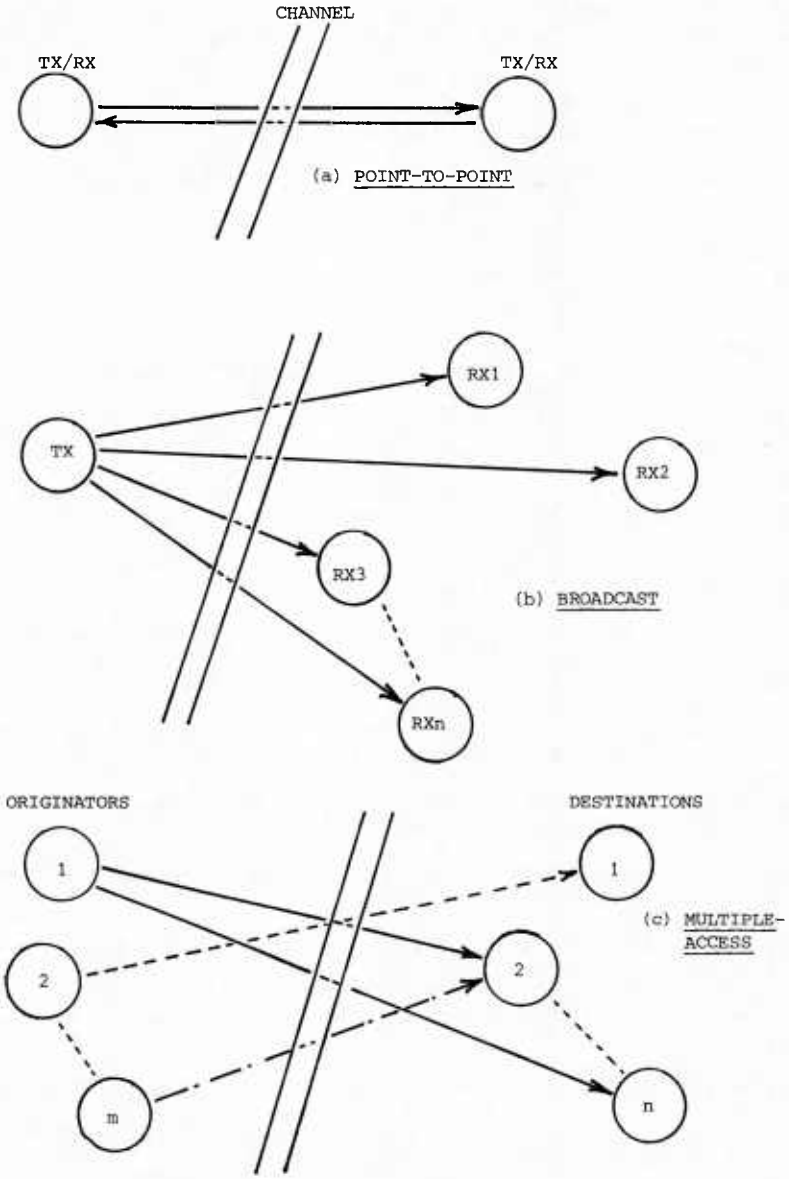


FIG. 1 COMMUNICATION SYSTEM ARCHITECTURES

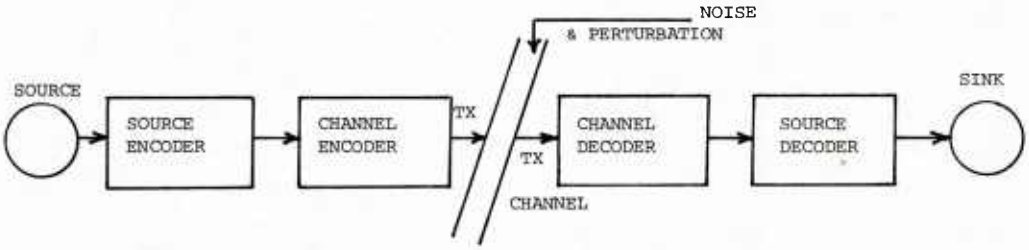


FIG. 2 GENERALISED COMMUNICATION SYSTEM

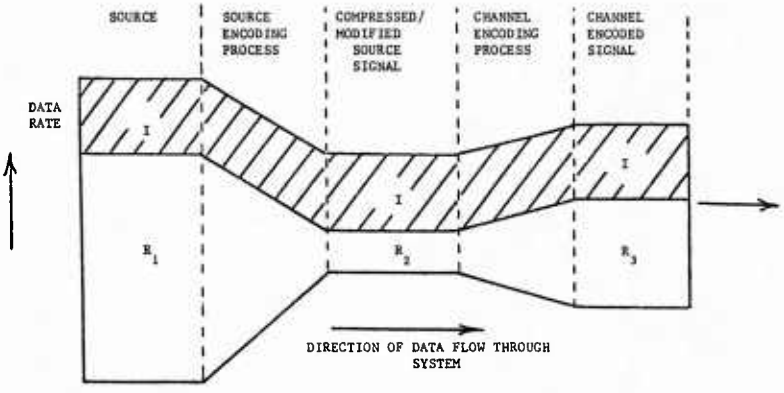


FIG. 3 COMBINED EFFECTS OF SOURCE & CHANNEL CODING

a) Performance in Gaussian white noise with steady signal conditions.

MODULATION TYPE	P_e FOR COHERENT DETECTION	P_e FOR NON-COHERENT DETECTION
BINARY OOK NB: γ defined only in MARK when signal is present	a) $1/2 \operatorname{erfc} [\sqrt{\gamma}/2]$ at optimum threshold b) $\doteqdot 1/\sqrt{\pi\gamma} \exp [-\gamma/4]$ at high γ and optimum threshold	a) $1/2 \exp [-V_T'^2/2]$ $+1/2[1 - Q(\sqrt{2\gamma}, V_T')]$ b) $\doteqdot 1/2 \exp [-\gamma/4]$ at high γ and optimum threshold
BINARY FSK	$1/2 \operatorname{erfc} \sqrt{\gamma/2}$	$1/2 \exp [-\gamma/2]$
BINARY DPSK	$1/2 \exp [-\gamma]$	
BINARY PSK	$1/2 \operatorname{erfc} \sqrt{\gamma}$	

Where

γ = Signal-to-Noise Power Ratio

P_e = Error Probability

V_T' = Normalised Threshold Value

FIG. 4 ERROR PROBABILITIES FOR BASIC MODULATION TYPES

b) Performance in Gaussian white noise with a Rayleigh flat-fading signal.

MODULATION TYPE	P_e FOR COHERENT DETECTION	P_e FOR NON-COHERENT DETECTION
BINARY OOK NB: γ_o only defined in MARK when signal is present	$1/2 \left\{1 - \frac{1}{\left[1 + \frac{4}{\gamma_o}\right]^{\frac{1}{2}}}\right\}$ at optimum threshold For $\gamma_o \gg 1$ $1/\gamma_o$ at optimum threshold	For $\gamma_o \gg 1$ $2/\gamma_o$ at optimum threshold Otherwise, no convenient analytical expression
BINARY FSK	$1/2 \left\{1 - \frac{1}{\left[1 + \frac{2}{\gamma_o}\right]^{\frac{1}{2}}}\right\}$ For $\gamma_o \gg 1$ $1/(2\gamma_o)$	$\frac{1}{[2 + \gamma_o]}$ For $\gamma_o \gg 1$ $1/\gamma_o$
BINARY DPSK	$\frac{1}{[2 + 2\gamma_o]}$ For $\gamma_o \gg 1$ $1/(2\gamma_o)$	
BINARY PSK	$1/2 \left\{1 - \frac{1}{\left[1 + \frac{1}{\gamma_o}\right]^{\frac{1}{2}}}\right\}$ For $\gamma_o \gg 1$ $1/(4\gamma_o)$	

γ_o = Mean Signal-to-Noise Ratio (averaged over fading)

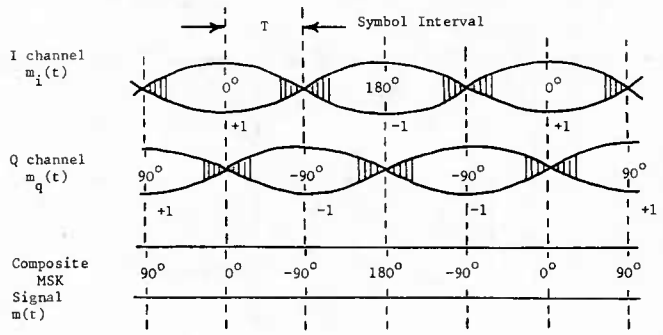


FIG. 5 PRINCIPLE OF MSK MODULATION

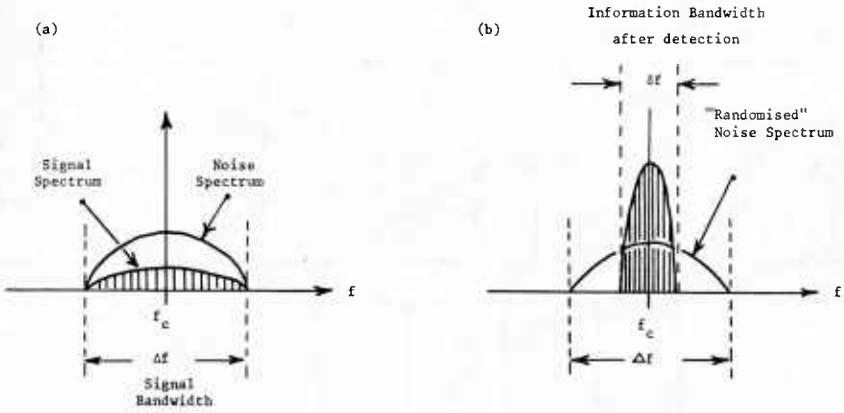


FIG. 6 PRINCIPLE OF DSSS PROCESSING

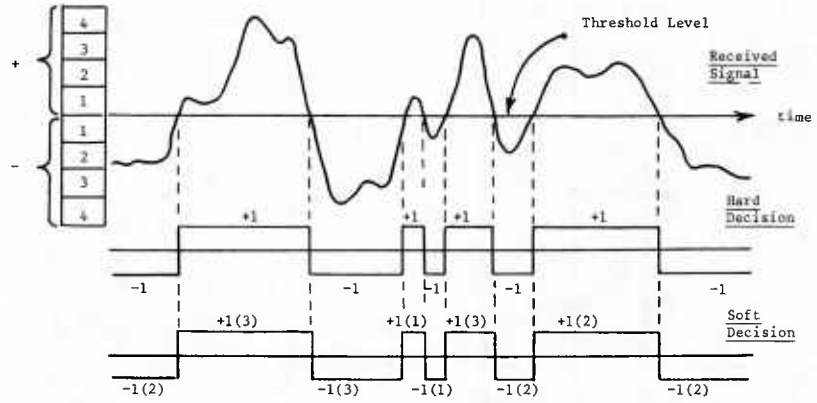


FIG. 7 PRINCIPLE OF SOFT-DECISION DECODING

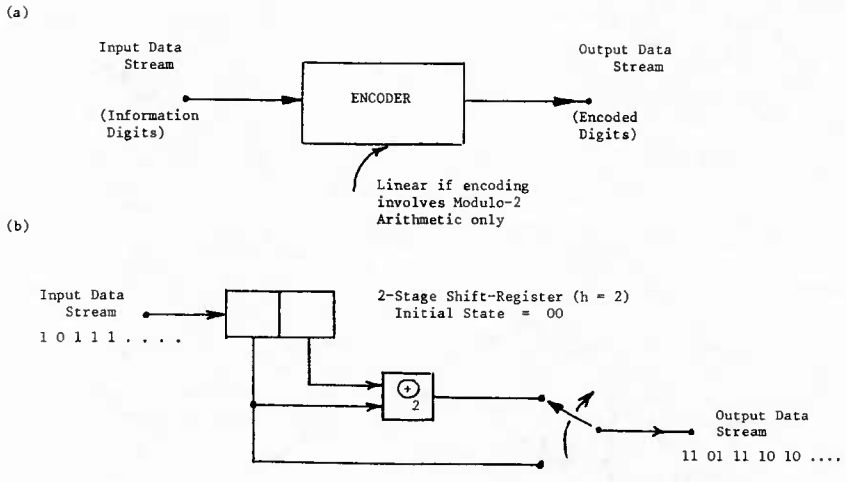


FIG. 8 CONVOLUTIONAL ENCODING

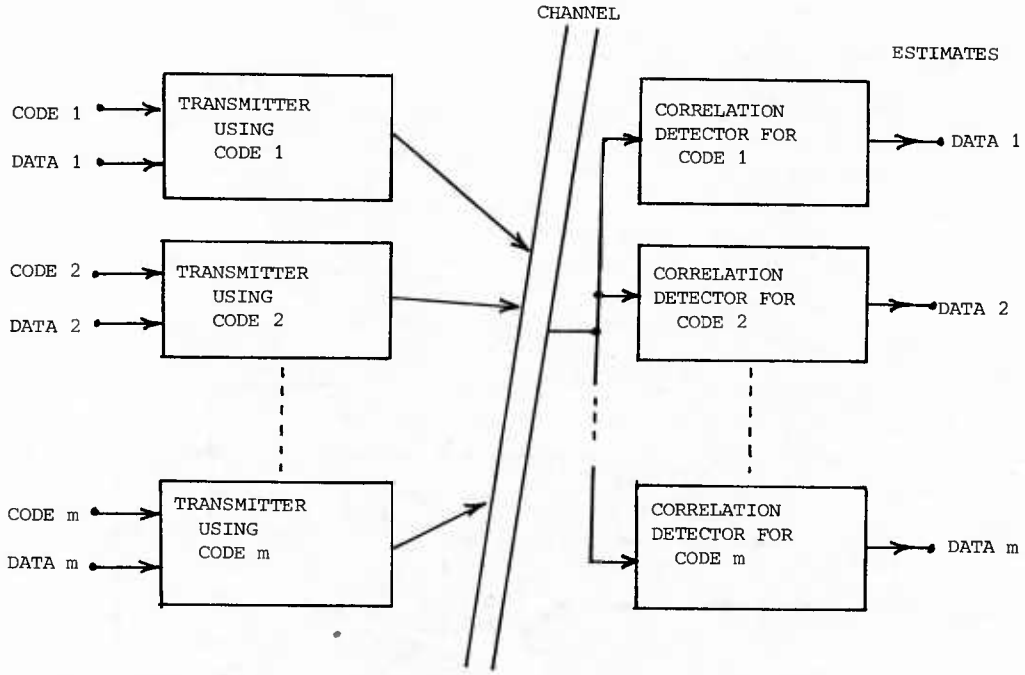


FIG. 9 CDMA SYSTEM

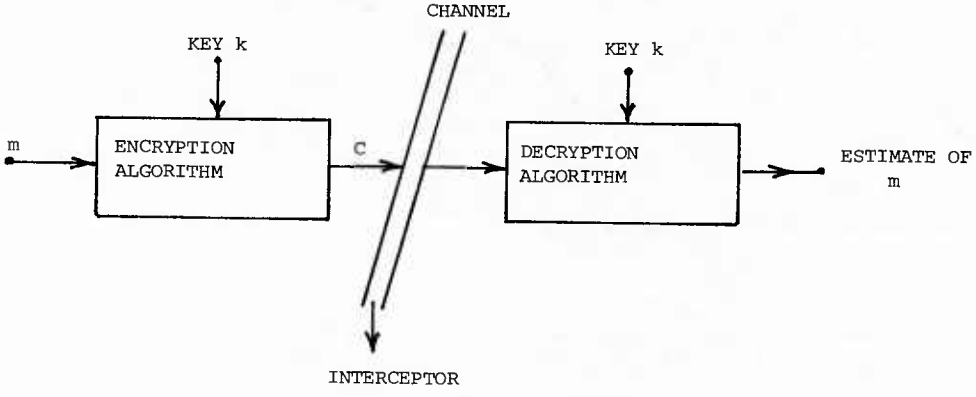


FIG.10 SIMPLIFIED SECURE COMMUNICATION SYSTEM

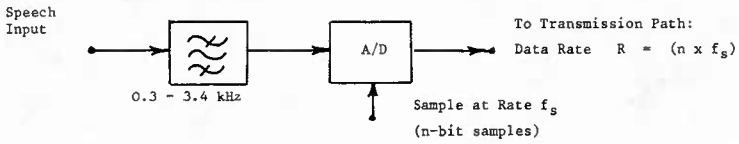


FIG. 11 BASIC WAVEFORM ENCODING PROCESS

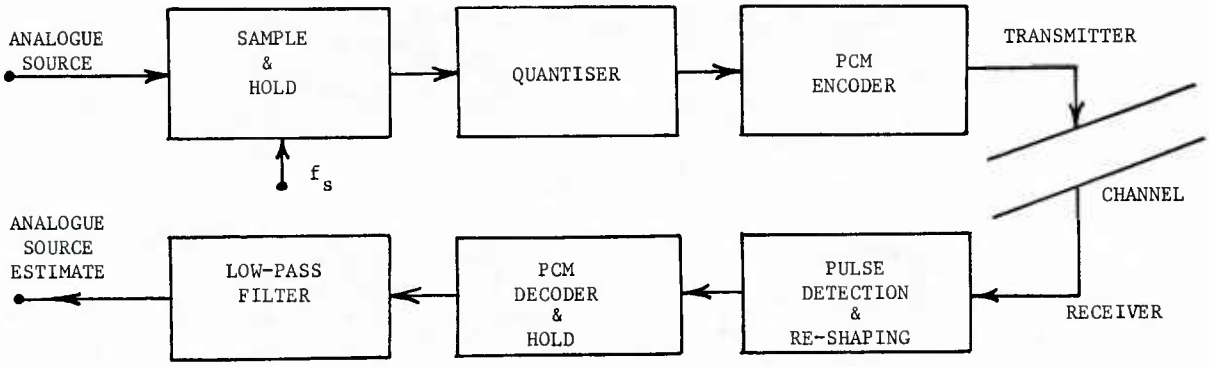


FIG. 12 BLOCK DIAGRAM OF PCM SYSTEM

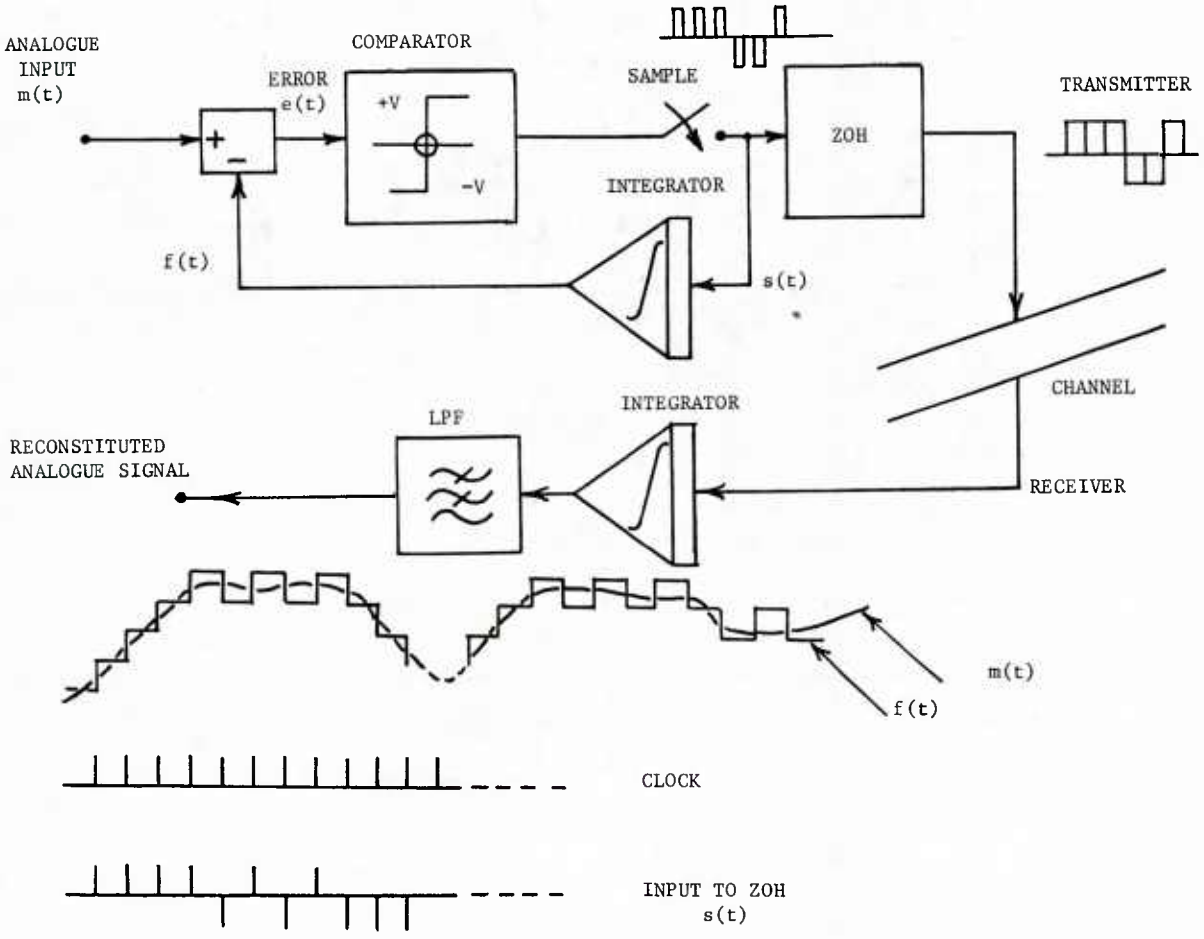


FIG. 13 DELTA MODULATION SYSTEM

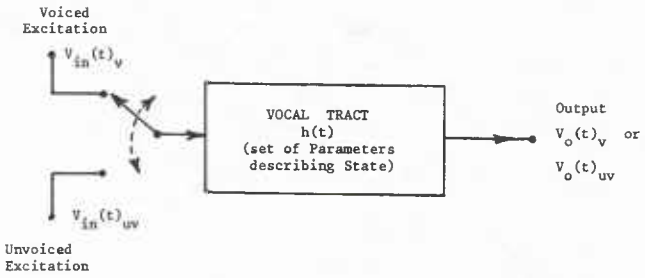


FIG. 14 PARAMETER ENCODING TECHNIQUE

ELF/VLF/LF PROPAGATION AND SYSTEM DESIGN

Francis J. Kelly

Ionospheric Effects Branch
Space Sciences Division
Naval Research Laboratory
Washington, DC

ABSTRACT

The effects of propagation conditions in the ELF/VLF/LF frequency band on the design and performance of digital communications systems are described. The communications channel from a ground-based or airborne transmitting antenna through the earth-ionosphere waveguide to the receiver is discussed. Questions of coverage, noise, dispersion, time variance and propagation disturbances are summaried.

1.0 INTRODUCTION

The utilization of the longwave channel (5Hz to 300kHz) for digital radio transmission has greatly increased during the past 20 years. Several large VLF ground based transmitting antennas have been rebuilt, and new facilities have been constructed. The ELF band has been opened up for use. Squadrons of longwave broadcasting aircraft have been procured and operated. New modulation techniques and receivers have been introduced into widespread use. The importance of systems utilizing the longwave band is not widely appreciated, but the statement by Donald C. Latham, Assistant Secretary of Defense for Command, Control, Communication and Intelligence in Appendix I (from Ref. 1) is abundantly clear. It is impossible to present a comprehensive picture of the long wavelength situation today without producing a book as large as A. D. Watt's monumental treatise (Ref. 2). However, because many students and readers will be unfamiliar with the longwave area, we will present a brief introductory discussion of the total longwave communications system including transmitting antennas and radio receivers before discussing in detail the impact of propagation on system design and performance.

2.0 ANTENNAS

In the longwave channel antennas are an important component of the overall propagation medium because they provide a basic restriction on types of signals that can be launched. The longwave channel is characterized by the fact that efficient radiation from conventional size man-made structures is difficult to achieve. Great ingenuity has been employed to produce adequately large antenna configurations and to obtain the greatest possible efficiency from antenna systems that are necessarily small in comparison to a wavelength. Figure 1 shows the types of longwave antenna in common use.

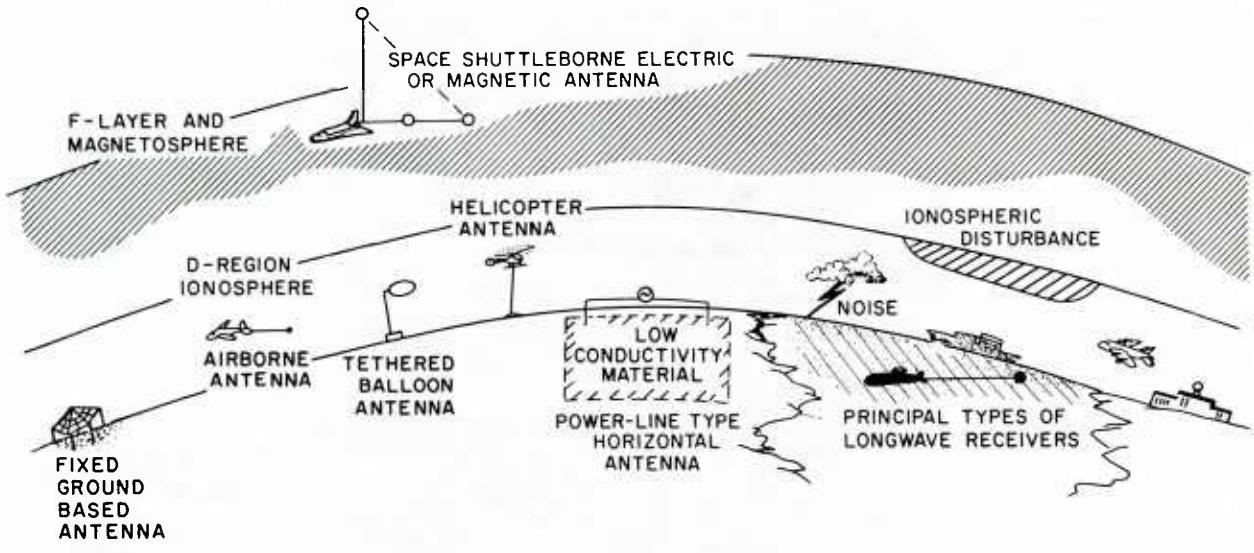


Fig. 1: Longwave Broadcast Systems

Many of the fixed/ground based vertical monopole antennas are of the "small antenna" variety characterized by a very small radiation resistance (less than 1 ohm). These are consequently quite inefficient unless special precautions are taken to reduce utterly all unnecessary losses. One common technique is the construction of a large ground plane to suppress losses

in the earth near the antenna. A second feature is the deployment of an elaborate ''top hat'' consisting of cables connected together and suspended from towers to provide additional antenna capacity to wirease the antenna's ability to carry current for a given voltage. In this way the losses can be reduced to a minimum and a higher efficiency can be achieved. These procedures however result in a high ''Q'' antenna system with a consequent low bandwidth. The design considerations for such antennas are summarized in Refs. 1 through 3. Figures 2 through 5 (from Ref. 3) show the four principal ground based antenna types: Triatic, Golaith, Trideco and valley span. Figures 6 through 11 illustrate the designs of U.S. Navy VLF fixed ground based antennas.

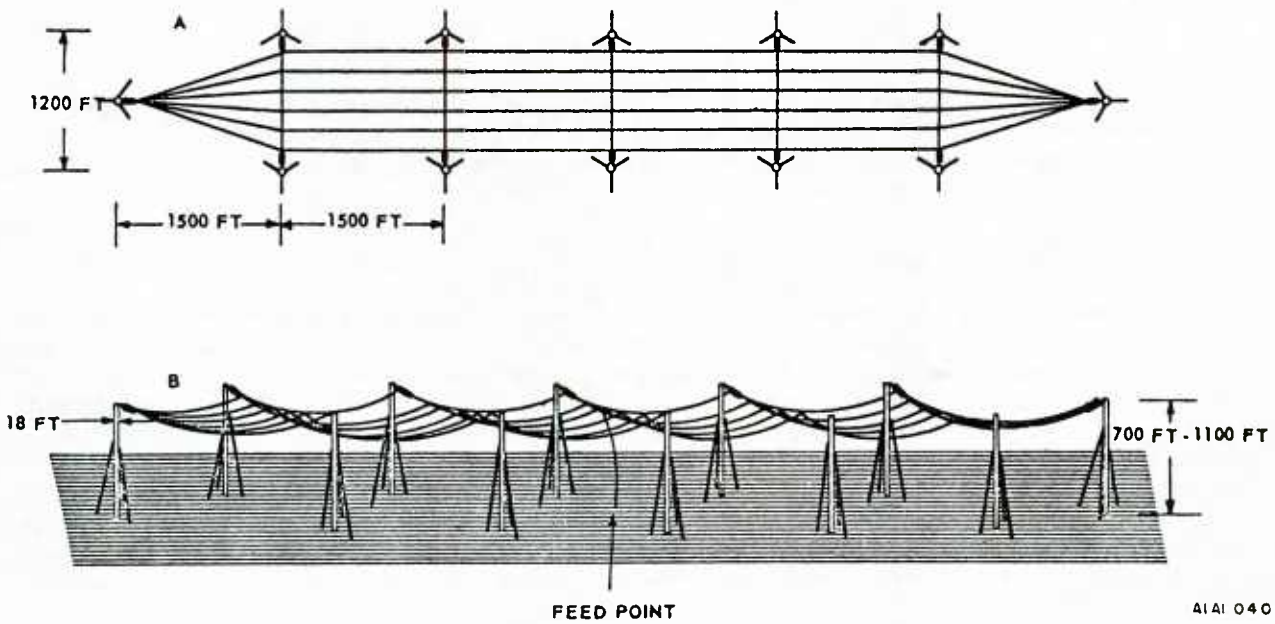


Fig. 2: Triatic Type Antenna

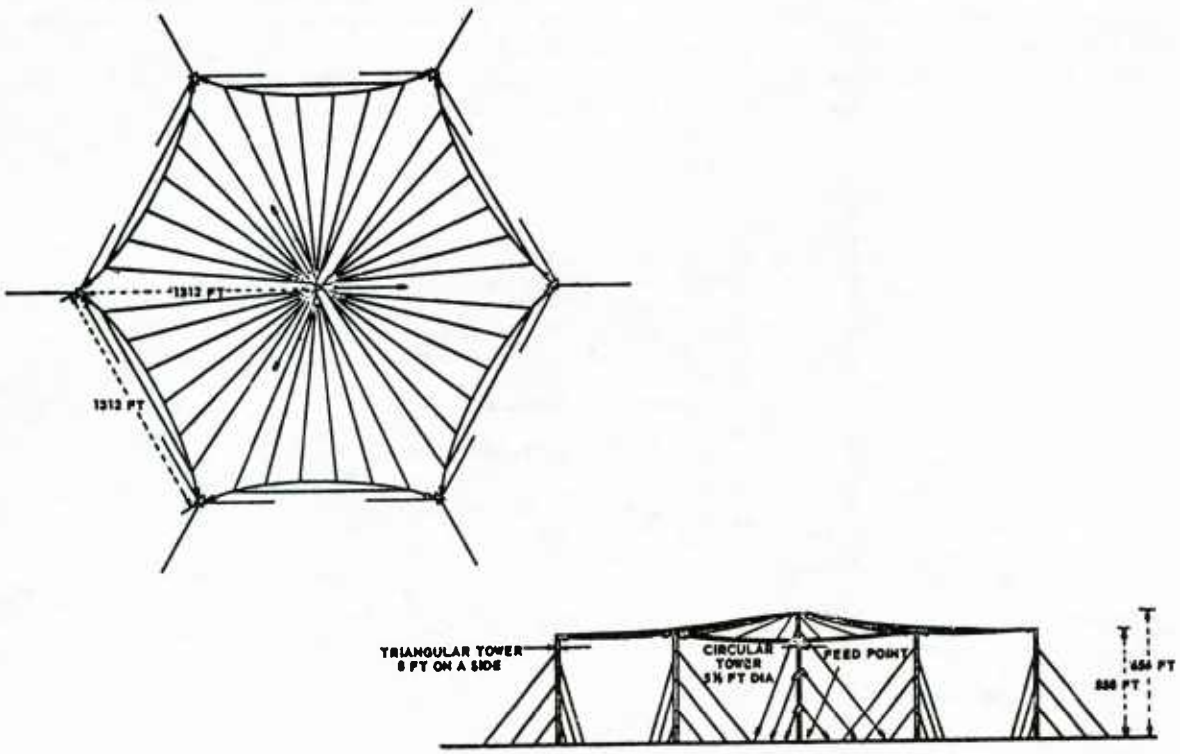


Fig. 3: Goliath Type Antenna

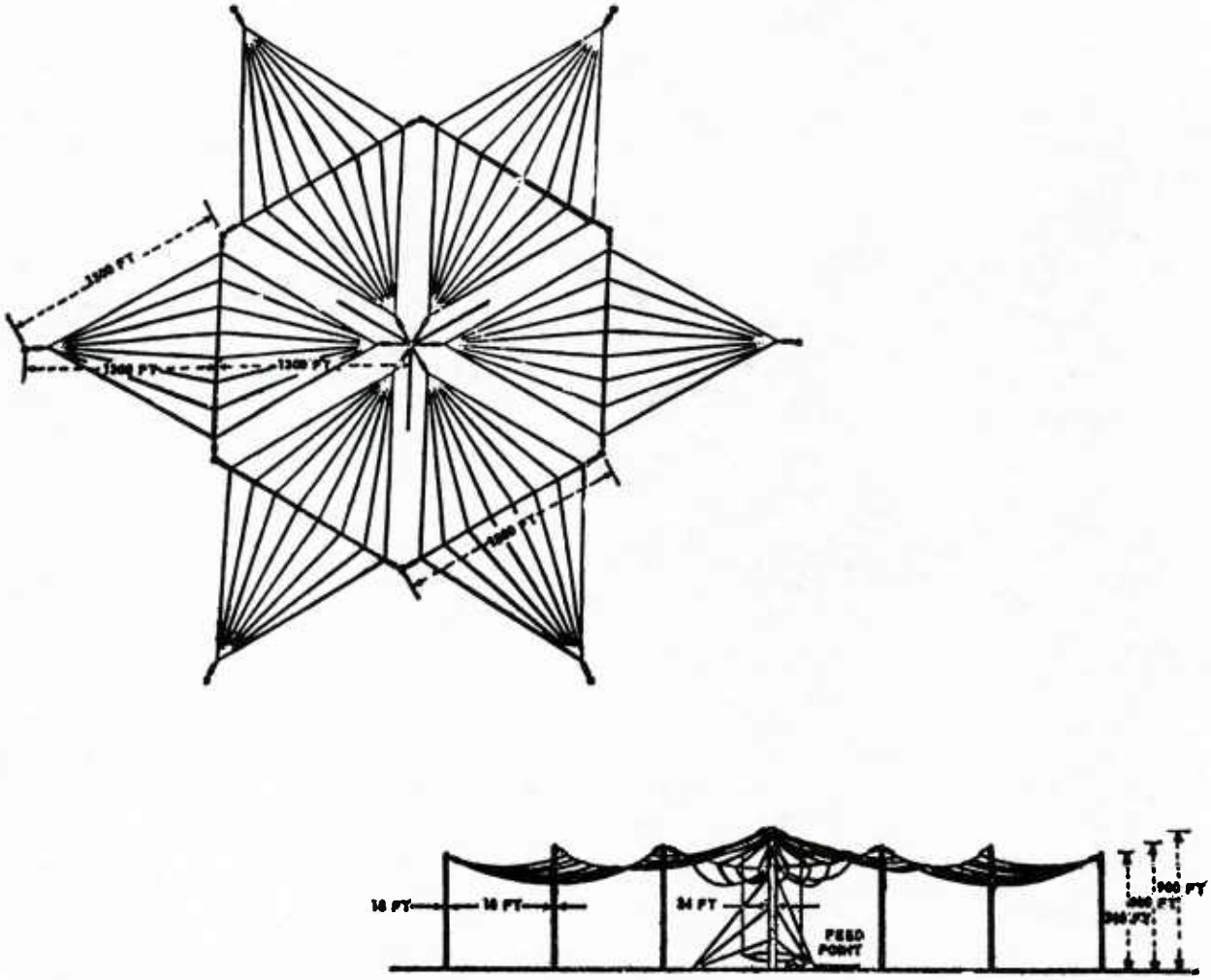


Fig. 4: Trideco Type Antenna

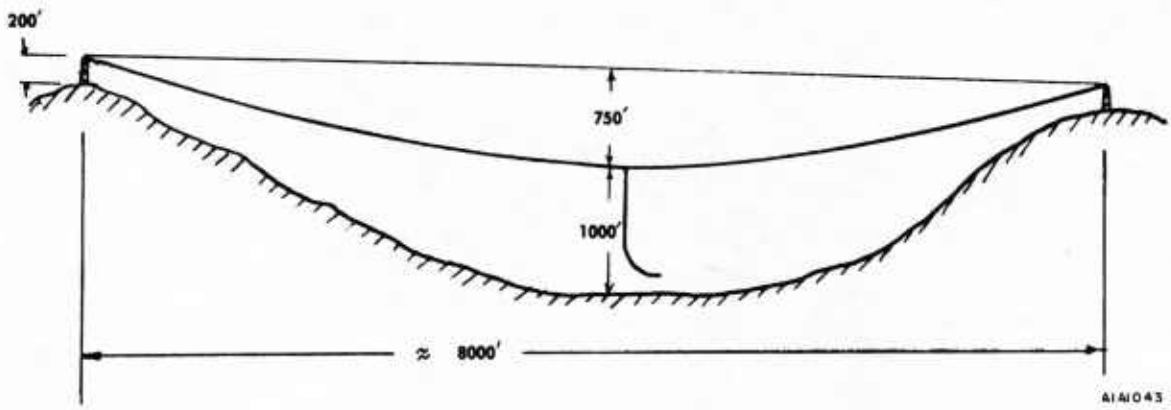


Fig. 5: Natural Valley-Span Antenna

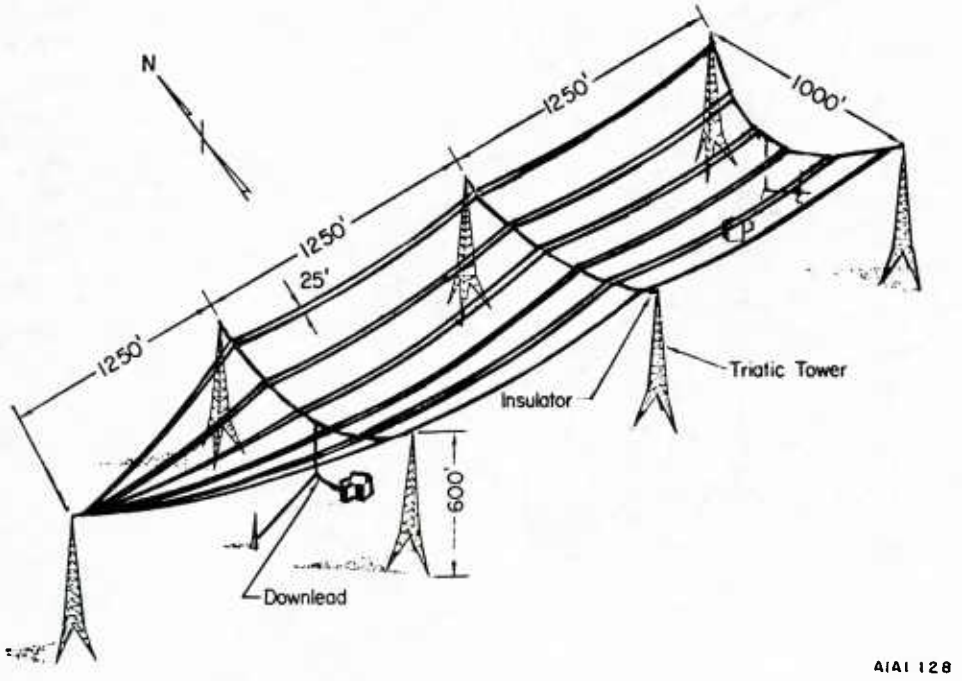


Fig. 6: Old Lualualei Antenna, Pictorial View

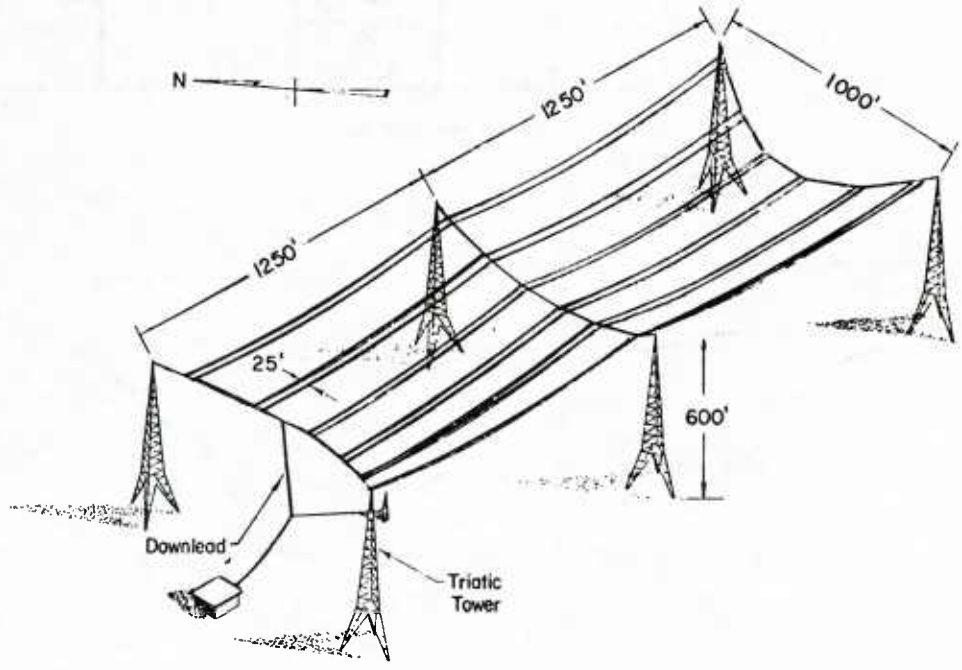


Fig. 7: Balboa (Summit) Antenna, Pictorial View

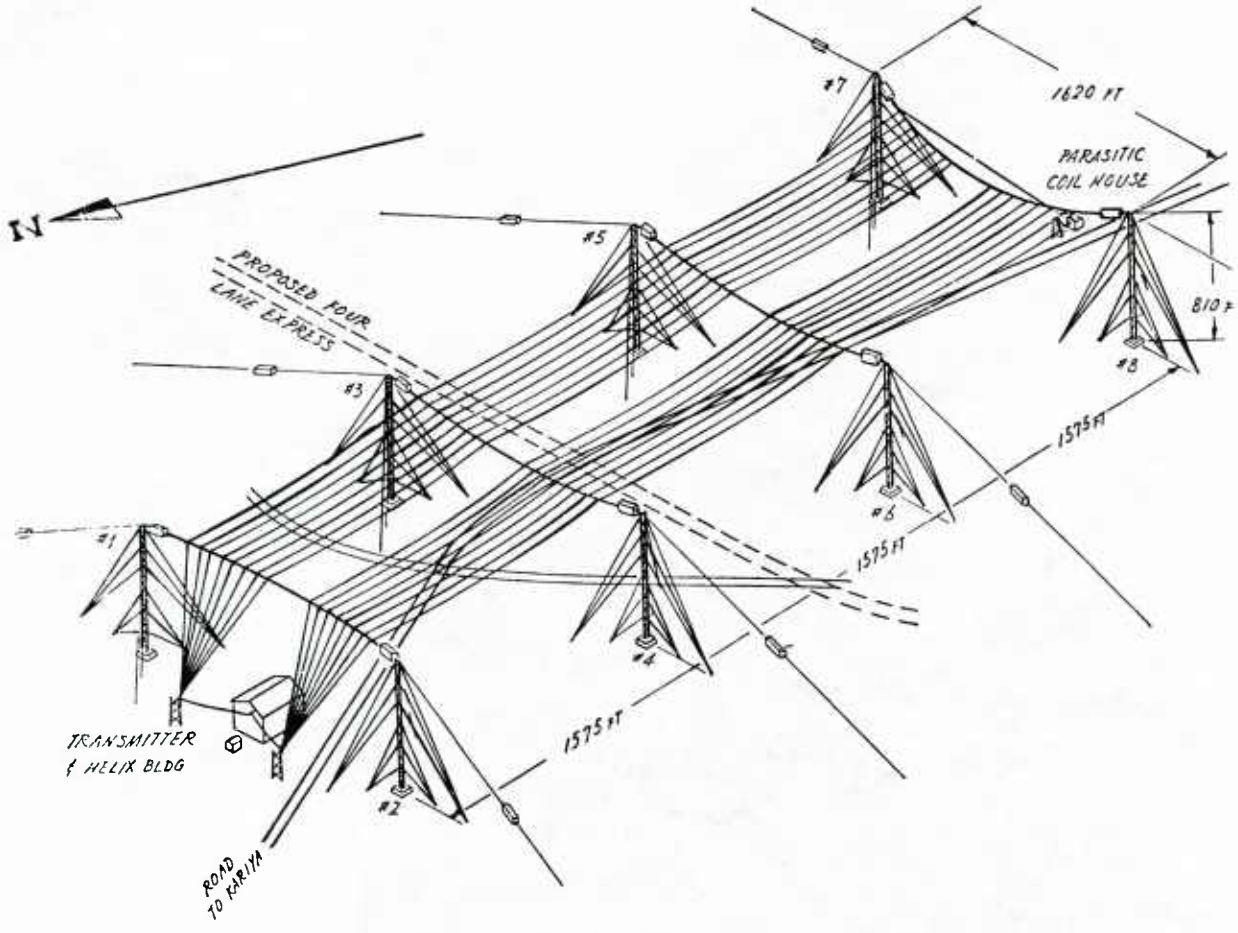


Fig. 8: Yosami, Japan Antenna, Pictorial View

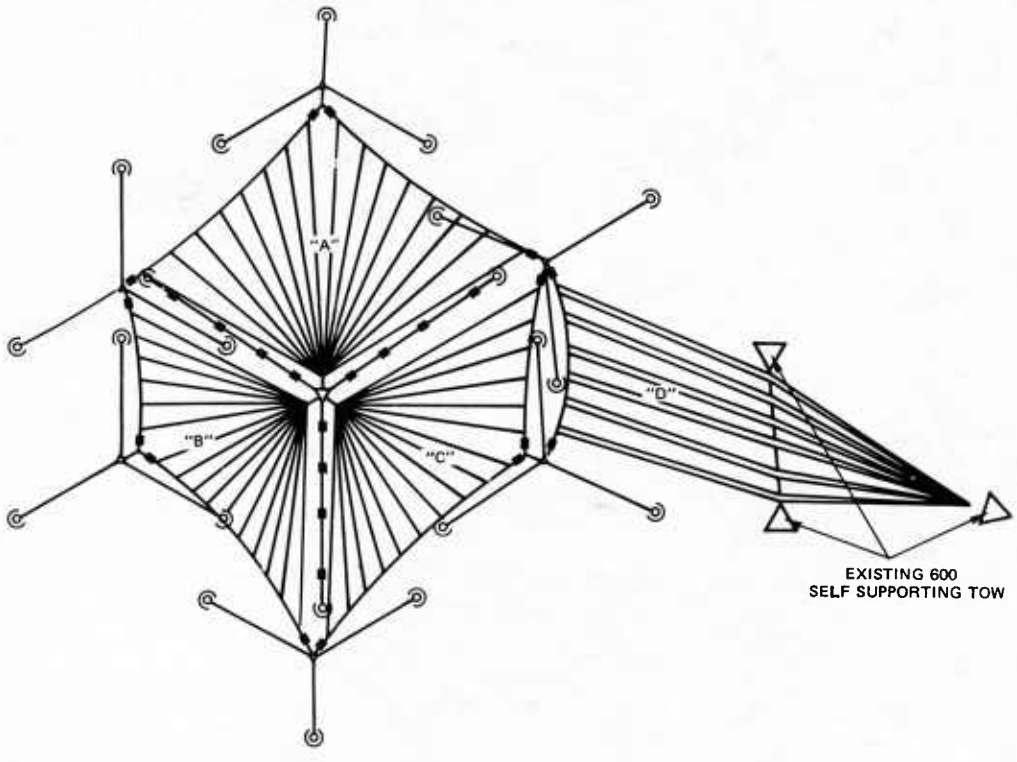


Fig. 9: Annapolis Top Hat Arrangement

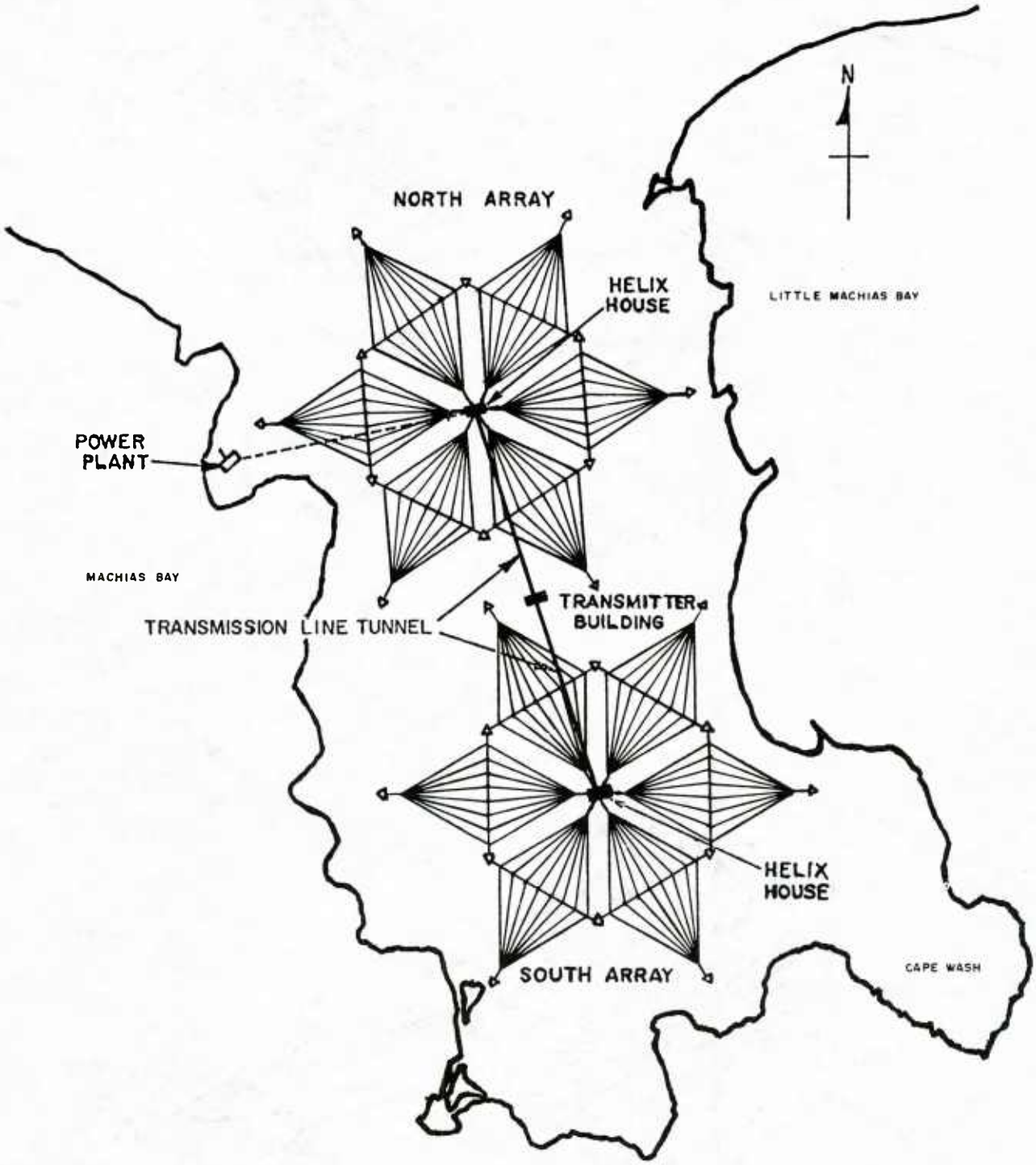


Fig. 10: Cutler, Maine Antenna-Installation, Plan View

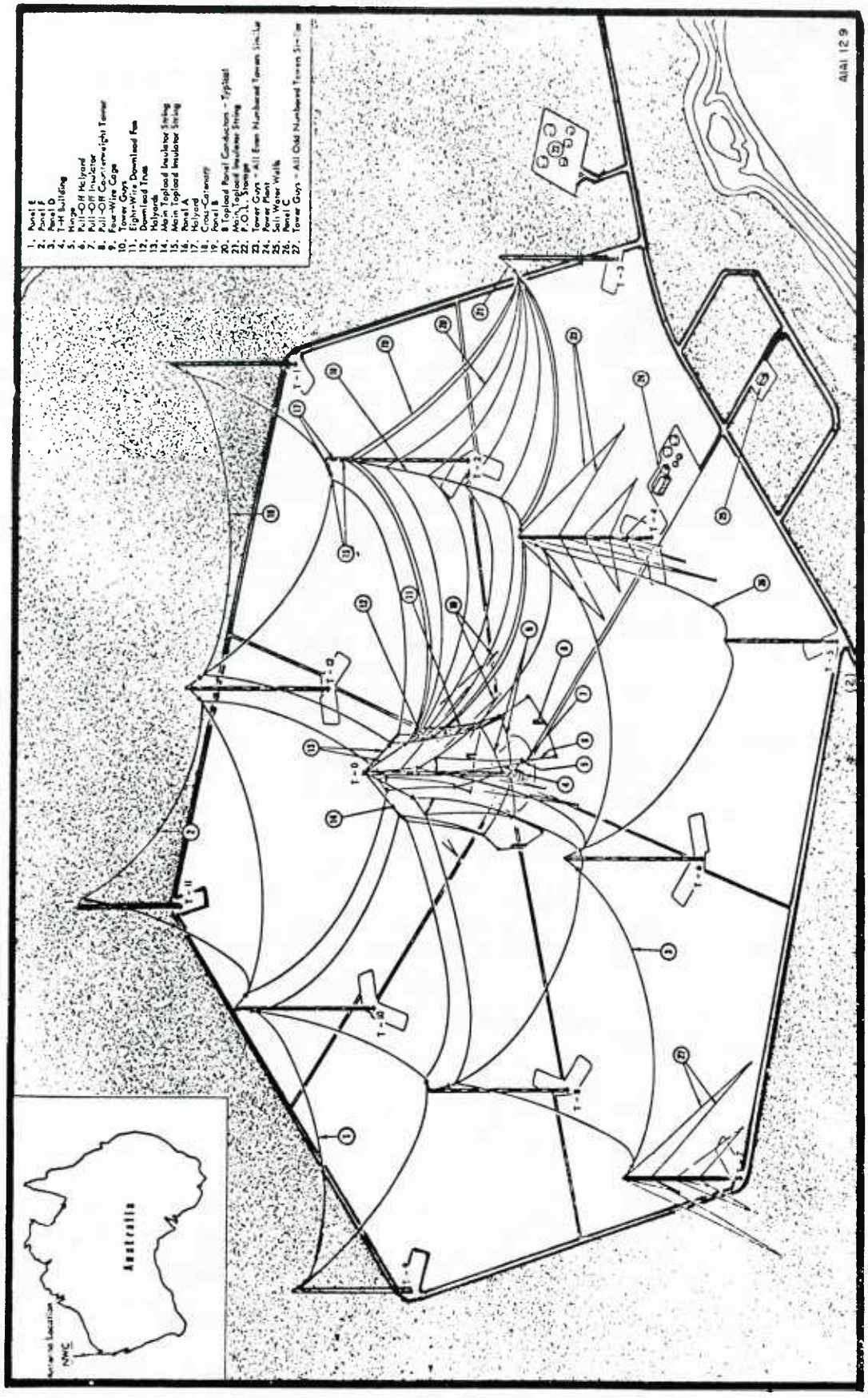


Fig. 11: NWC Australia Antenna

An alternative method for making an efficient radiator at long wavelengths is to suspend a long conducting cable from an aircraft, or other lifting vehicle. In this way a half wave dipole antenna can be achieved with a large radiation resistance and high radiation efficiency without the high "Q" and limited bandwidth of the ground based tower antennas. This technology is the basis of the TACAMO and airborne command post longwave communication capability as presented in Refs. 5, 6 and 7. The airborne facilities are illustrated in Fig. 12 taken from Ref. 5.

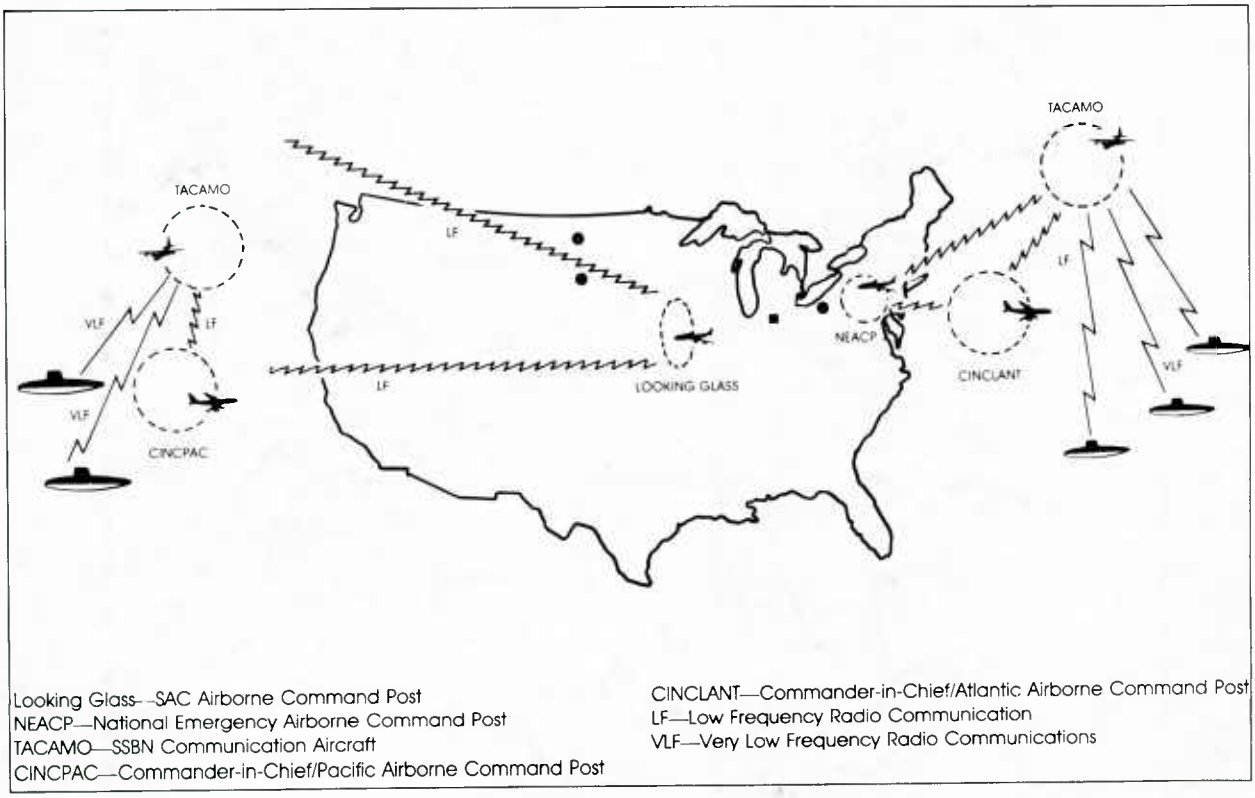


Fig. 12: Airborne submarine communications. (Figure 3 on p. 28, Vol. 39, No. 7, entitled "Strategic Command and Control and National Security," by Dr. Bruce G. Blair from Signal, published by The Armed Forces Communications and Electronics Association Copyright (c) 1985, used by permission.)

A horizontal electric antenna can be used at long waves also. Such an antenna is utilized for the ELF communication system (Refs. 8, 9). In this system a powerful current is driven along horizontal wires like power lines stretched over the earth's surface and grounded at each end. When the conductivity of the underlying earth is low, it is as if the antenna is suspended at a modest height above the "electrical" earth. The effective image currents flow from one terminal to the other at a depth proportional to the electromagnetic skin depth of the underlying material. This mechanism permits the horizontal antennas to operate with modest efficiency.

Soon it will be feasible to deploy a large longwave antenna from an earth orbiting satellite. Italy and the USA(NASA) have a joint program to develop a 60 km long conducting tether system for use in space. One of the applications of this tether will be the generation of long radio waves (Ref. 10).

3.0 PROPAGATION

3.1 Summary

The earliest trans-Atlantic radio propagation experiments of Marconi were conducted in the long wavelength band (Ref. 2). In comparison to the short-wave band the long wavelength band is highly reliable. Long-distance propagation of a long radio wave relies on a reflection from the lower ionosphere to bring the radio wave back to earth. On the other hand short-waves are gradually bent back to earth because of a refraction process that depends on the detailed horizontal and vertical gradients of an upper-ionosphere F-layer. A VLF wave spends most of its journey in the free space between the earth's surface and the lower ionosphere. A short-wave propagates mostly through the ionosphere between 70 km and 300 km height between the D-layer and the F-layer. This is probably the basic reason that long-wavelength waves show greater stability of reception. The disadvantages of long waves (which were deadly from an ordinary economic point of view) are:

(a) Large antenna structure;

(b) Narrow bandwidths;

(c) Limited total spectrum;

(d) High powers required to overcome noise.

The advantages of long waves are:

(a) Great propagation range (even to the antipodes)

(b) High phase stability

(c) Significant ability to penetrate earth and seawater.

The long propagation range and phase stability for long wavelength waves makes their use for navigation feasible (Omega, Decca and Loran). Their penetration into the earth makes the waves useful to prospectors looking for ore-bodies or other deposits beneath the earth.

3.2 The Ionosphere and Reflection Coefficients:

The layers ascribed to the ionosphere and atmosphere are illustrated in Fig. 13 taken from Ref. 11. The ionospheric reflection coefficients are defined as the ratio of downcoming electric field to the up going field. Because there are two orthogonal up going and downcoming field components a total of four complex reflection coefficients characterize the process at any given angle of incidence. In matrix form we have

$$\begin{pmatrix} E_{\parallel}^{d} \\ E_{\perp}^{d} \end{pmatrix} = \begin{pmatrix} {}_{\parallel}R_{\parallel} & {}_{\perp}R_{\parallel} \\ {}_{\parallel}R_{\perp} & {}_{\perp}R_{\perp} \end{pmatrix} \begin{pmatrix} E_{\parallel}^{u} \\ E_{\perp}^{u} \end{pmatrix} \tag{1}$$

where

$E_{\parallel}^{u}$ is the upgoing electric field in the plane of incidence;
$E_{\perp}^{u}$ is the upgoing electric field perpendicular to the plane of incidence (i.e. the horizontal electric field);
$E_{\parallel}^{d}$ is the downcoming electric field in the plane of incidence;
$E_{\perp}^{d}$ is the downcoming electric field perpendicular to the plane of incidence.

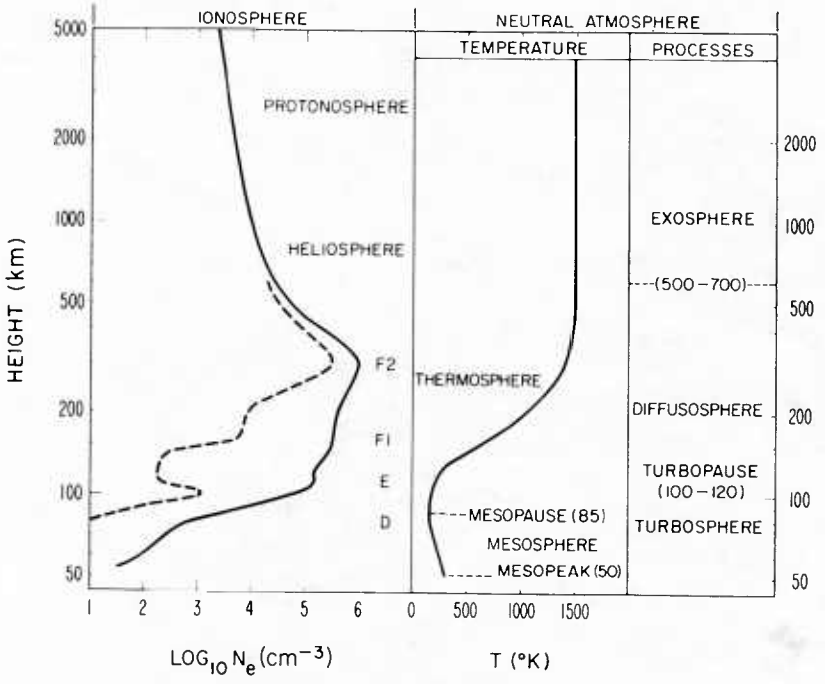


Fig. 13: Atmospheric layers and nomenclature

The reflection coefficients are functions of the electron and ion density profiles of the ionosphere. The collision frequency profiles, geomagnetic field and angle of incidence must also be specified. Several full-wave techniques can be used to calculate the reflection coefficient matrix. It is beyond the range of this lecture to describe these methods in depth. Figures 14 through 16 show sample calculations of reflection coefficients versus incidence angle at 3 kHz for a geomagnetic dip angle

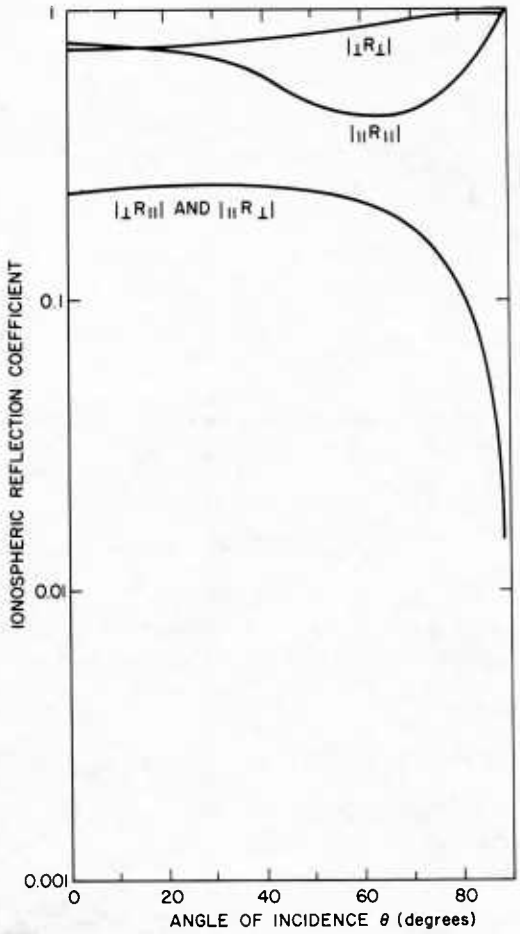


Fig. 14: 3-kHz nighttime ionospheric reflection coefficients, west-to-east propagation

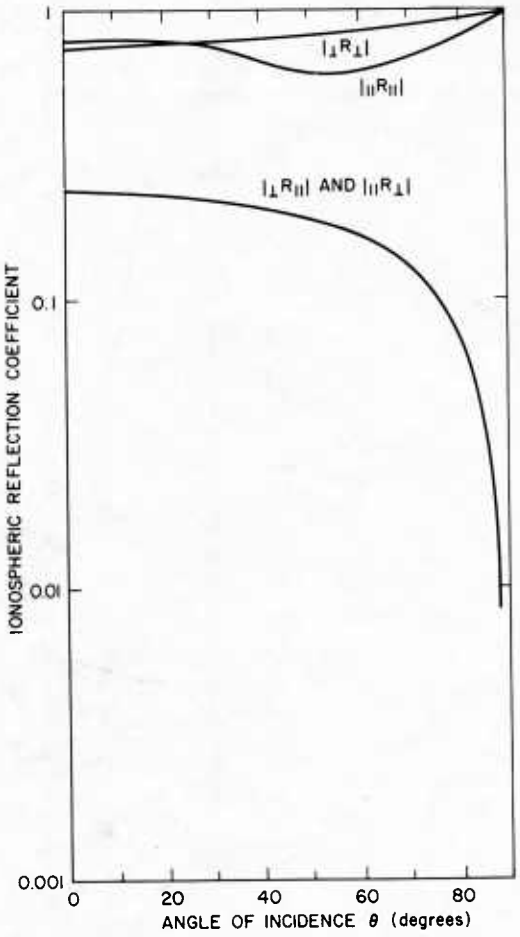


Fig. 15: 3-kHz nighttime ionospheric reflection coefficients, south-to-north propagation

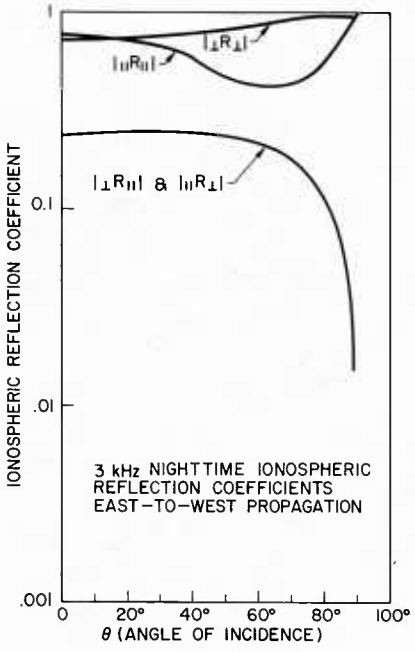


Fig. 16: 3-kHz nighttime ionospheric reflection coefficients, east-to-west propagation

of 60°. (Ref. 12) Figure 17 shows two reflection coefficients as a function of frequency for fixed incidence angle of approximately 30°. (Ref. 13) The ionospheric profile used for each calculation is shown in Fig. 18. Notice that the calculated reflection coefficients above about 30 kHz are strongly varying with frequency. They seem to reflect the detailed structure of the ionosphere. The smooth approximation by Allcock and Belrose (Ref. 14) seems adequate for a first approximation; but it lacks the oscillations noted in the fullwave results. During nighttime long radio waves are transmitted through the lower ionosphere with small loss. Space-borne receivers can readily hear ground based transmitters. Likewise receivers on the ground can sense many curious electrical noises that are generated and stimulated in the ionosphere and magnetosphere. For these ground-to-space and space-to-ground paths the lower ionosphere transmission coefficient is significant in forming a signal power budget. This transmission coefficient can be calculated using Pitteway's full wave method. (Ref. 15). Sample transmission coefficient results for a 3 kHz case are shown in Figs. 19 and 20.

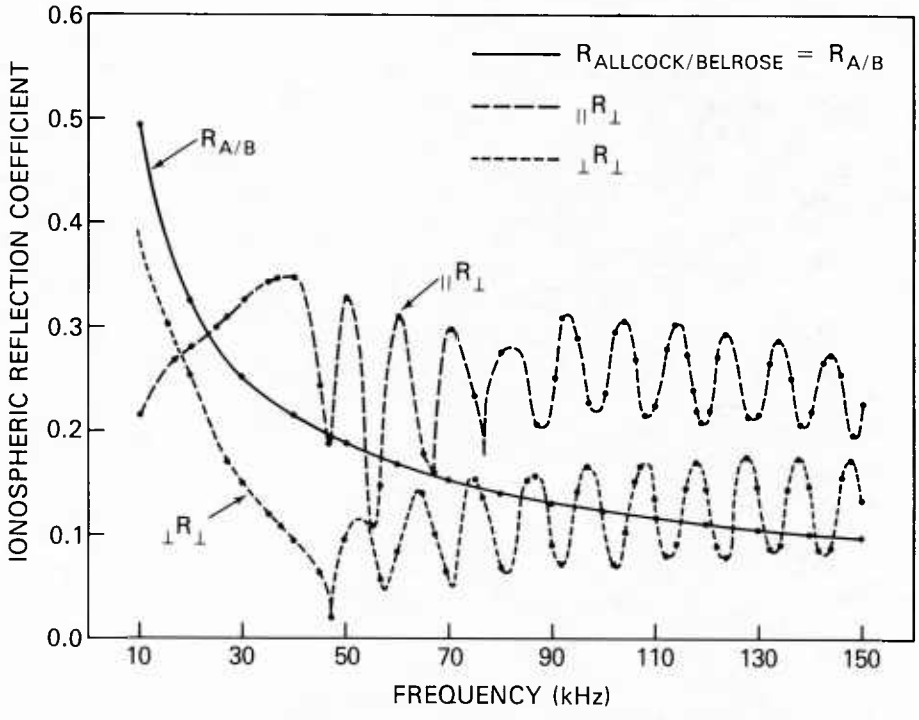


Fig. 17: Graph of reflection coefficients vs frequency. The $R_{A/B}$ curve is a plot of an empirical reflection coefficient approximation given by Belrose [14] using an equivalent frequency concept of Allcock. The $_{\|}R_{\perp}$ and $_{\perp}R_{\perp}$ reflection coefficients follow the convention of [15].

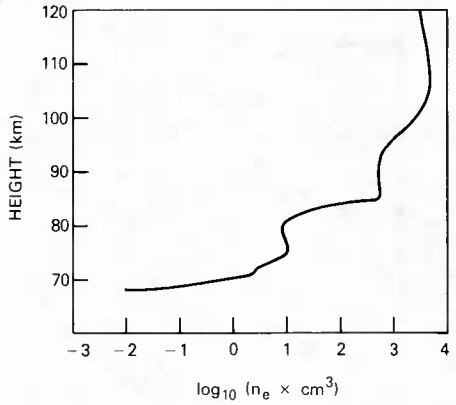


Fig. 18: Electron density profile for nighttime calculations

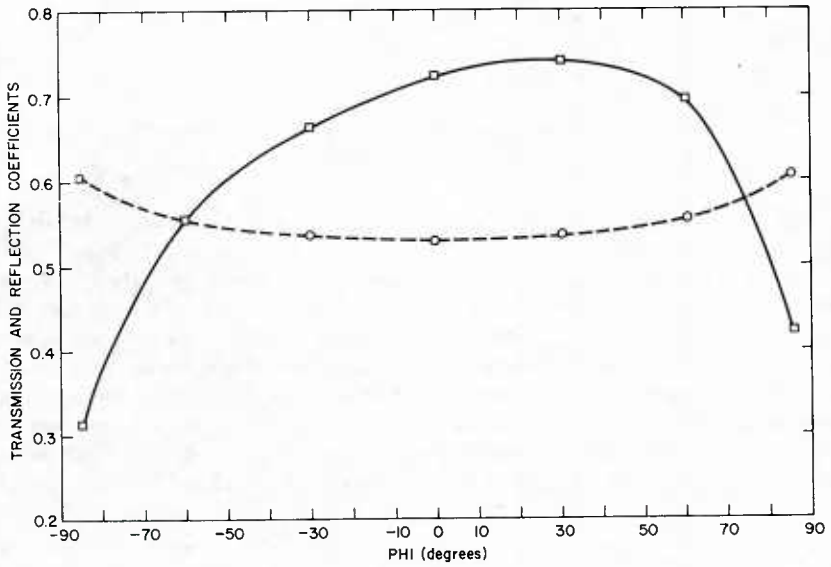


Fig. 19: Transmission coefficient (solid line) and reflection coefficient (dashed line) vs angle of incidence from vertical (PHI) for frequency of 3 kHz, dip angle = 60°, and azimuth = 0° (measured from magnetic north)

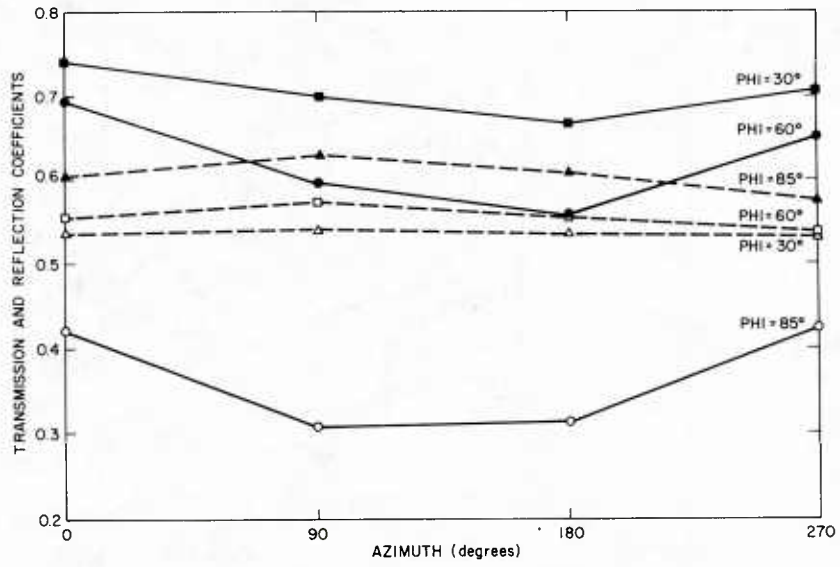


Fig. 20: Transmission coefficients (solid lines) and reflection coefficients (dashed lines) vs azimuth (measured from magnetic north) for frequency of 3 kHz, dip angle = 60°, and angle of incidence from vertical (PHI) = 30°, 60°, and 85°

3.3 Field Strength Calculations using Waveguide Modes:

For long range propagation calculations it is customary (Ref. 16) to treat the space between the earth and the ionosphere as a waveguide and to calculate the field strengths from a transmitter as a summation of waveguide modes. A popular method for achieving this is by use of the WAVEGUID computer program.

3.3.1 Description of WAVEGUID

The WAVEGUID computer program was originally developed at the Naval Electronics Laboratory Center to predict the vertical electric field received at a point on the earth's surface produced by a vertical electric transmitting antenna at another point on the earth's surface. It was later extended to give the crosspolarized (horizontal) components of the field at any height in the earth-ionosphere waveguide and even in the ionosphere itself [17, 18, 19]. The basic formulation of the equations for the fields are given in Ref. 20. The net resultant vertical electric field at the earth's surface is represented as a summation of waveguide-mode fields:

$$E_z(0) = \frac{i\sqrt{\mu_0/\epsilon_0}}{\sqrt{d\lambda h}}\, Ids \sqrt{\frac{d/a}{\sin(d/a)}} \sum_n [\sin^{3/2}\theta_n\ \Lambda_n\ \exp(i\pi/4 + ik_0 d\sin\theta_n)], \tag{2}$$

where

θ_n is the eigenangle of the nth waveguide mode,
Ids is the dipole moment (ampere-meter) of the vertical antenna located on the earth's surface,
Λ_n is the excitation factor of the nth waveguide mode,
a is the radius of the earth,
d is the great-circle distance between the transmitter and the receiver,
h is the reference height of the ionosphere (used in the definition of Λ_n),
k_0 $(= \omega/c)$ is the free-space propagation constant of the wave,
λ is the wavelength of the wave,

and

μ_0 and ϵ_0 are the magnetic permeability and dielectric constant of free space.

Each waveguide mode is characterized by its own value of attenuation rate α_n and phase velocity v_p^n, which are related to the eigenangle of the nth waveguide mode θ_n according to

$$\alpha_n = 0.02895\ \omega\ \mathrm{Im}(\sin\theta_n) \tag{3}$$

and

$$v_p^n = \frac{c}{\mathrm{Re}(\sin\theta_n)}, \tag{4}$$

where α_n is measured in decibels per 1000 km (dB/Mm) and c is the speed of light.

The values of the waveguide-mode eigenangles depend on the reflection coefficients of the ground and the ionosphere. Because of its anisotropy, the ionosphere has four reflection coefficients ${}_{\parallel}R_{\parallel}(\theta_n)$, ${}_{\perp}R_{\perp}(\theta_n)$, ${}_{\parallel}R_{\perp}(\theta_n)$, and ${}_{\perp}R_{\parallel}(\theta_n)$ for a given angle of incidence θ_n. The ground is assumed to be isotropic and to have reflection coefficients ${}_{\parallel}\bar{R}_{\parallel}(\theta_n)$ and ${}_{\perp}\bar{R}_{\perp}(\theta_n)$, which are calculable from the eigenangle, ground conductivity, and dielectric constant. The WAVEGUID program uses a procedure for calculating the ground and ionospheric reflection coefficients referenced to any height z in the waveguide. References 26, 27, and 28 discuss the reflection coefficients further. The eigenangle θ_n for a given waveguide mode is obtained by satisfying the mode equation

$$0 = \left| \begin{bmatrix} {}_{\parallel}R_{\parallel}(\theta_n) & {}_{\perp}R_{\parallel}(\theta_n) \\ {}_{\parallel}R_{\perp}(\theta_n) & {}_{\perp}R_{\perp}(\theta_n) \end{bmatrix} \begin{bmatrix} {}_{\parallel}\bar{R}_{\parallel}(\theta_n) & 0 \\ 0 & {}_{\perp}\bar{R}_{\perp}(\theta_n) \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right|. \tag{5}$$

This matrix equation is simply rewritten

$$F(\theta_n) = ({}_{\parallel}R_{\parallel}\,{}_{\parallel}\bar{R}_{\parallel} - 1)\ ({}_{\perp}R_{\perp}\,{}_{\perp}\bar{R}_{\perp} - 1) - {}_{\perp}R_{\parallel}\,{}_{\perp}\bar{R}_{\perp}\,{}_{\parallel}R_{\perp}\,{}_{\parallel}\bar{R}_{\parallel} = 0, \tag{6}$$

where the dependence of the reflection coefficient on θ_n has been suppressed to simplify the notation. The new variable $F(\theta_n)$ is defined also. Each value of θ_n which satisfies Eq. (6) is the eigenangle of the nth waveguide mode. The excitation factor Λ_n for the nth waveguide mode is obtained from

$$\Lambda_n = -i\,\frac{kh}{2}\sin\theta_n\,\frac{(1 + {}_{\parallel}\bar{R}_{\parallel})^2\,(1 - {}_{\perp}\bar{R}_{\perp}\,{}_{\perp}R_{\perp})}{{}_{\parallel}\bar{R}_{\parallel}\,\dfrac{\partial F(\theta)}{\partial\theta}\Big|_{\theta=\theta_n}}, \tag{7}$$

where the variable $F(\theta)$ is as defined in Eq. (6). In addition to the vertical electric field E_z for a given waveguide mode having eigenangle θ_n, in general there exist five other nonzero field components E_x, E_y, H_x, H_y, and H_z within the waveguide for each mode.

These extra fields arise because the ionosphere is anisotropic on account of the earth's magnetic field. Propagation through and reflection from such an anisotropic medium rotates the plane of polarization of the incident wave to generate crosspolarized reflected fields. Alternatively one could consider the linearly polarized wave transmitted from the vertical electric dipole antenna to be a linear superposition of right-handed and left-handed circularly polarized waves, each reflected with a different amplitude at the ionosphere. The ionospheric reflection causes a mixture of wave polarizations at a reception point on or above the earth's surface. The fields at height z are related to the vertical electric fields at the earth's surface $E_z(0)$ in a given waveguide mode by the following functions of reflection coefficients and Hankel functions:

$$E_z(z)/E_z(0) = f_{\parallel}(z), \tag{8}$$

$$E_x(z)/E_z(0) = g(z)/S, \tag{9}$$

$$E_y(z)/E_z(0) = \frac{{}_{\parallel}R_{\perp}(1+{}_{\perp}\bar{R}_{\perp})\, f_{\perp}(z)}{S(1+{}_{\parallel}\bar{R}_{\parallel})\,(1-{}_{\perp}\bar{R}_{\perp}{}_{\perp}R_{\perp})}, \tag{10}$$

$$H_z(z)/E_z(0) = -\frac{{}_{\parallel}R_{\perp}(1+{}_{\perp}\bar{R}_{\perp})\, f_{\perp}(z)}{\eta(1+{}_{\parallel}\bar{R}_{\parallel})\,(1-{}_{\perp}\bar{R}_{\perp}{}_{\perp}R_{\perp})}, \tag{11}$$

$$H_y(z)/E_z(0) = \frac{f_{\parallel}(z)}{\eta S}, \tag{12}$$

and

$$H_x(z)/E_z(0) = \frac{{}_{\parallel}R_{\perp}(1+{}_{\perp}\bar{R}_{\perp})}{i\,\eta\,k\,S(1+{}_{\parallel}\bar{R}_{\parallel})\,(1-{}_{\perp}\bar{R}_{\perp}{}_{\perp}R_{\perp})}\,\frac{df_{\perp}(z)}{dz}, \tag{13}$$

where

$$f_{\parallel}(z) = \exp\left(\frac{z-D}{a}\right)\,\frac{F_1 h_1(q) + F_2 h_2(q)}{F_1 h_1(q_d) + F_2 h_2(q_d)}, \tag{14}$$

$$f_{\perp}(z) = \frac{F_3 h_1(q) + F_4 h_2(q)}{F_3 h_1(q_d) + F_4 h_2(q_d)}, \tag{15}$$

and

$$g(z) = \frac{1}{ik}\frac{d}{dz}[f_{\parallel}](z)], \tag{16}$$

in which

$$F_1 = -\left[H_2(q_0) - i\,\frac{n_0^2}{N_g^2}\left(\frac{ak}{2}\right)^{1/3}(N_g^2 - S^2)^{1/2}\,h_2(q_0)\right], \tag{17}$$

$$F_2 = H_1(q_0) - i\,\frac{n_0^2}{N_g^2}\left(\frac{ak}{2}\right)^{1/3}(N_g^2 - S^2)^{1/2}\,h_1(q_0), \tag{18}$$

$$F_3 = -\left[h'_2(q_0) - i\left(\frac{ak}{2}\right)^{1/3}(N_g^2 - S^2)^{1/2}\,h_2(q_0)\right], \tag{19}$$

$$F_4 = h'_1(q_0) - i\left(\frac{ak}{2}\right)^{1/3}(N_g^2 - S^2)^{1/2}\,h_1(q_0), \tag{20}$$

$$q = \left(\frac{2}{ak}\right)^{-2/3}\left[C^2 - \frac{2}{a}(h - z)\right], \tag{21}$$

and

$$q_d = \left(\frac{2}{ak}\right)^{-2/3}\left[C^2 - \frac{2}{a}(h - D)\right], \tag{22}$$

with

$$q_0 = \left(\frac{2}{ak}\right)^{-2/3}\left[C^2 - \frac{2h}{a}\right], \tag{23}$$

$$H_j(q) = h'_j(q) + \frac{1}{2}\left(\frac{2}{ak}\right)^{2/3}h_j(q), \quad j = 1,2, \tag{24}$$

$$n^2 = 1 - \frac{2}{a}(h - z), \tag{25}$$

$$n_0^2 = 1 - \frac{2}{a}h, \tag{26}$$

and

$$N_g^2 = \epsilon/\epsilon_0 - j\,\frac{\sigma}{\omega\epsilon}. \tag{27}$$

In Eq. 27 σ and ϵ refer to the conductivity and dielectric constant of the earth.

3.3.2 Typical Waveguide Mode Parameters

Using the WAVEGUID program, one can calculate the propagation parameters of the various modes. Figures 21 through 23 from Ref. 21 give the excitation factors and attenuation rates for the three lowest order waveguide modes in the ELF band and lower VLF band. Figures 24 through 26 show the behavior of the magnetic fields versus height in the waveguide and inside the lower ionosphere. The nighttime ionosphere of Fig. 18 was used. The direction of propagation is toward magnetic north for a dip angle of 60°. Using these propagation parameters and Eqs. 7 through 27, one can estimate the field strength produced by a transmitter. Additional equations and sample parameters for this task are given in Ref. 16.

Figures 27 through 35 give propagation parameters for the first three waveguide modes calculated for a typical daytime isotropic ionosphere and used in the noise prediction computer program described in Ref. 22. These are similar to but more comprehensive than the parameters given in Ref. 23.

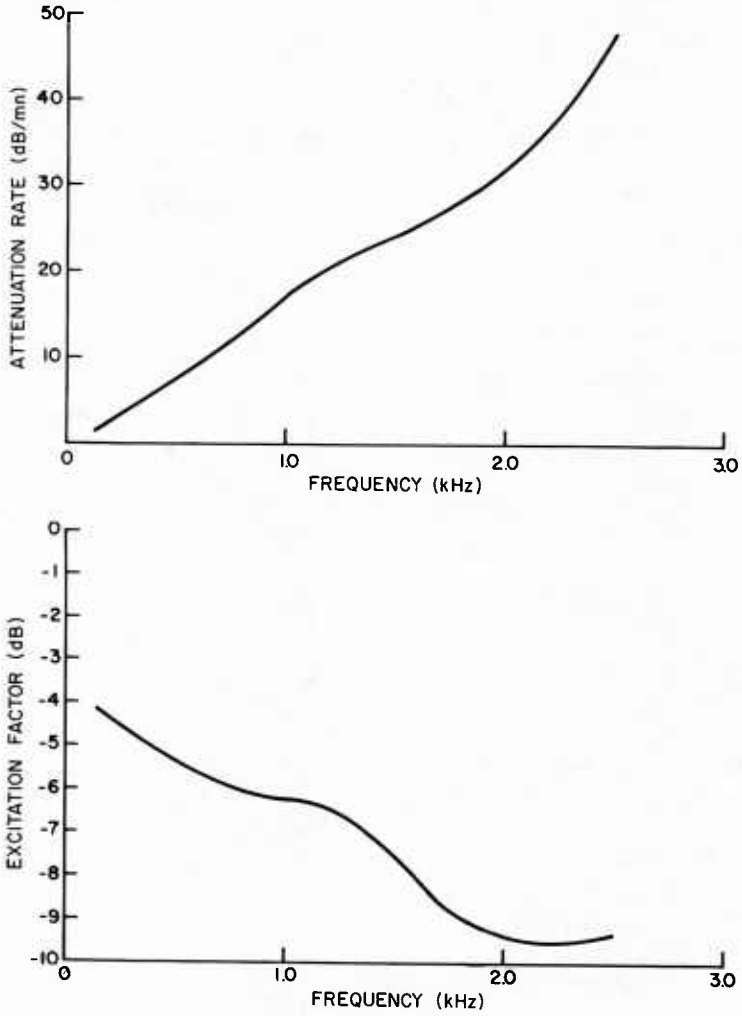


Fig. 21: Lowest-order TEM mode propagation parameters vs frequency

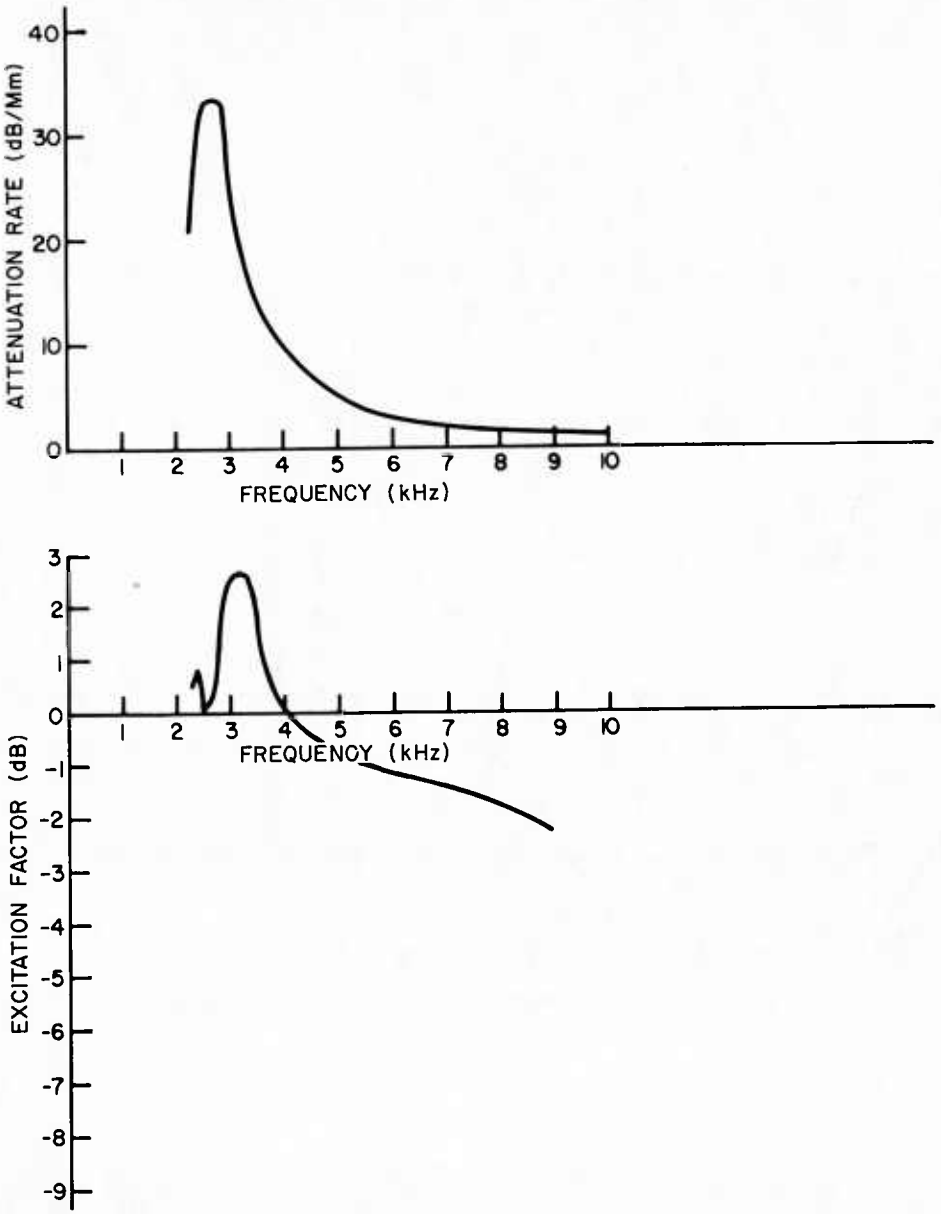


Fig. 22: First-order TM mode propagation parameters vs frequency

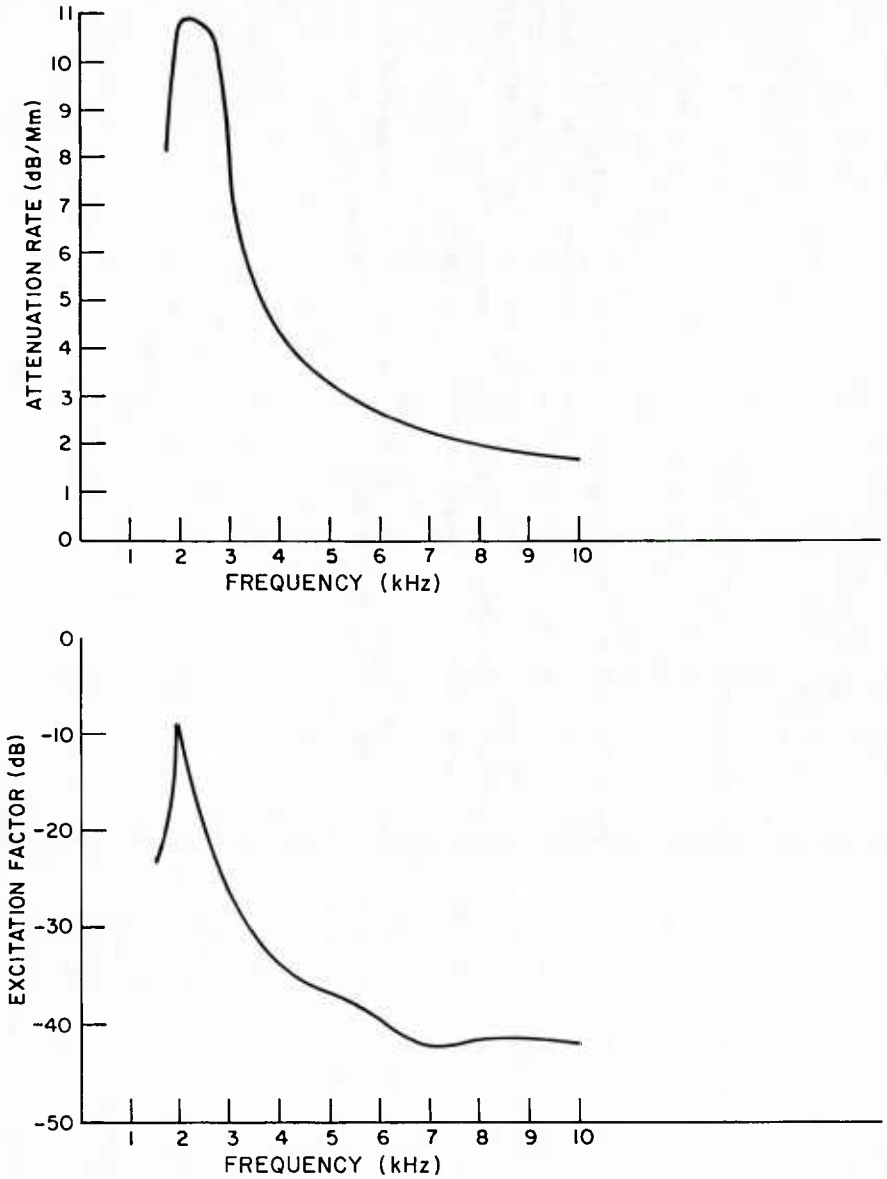


Fig. 23: First-order TE mode propagation parameters vs frequency

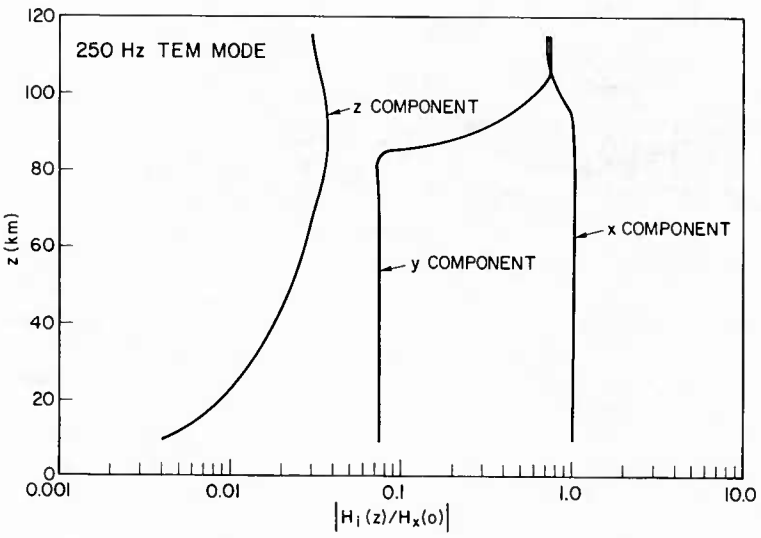


Fig. 24: Relative excitation efficiencies vs height of the zero-order TEM mode at 0.25 kHz

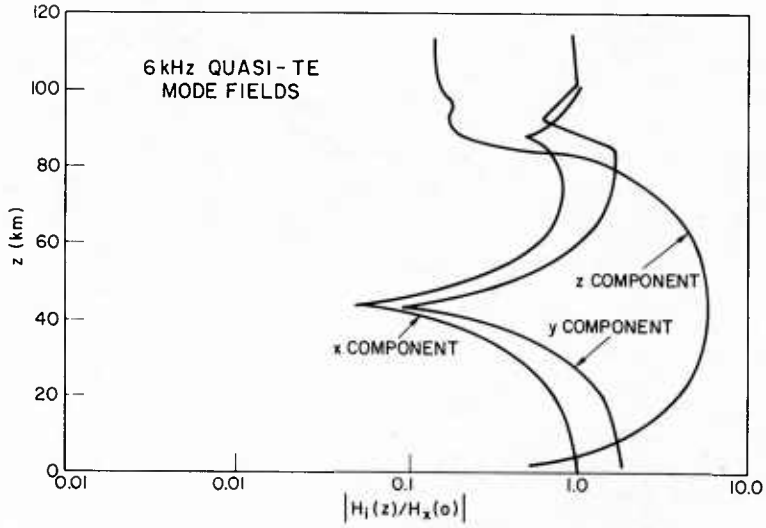


Fig. 25: Relative excitation efficiency vs height of the first-order TE mode at 6.0 kHz

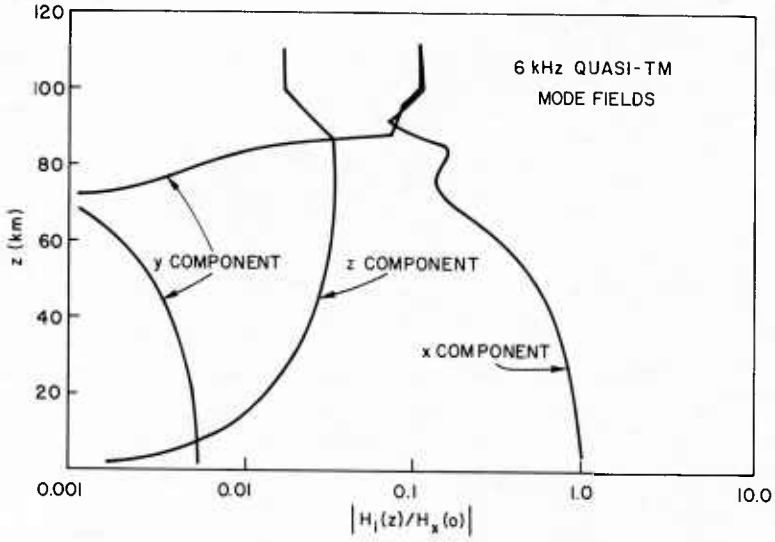


Fig. 26: Relative excitation efficiency vs height of the first-order TM mode at 6.0 kHz

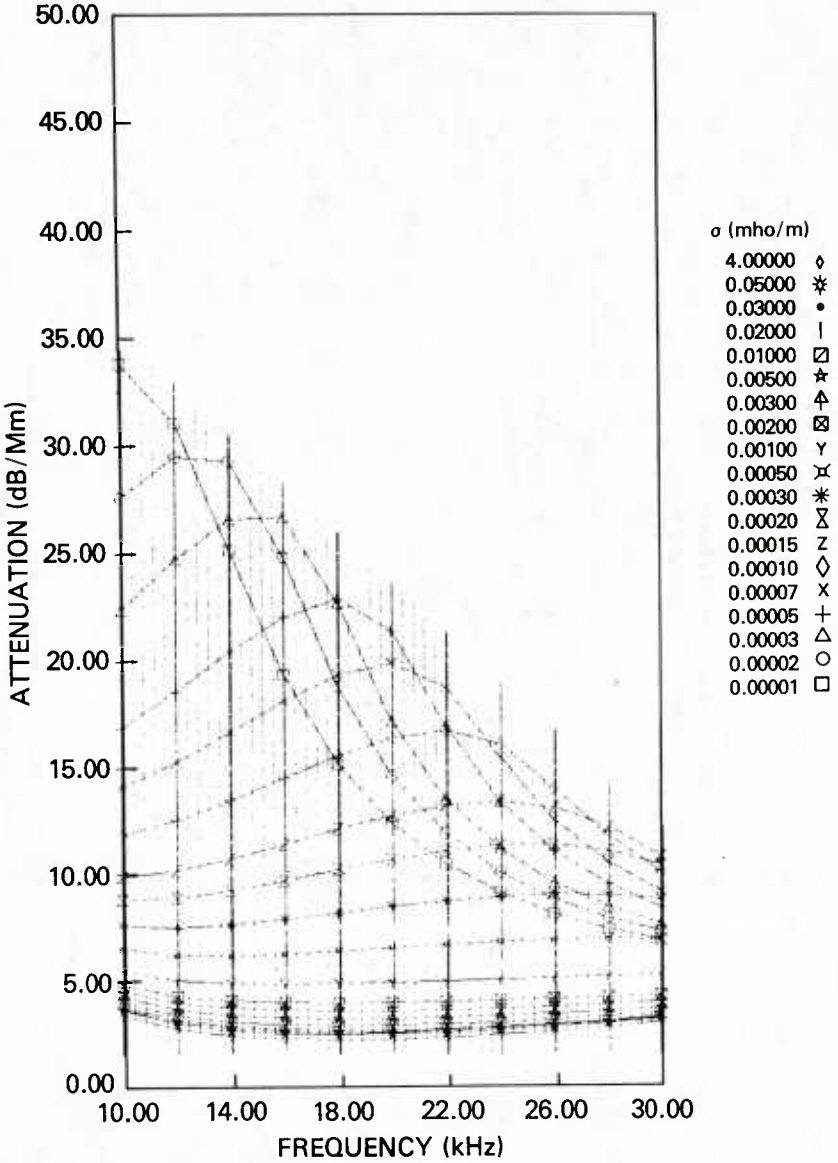


Fig. 27: Attenuation rate for the first order TM mode vs frequency and ground conductivity for an isotropic daytime ionosphere

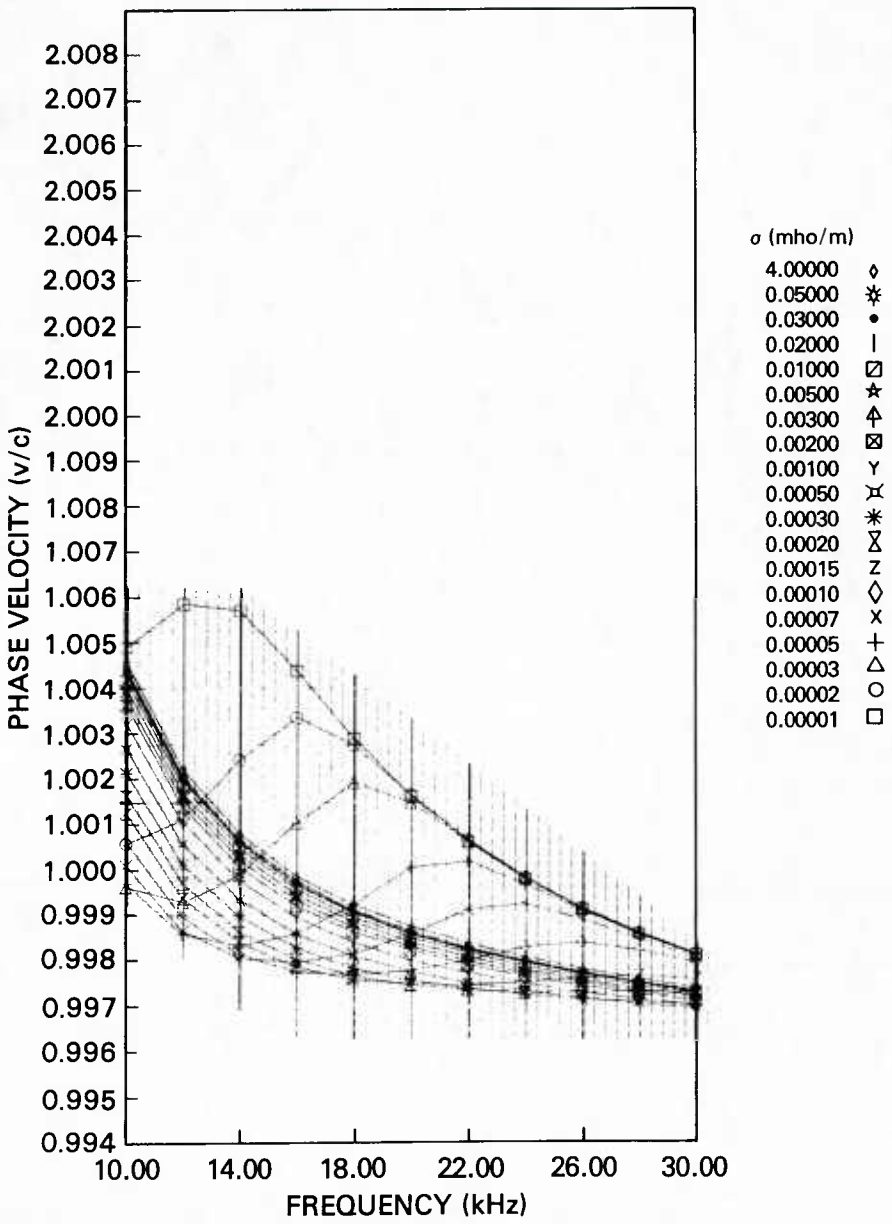


Fig. 28: Phase velocity for the first order TM mode vs frequency and ground conductivity from isotropic daytime ionosphere

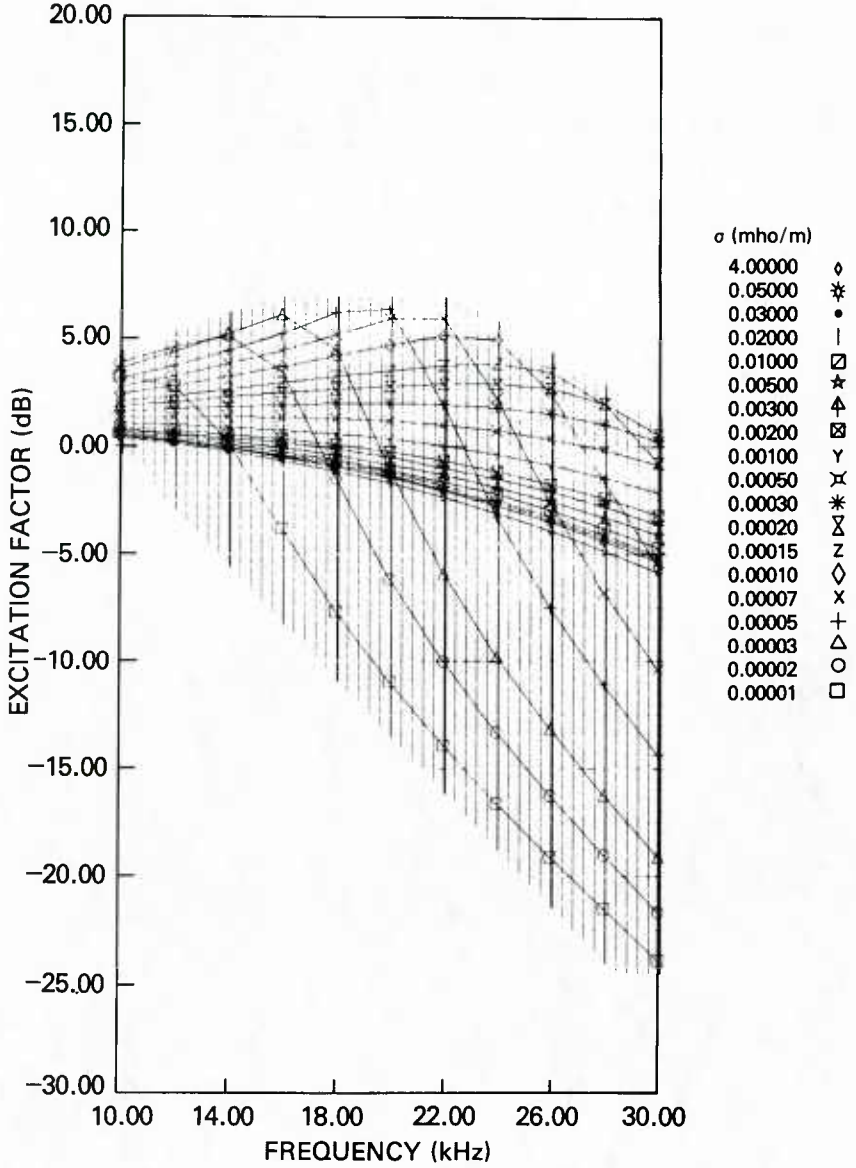


Fig. 29 Excitation factor for the first order TM mode vs frequency and ground conductivity from isotropic daytime ionosphere

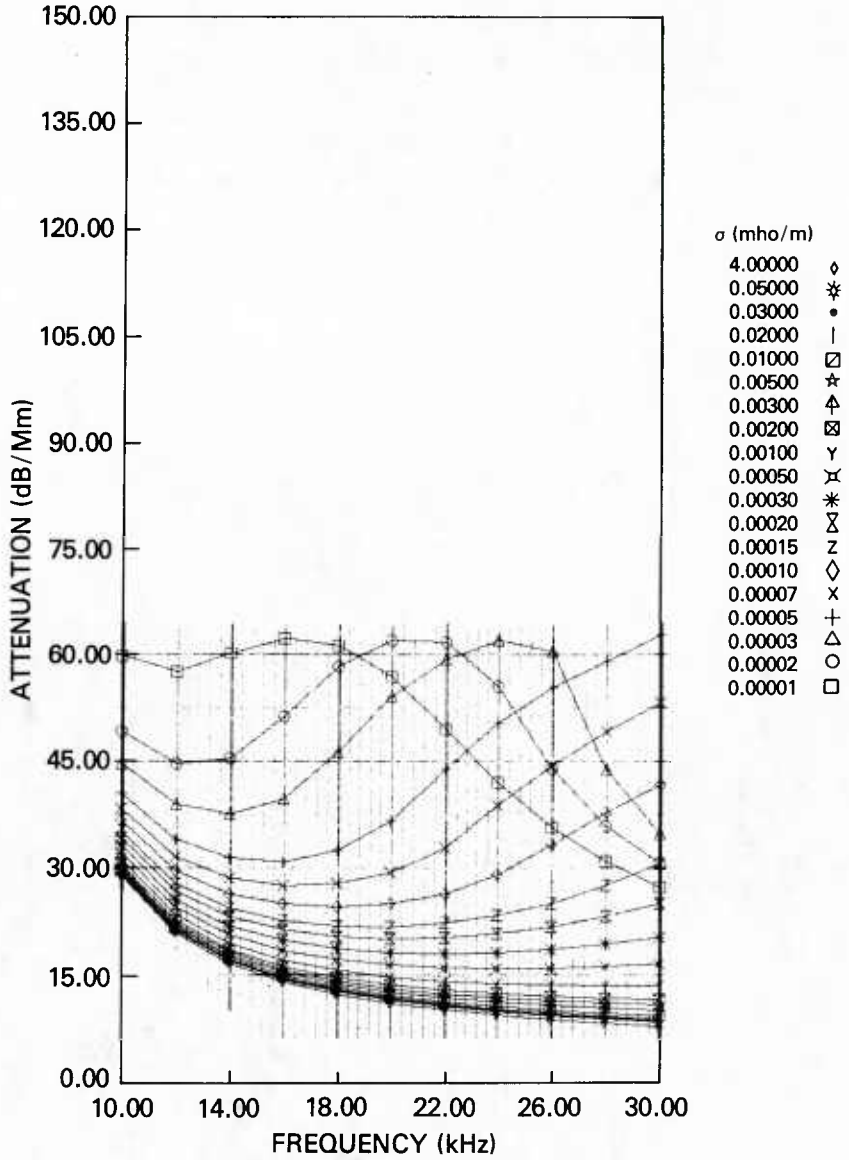


Fig. 30: Attenuation rate for the second order TM mode vs frequency and ground conductivity for an isotropic daytime ionosphere

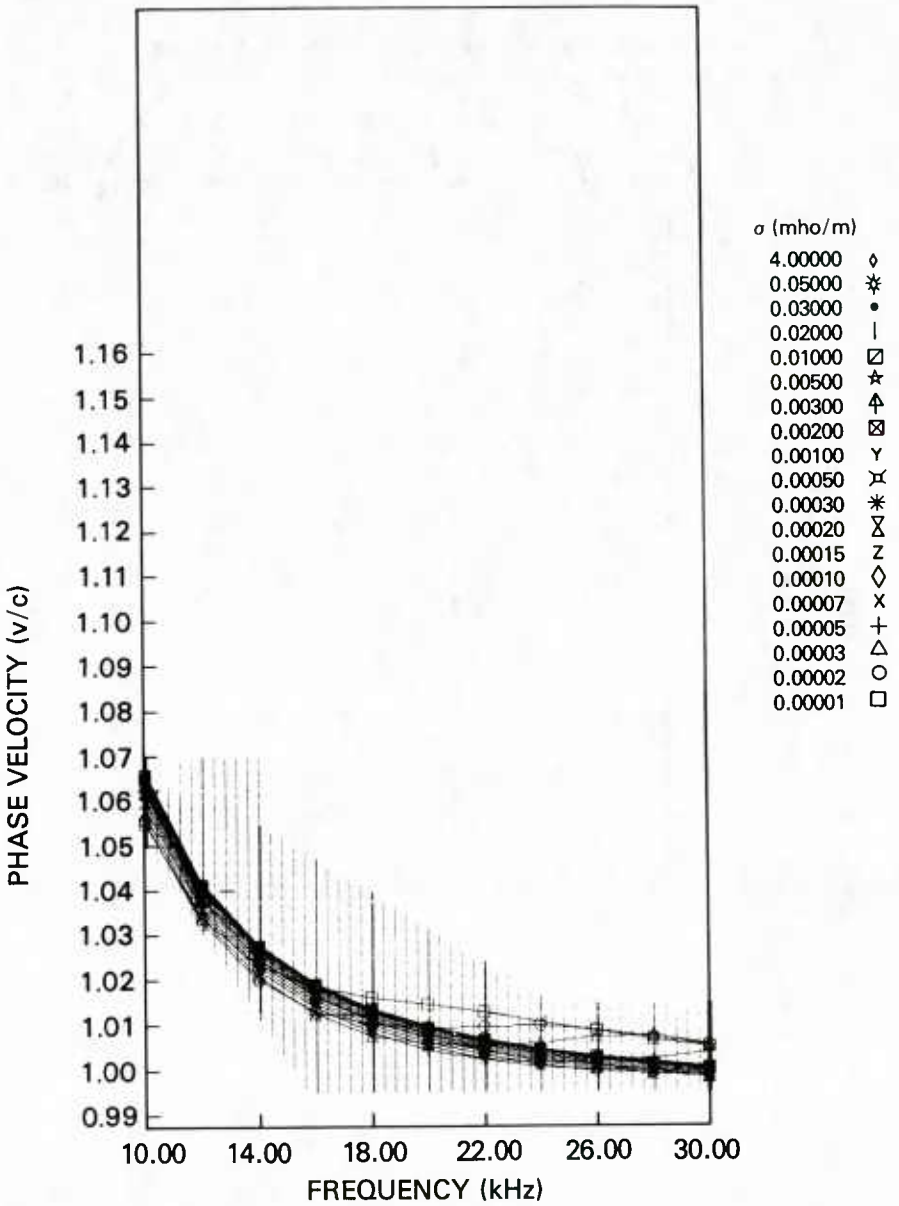


Fig. 31: Phase velocity for the second order TM mode vs frequency and ground conductivity for an isotropic daytime ionosphere

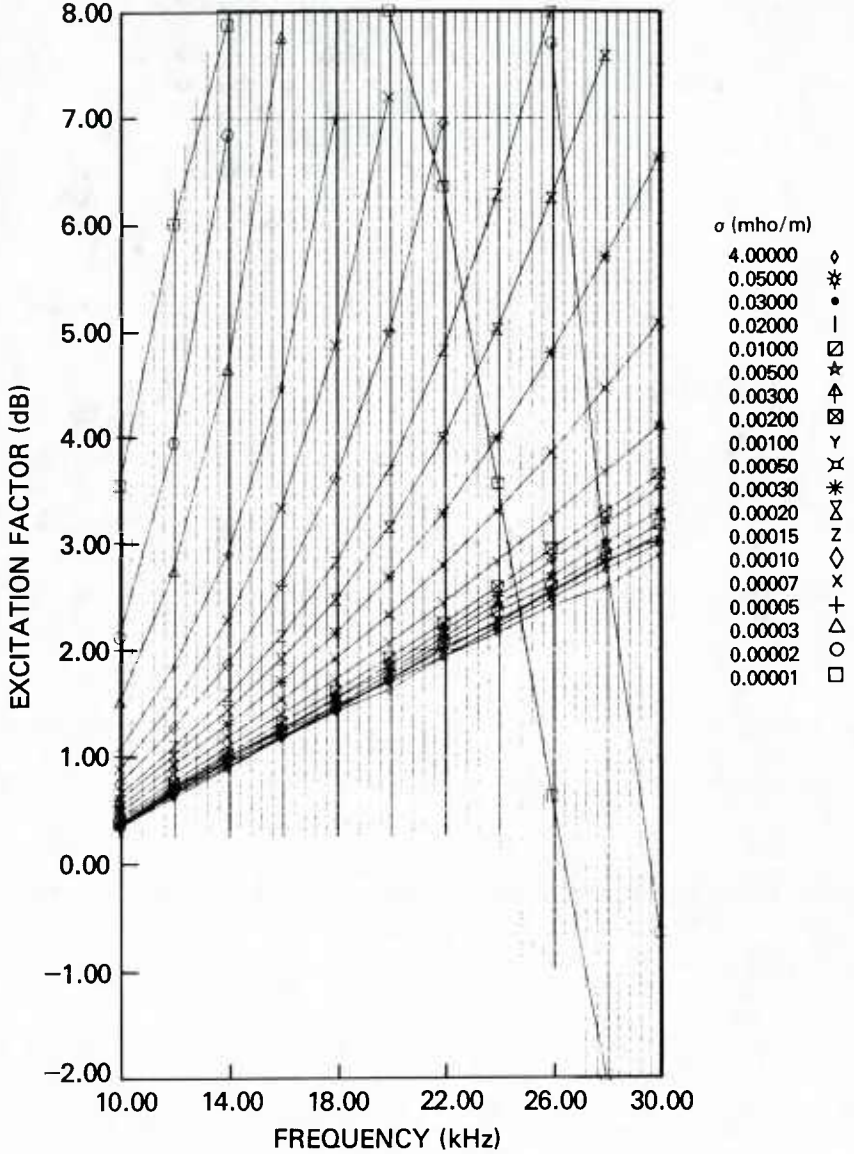


Fig. 32: Excitation factor for the second order TM mode vs frequency and ground conductivity for an isotropic daytime ionosphere

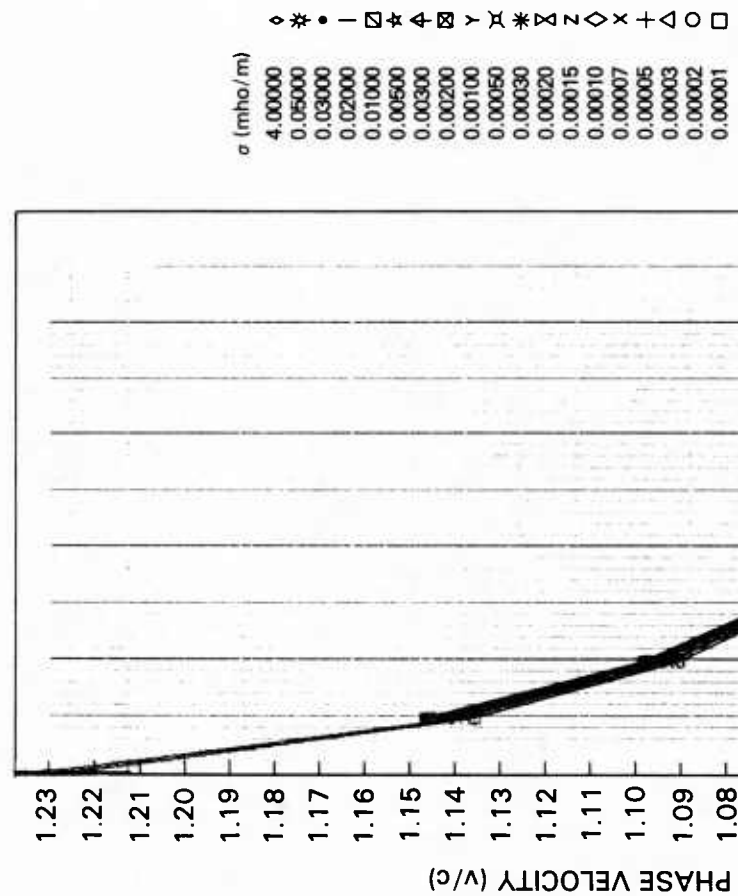


Fig. 34: Phase velocity for the third order TM mode vs frequency and ground conductivity from isotropic daytime ionosphere

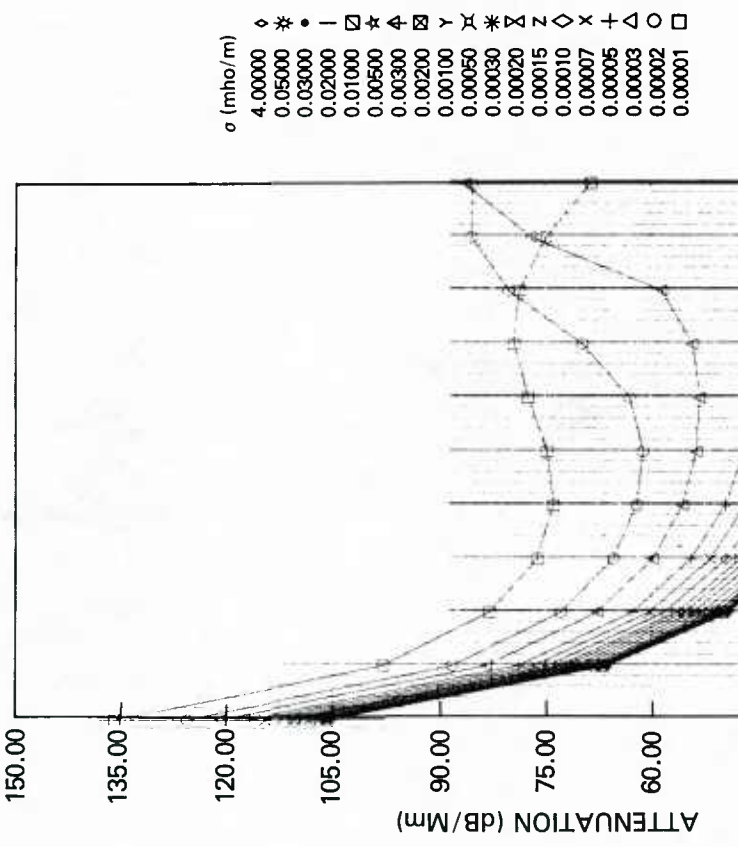


Fig. 33: Attenuation rate for the third order TM mode vs frequency and ground conductivity for an isotropic daytime ionosphere

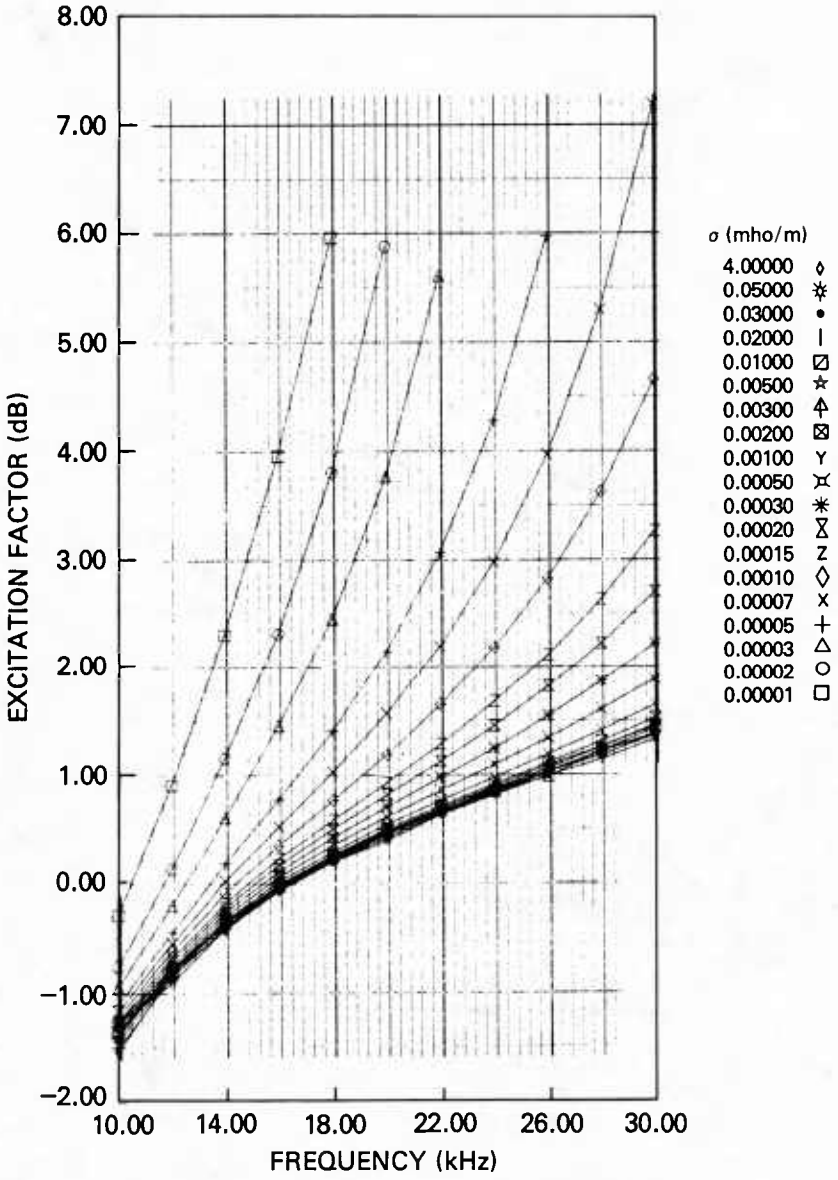


Fig. 35: Excitation factor for the third order TM mode vs frequency and ground conductivity for an isotropic daytime ionosphere

3.3.3 Field Strength Calculations and Measurements:

As shown in Ref. 24 field strength versus distance curves can be readily calculated from these parameters. Figures 36 through 41 taken from Ref. 25 illustrate the pattern of field strength versus distance for a one kilowatt radiated power transmitter propagating over sea water at 24.0 kHz. The very sharp predicted interference hull at 2.4 Mm is to be noted are Fig. 36. Using the parameters given in Ref. 23, a statistical prediction of field strength versus distance was made by a Monte Carlo technique. A gaussian distribution of height, h, and conductivity gradient, β, were assumed having an average $h = 70$ km and average $\beta = .5\ \text{km}^{-1}$. The standard deviation of the height was 1 km and of β was $.05\ \text{km}^{-1}$. This method predicted the greatest variability of signal level to occur near the interference nulls at 2.4 Mm and 3.8 Mm. Experimental data was taken for a period during summer 1970 near the 2.4 Mm null produced by the NBA, Balboa transmitter which was operating on 24.0 kHz (Ref. 26). The measured standard deviation was much smaller than that predicted from the model. It was also smaller than is usually observed with long term data-taking. It was concluded from this data that during the short period of this experiment the ionosphere was highly repeatable, even though it shows greater variability over longer durations.

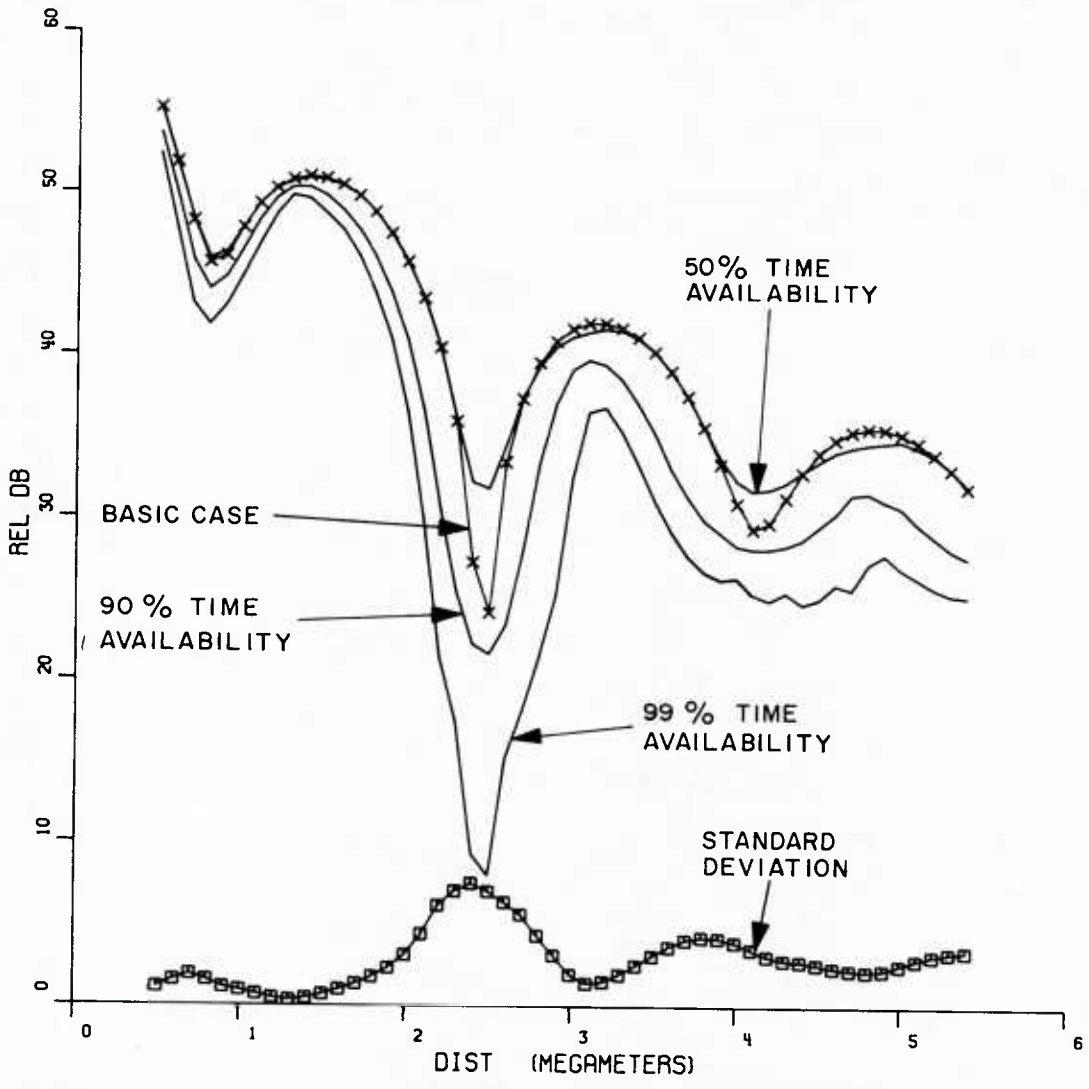


Fig. 36: Plot of field strengths having 99%, 90%, and 50% time availabilities vs distance. Plot of standard deviation of field strength vs distance. Plot of basic multimode field strength ($\beta = 0.5\ \text{km}^{-1}$ $h = 70$ km) case.

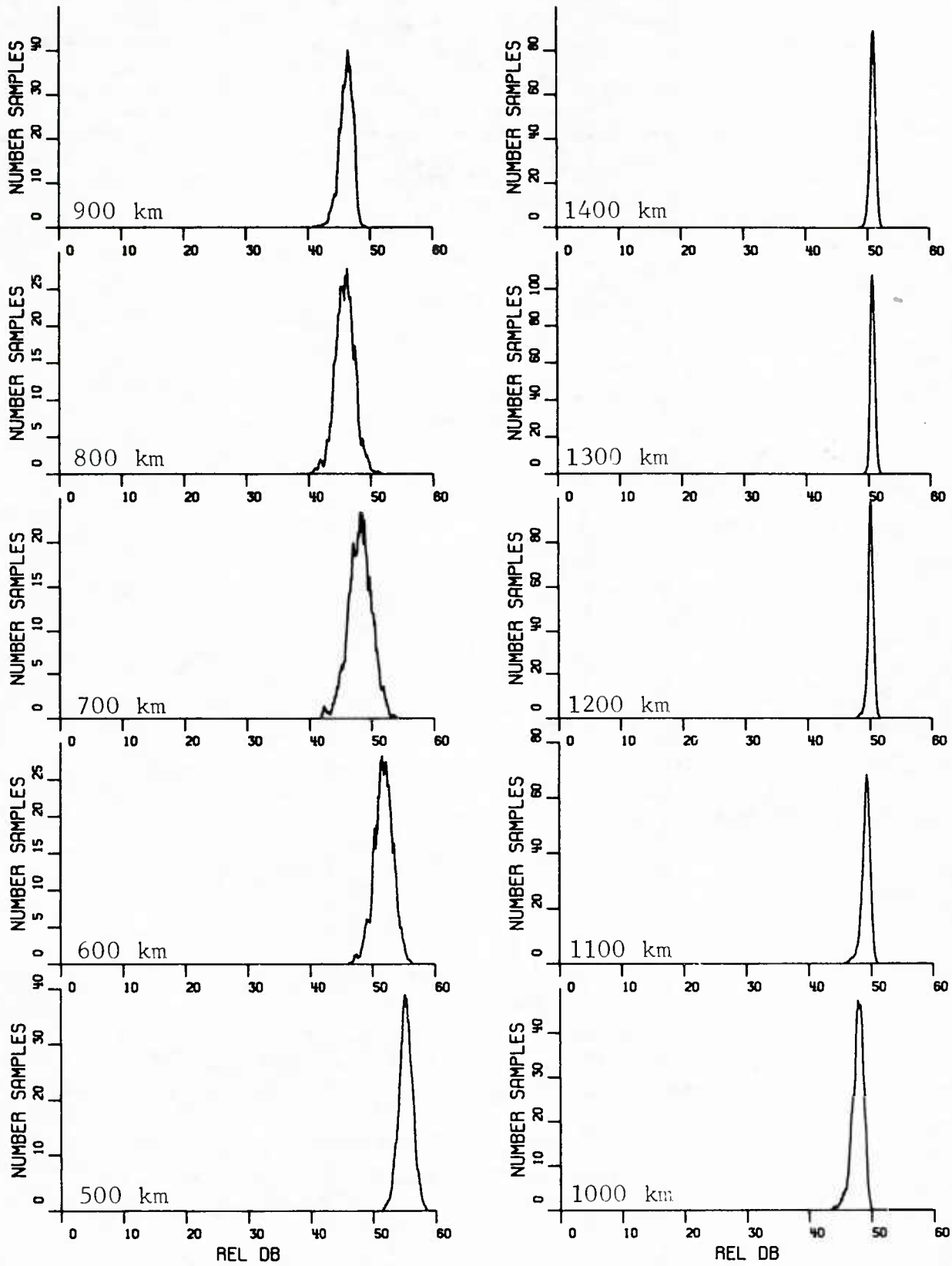


Fig. 37: Histograms of calculated field strengths

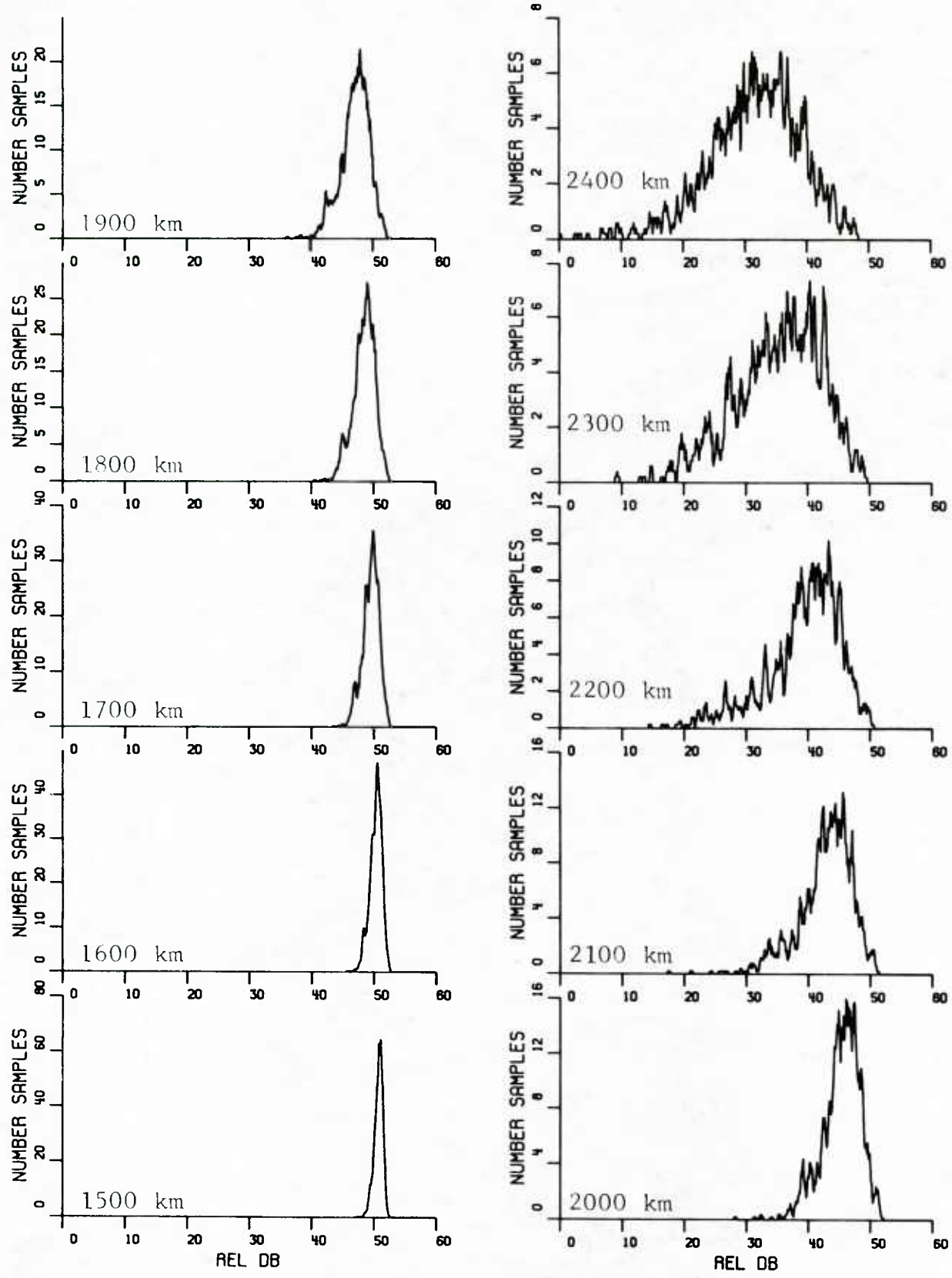


Fig. 38: Histograms of calculated field strengths

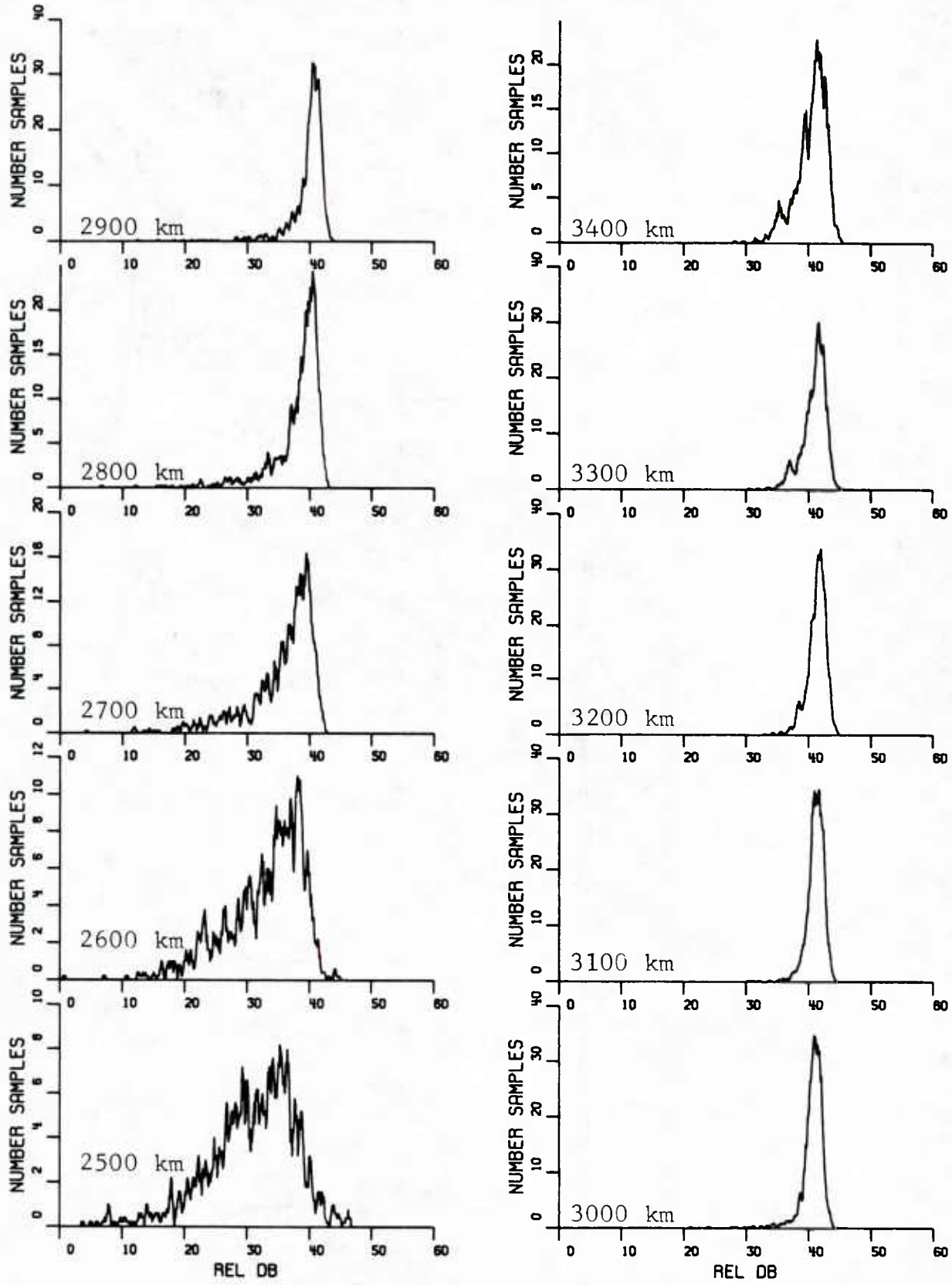


Fig. 39: Histograms of calculated field strengths

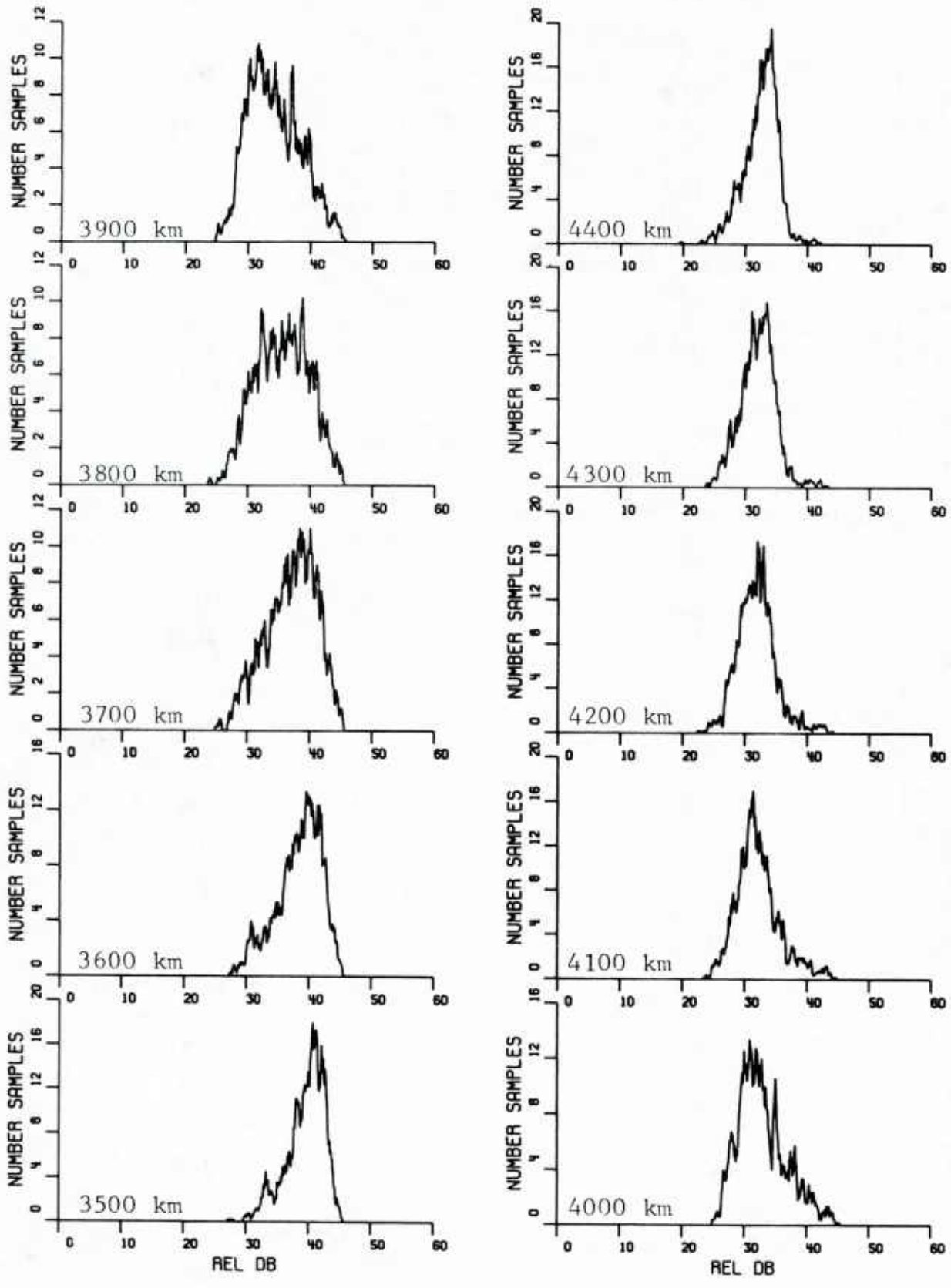


Fig. 40: Histograms of calculated field strengths

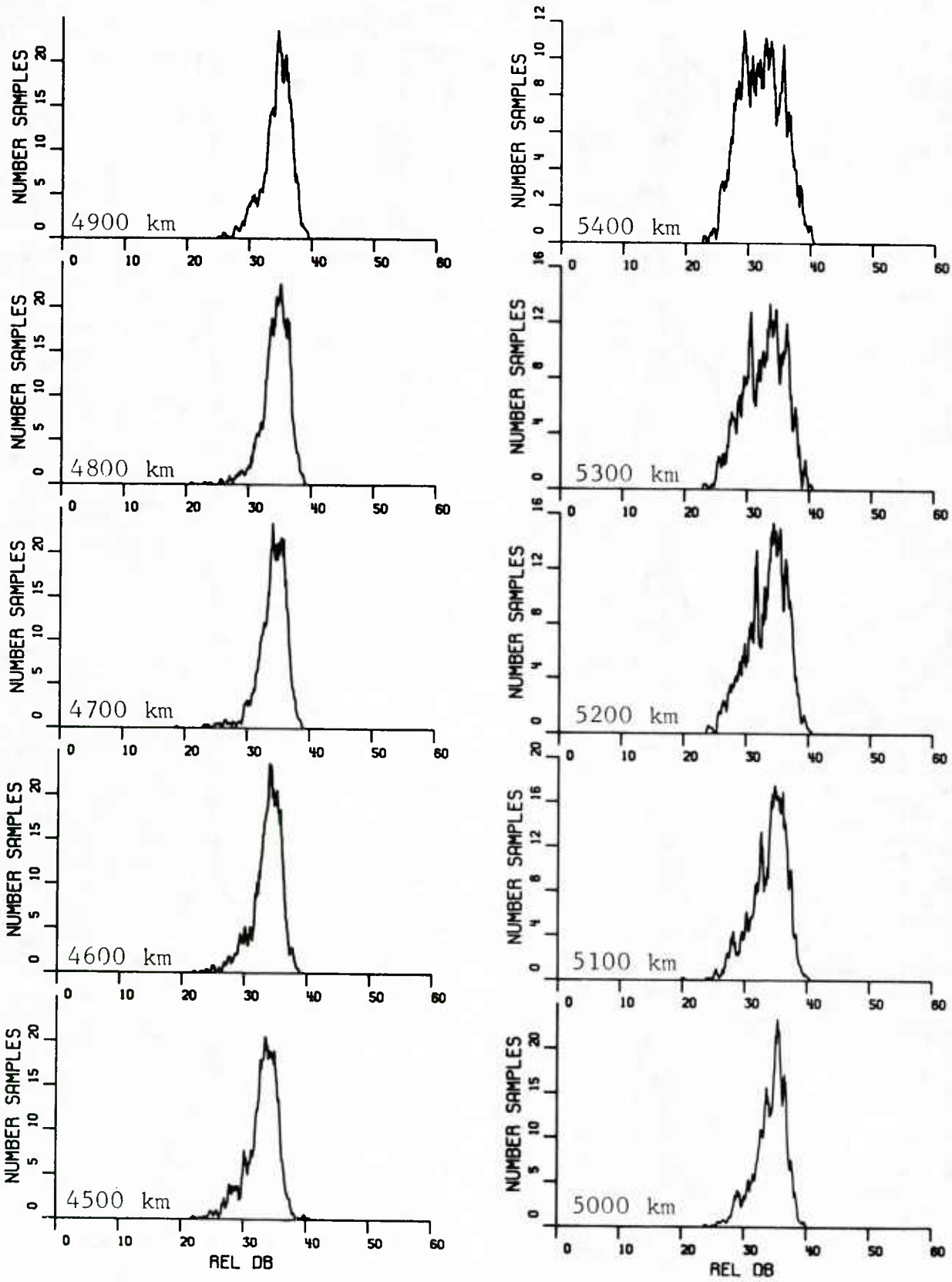


Fig. 41: Histograms of calculated field strengths

3.3.4 DCA Recommended Ionospheres:

From an analysis of field strength versus distance data similar to those shown in Figs. 42 and 43, the U.S. Defense Communications Agency (Ref. 27) recommended the ionosphere profiles shown in Table 1. Graphs showing these ionosphere parameters are shown in Figs. 44 and 45 (taken from Ref. 27). The electron density $N(z)$ versus height, z, in kilometers for a particular β and h is given by Eq. 28.

$$N(z) = \left\{1.43 \cdot 10^7 \frac{e^{-1}}{\text{cm}^3} \exp(-0.15h)\right\} \left\{\exp[(\beta - 0.15)(z - h)]\right\} \tag{28}$$

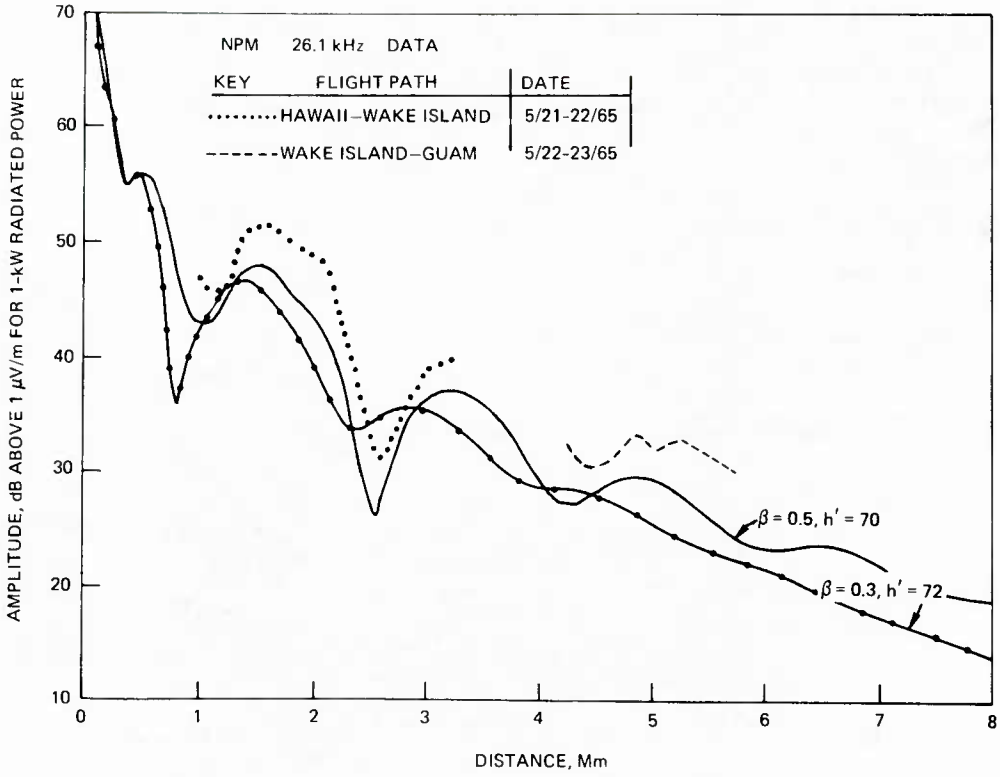
The collision frequency versus height is given by Eq. 29.

$$\nu(z) = \nu_o \exp(-\alpha z) \tag{29}$$

where

$$\nu_o = 1.82 \times 10^{11} \text{ collision/sec} \qquad \alpha = 0.15 \text{ km}^{-1}$$

From a later analysis of multifrequency mid-latitude day and nighttime data (Ref. 28) taken on flights from Hawaii to California, the recommendations of Table 2 are given. It is significant that for nighttime the laterally homogeneous ionospheric profile must be assumed to change with broadcast frequency. This unphysical situation reflects our current lack of understanding of the nighttime lower ionosphere.



Flight Path: Hawaii to Wake Island to Guam (daytime)
Ground Conductivity: Sea water
Transmitter: NPM, Hawaii at 26.1 kHz
Date: June 1965
Electron Density Profiles: $\beta = 0.5$ km^{-1}, h' = 70 km with base of the ionosphere at 50 km, and
$\beta = 0.3$ km^{-1}, h' = 72 km with base of the ionosphere at 30 km.
Method of Computation: Horizontally homogenous waveguide

Fig. 42: Propagation over the Pacific Ocean (daytime, summer) (26.1 kHz)

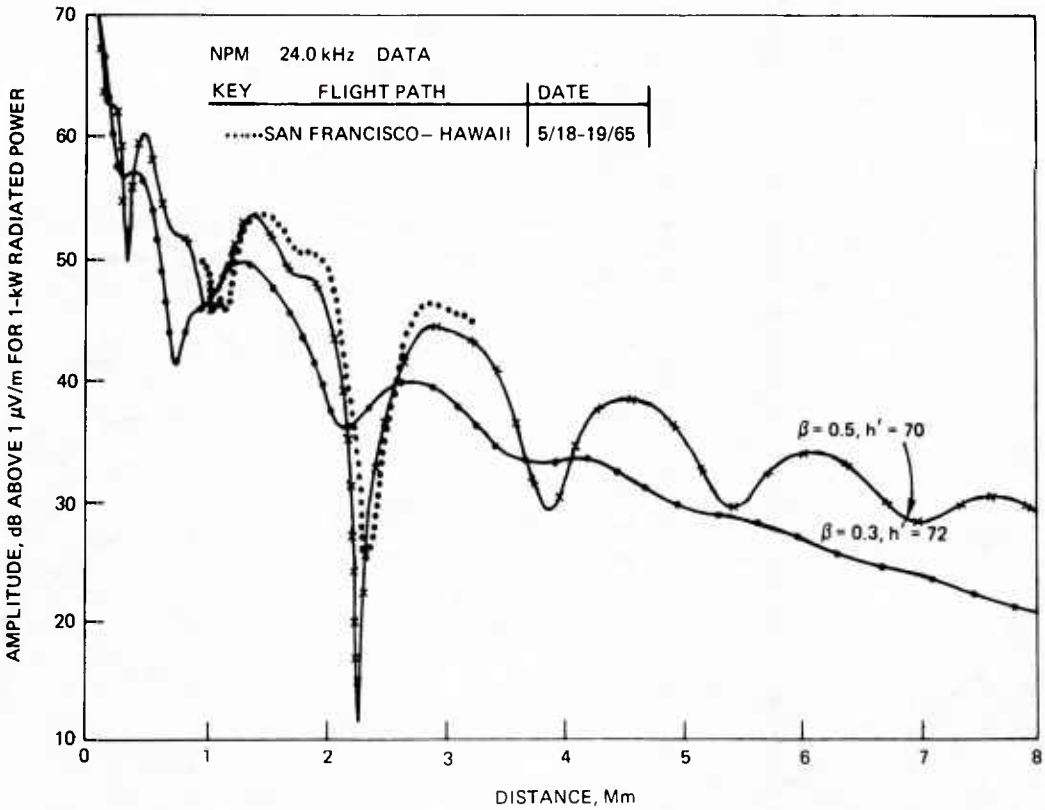


Fig. 43: Propagation over the Pacific Ocean (daytime, summer) (24 kHz)

Table 1 — DCA (Ref. 27) Recommended Profiles to Use in WAVEGUIDE OR WAVEHOP Propagation Programs

	DAYTIME		NIGHTTIME	
	Summer	Winter	Summer	Winter
High Latitude	$\beta = 0.3, h' = 72$	$\beta = 0.3, h' = 72$	$\beta = 0.5,$	$h' = 87$
Low Latitude	$\beta = 0.5, h' = 70$	$\beta = 0.3, h' = 72$		

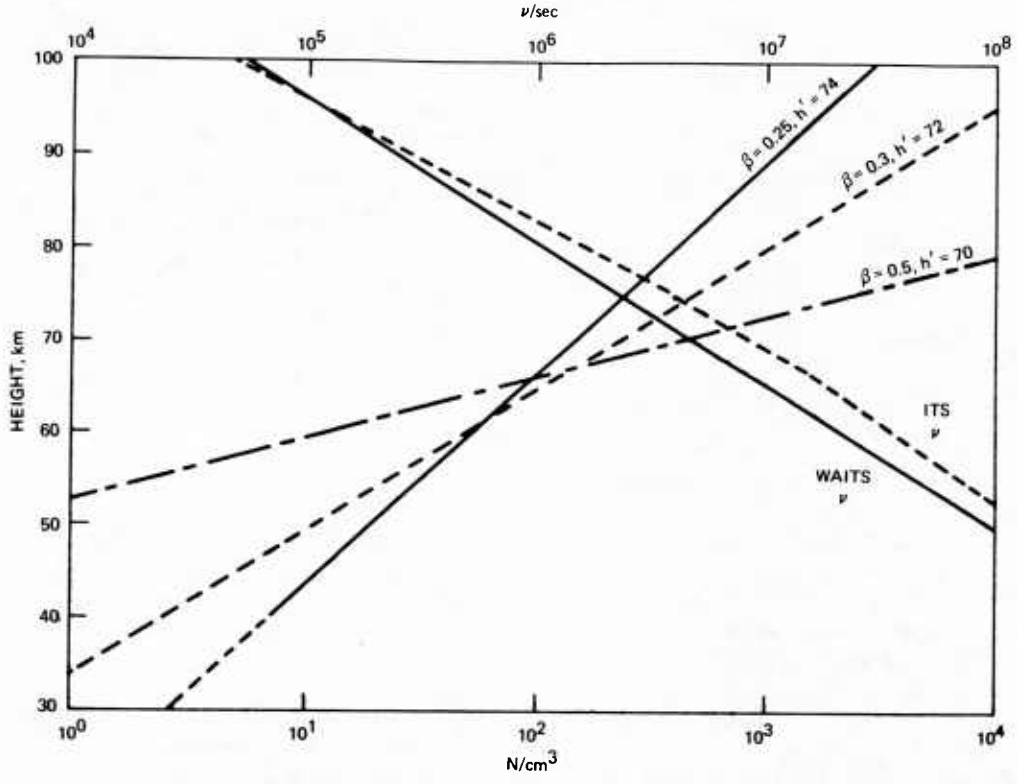


Fig. 44: Daytime Electron Density Profiles and Collision Frequency Profiles

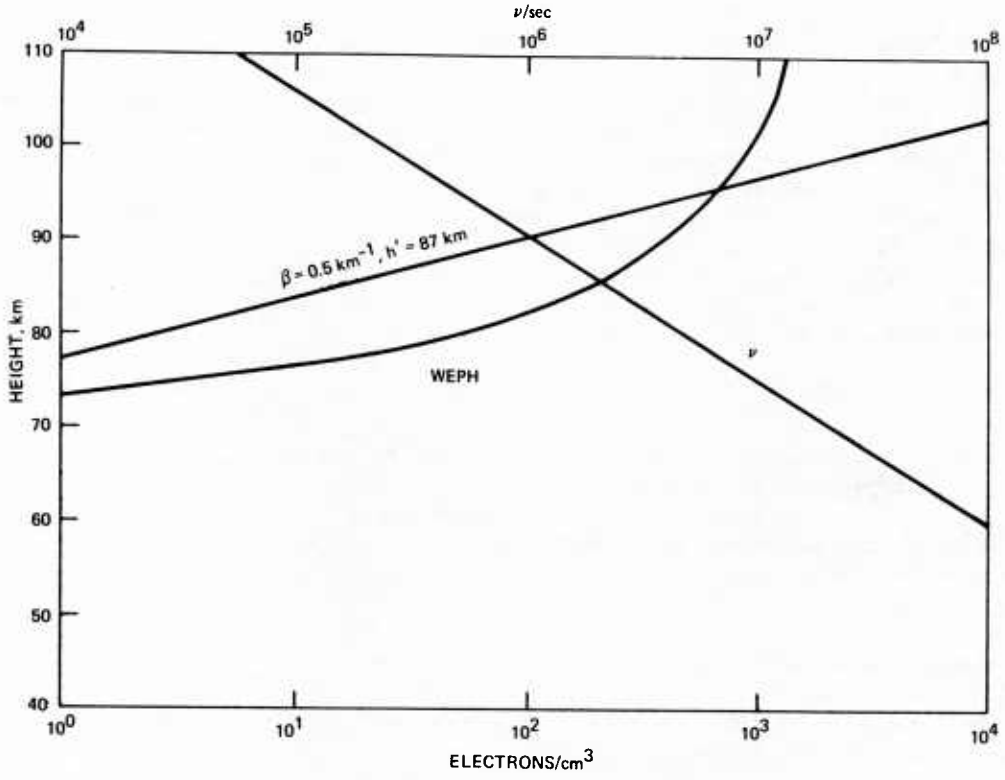


Fig. 45: Nighttime Electron Density Profiles and Collision Frequency Profile

Table 2 — DCA (Ref. 28) Recommended Electron Density Profiles for Use in Propagation Predictions (midlatitude)

DAYTIME Winter		NIGHTTIME Winter	
Frequency (kHz)	Profile β (km^{-1}), H' (km)	Frequency (kHz)	Profile β (km^{-1}), H' (km)
9-60	$\beta = 0.3, H' = 74$	below 10	$\beta = 0.3, H' = 87$
		10 - 15	$\beta = 0.4, H' = 87$
		15 - 25	$\beta = 0.5, H' = 87$
		25 - 30	$\beta = 0.6, H' = 88$
		30 - 40	$\beta = 0.7, H' = 88$
		40 - 60	$\beta = 0.8, H' = 88$

3.3.5 Coverage Prediction Programs for Inhomogeneous Paths

Up to this point we have discussed prediction methods for homogeneous paths only. However the earth's surface, earth's magnetic field and the ionosphere typically vary with distance along a propagation path. References 32 and 33 represent early efforts to apply mode theory to these inhomogeneous paths. References 34 through 37 describe further developments and elaboration of the waveguide mode propagation picture using mode-matching and the WKB approximation to propagate the waves through spatially varying portions of the path.

Reference 38 describes a Very Low Frequency Automatic Computation Method (VLFACM) computer model. It is approximate but it is inexpensive to run (Ref. 39). It makes predictions with built in easily-computed values for the attenuation and excitation of the signal. VLFACM only computes a single dominant waveguide mode and operates best from 15 kHz to 30 kHz. The model parameters were selected by comparison of the program predictions with a great many propagation paths. VLFACM contains expressions which permit it to give a complete diurnal prediction taking solar zenith angle, geomagnetic azimuth and ground conductivity into account.

3.3.6 DCA Arctic Propagation Recommendation and Recent Experiments:

The DCA recommendations (Ref. 27) for Arctic or high-latitude propagation conditions were based on the comparisons shown in Figs. 46 through 53. Notice that in all cases the field strength predictions are noticeably higher than the measured values for waves propagating across the Greenland Ice Cap and that the conductivity of the ice cap is assumed to be 10^{-5} mhos/meter in conformity with Refs. 29 and 30. However at 25 kHz Ref. 31 gives the measured conductivity of an antarctic glacier as 5.0 to 6.1 $\times$ 10^{-5} mho/1 m with a relative dielectric constant of approximately 11. Referring to Fig. 27 we see that this conductivity change could raise the attenuation rate over Greenland at 24 hKz from 9 dB for $\sigma = 10^{-5}$ mho/m to between 13 and 15 dB/Mm for the larger values of σ given by Ref. 31. Recent measurements of the 24.0 kHz signal from NAA indicate an attenuation rate over Greenland in better agreement with the higher attenuation rate, suggesting that the conductivity and dielectric constant of the Greenland Ice Cap may need modification.

3.3.7 Coverage Problems

The entire question of the actual coverage of longwave systems is currently in a status of reevaluation and flux. As one might guess from the meager amount of airborne propagation data available and the tendency of the ionosphere not to stand still, we find that our coverage prediction techniques are unable to achieve the accuracy that will readily satisfy a system designer. Because significantly large sums of money are involved in the purchase of aircraft squadrons or in the building of a huge ground based transmitters, greater emphasis and resources should be and are being allocated to longwave coverage research.

3.4 Wavehop Coverage Predictions

The solution of the Helmholtz equation for the Hertz vector between concentric spherical shells (Ref. 40) is given by a summation of products of spherical harmonics functions (Eq. 30).

$$u \sim \frac{\overline{ikC}}{4\pi} \sum_{q=o}^{\infty} (2q+1)\, h_q^{(2)}(kb)\, h_q^{(1)}(kr)\, P_q(\cos\Theta) \frac{F_q}{D_q} \tag{30}$$

Watson (Ref. 41) discovered a technique for transforming this slowly converging series into an integral in the complex plane and re-expressing the integral as a summation over residues.

$$u = -i \int_L \frac{(\nu+1/2)}{\sin \nu\pi} f(\nu) P_\nu[\cos(\pi-\Theta)]\, d\nu \tag{31}$$

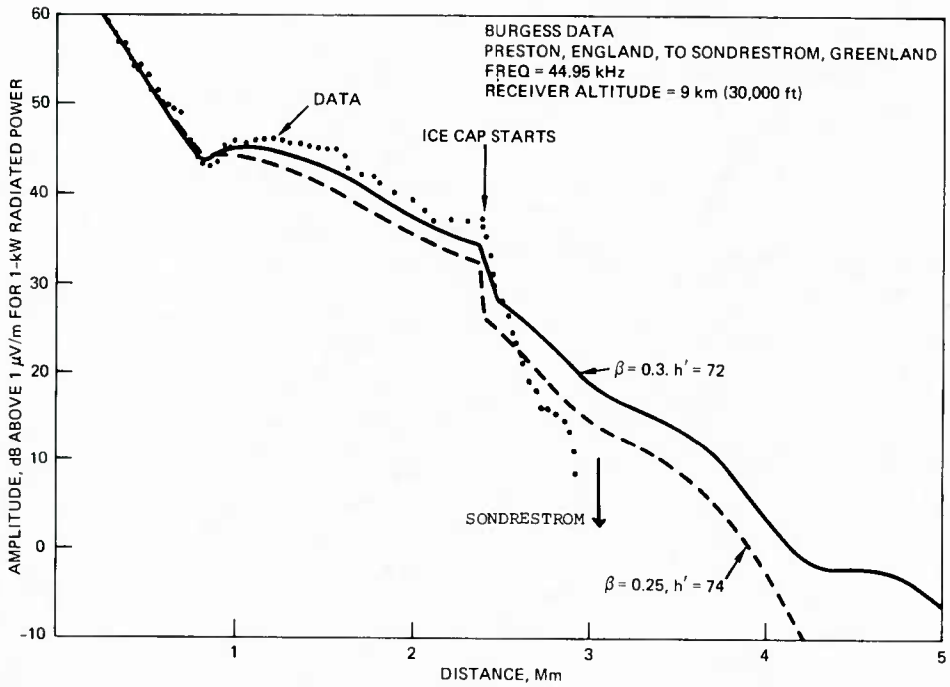


Fig. 46: High-Latitude Propagation

Flight Path: England to Sondrestrom, Greenland (daytime)

Ground conductivity: Sea water, 4 mhos/meter; Greenland ice cap, 10^{-5} mhos/meter

Transmitter: GYN Preston, England (45 kHz)

Date: July 1970

Electron Density Profiles: $\beta = 0.3\text{km}^{-1}$, $h' = 72$ km with the base of the ionosphere at 30 km and, $\beta = 0.25\text{km}^{-1}$, $h' = 74$ km with the base of the ionosphere at 40 km

Method of Computation: WKB approximation

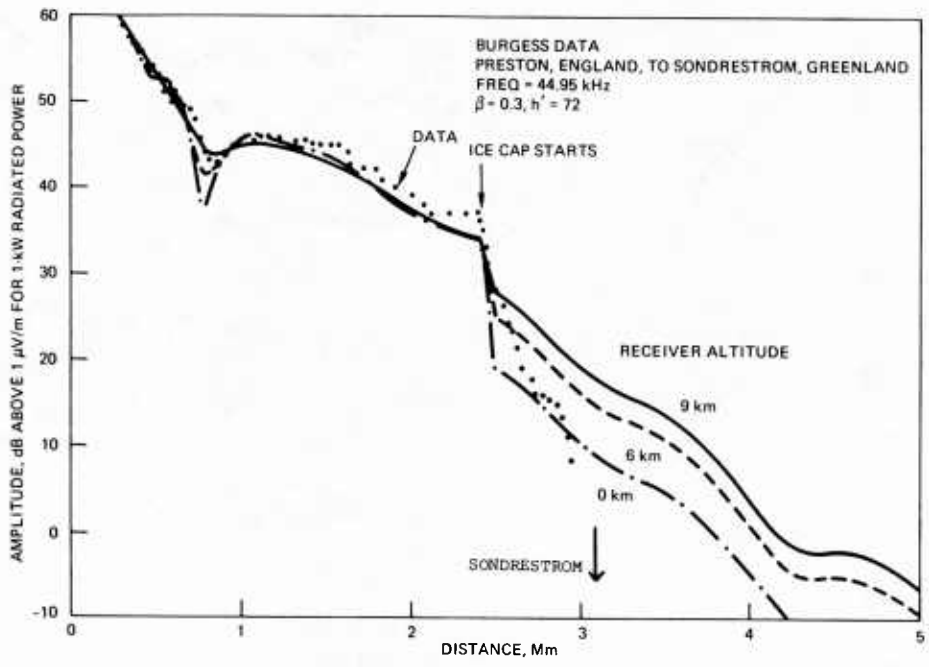


Fig. 47: High-Latitude Propagation

Propagation parameters are the same as Figure 46 except that only the $\beta = 0.3\ \text{km}^{-1}$, $h' = 72$ km profile is shown. Field strengths are computed for receiver altitudes (RALT) of 0, 6, and 9 km.

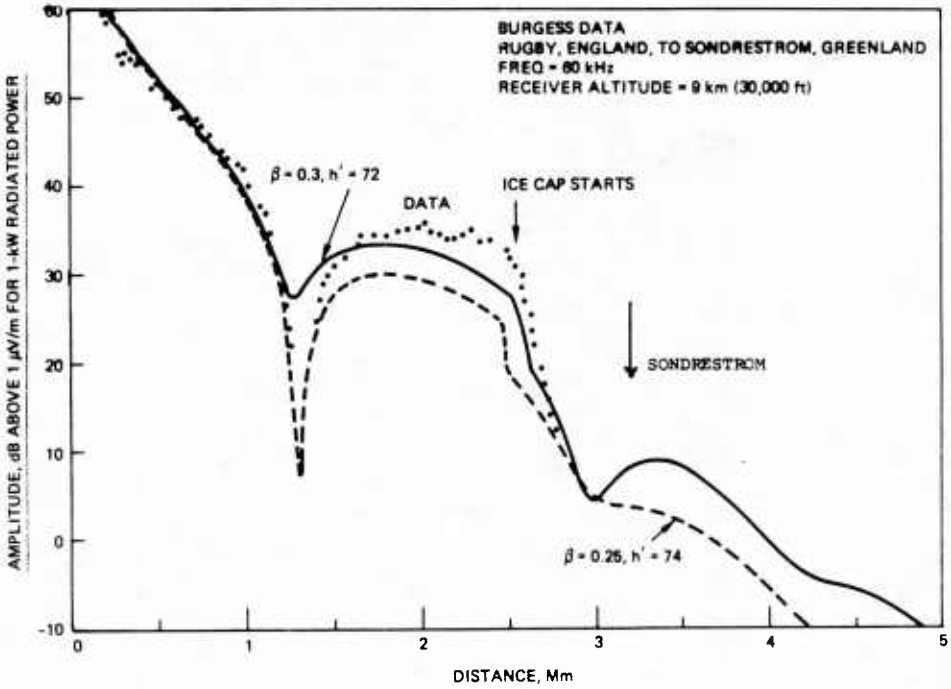


Fig. 48: High-Latitude Propagation
Flight Path: England to Sondrestrom, Greenland (daytime)
Ground conductivity: Sea water, 4 mhos/meter; Greenland ice cap, 10^{-5} mhos/meter
Transmitter: MSF Rugby, England (60 kHz)
Date: July 1970
Electron Density Profiles: $\beta = 0.3\text{km}^{-1}$, $h' = 72$ km with the base of the ionosphere at 30 km and, $\beta = 0.25\text{km}^{-1}$, $h' = 74$ km with the base of the ionosphere at 40 km
Method of Computation: WKB approximation

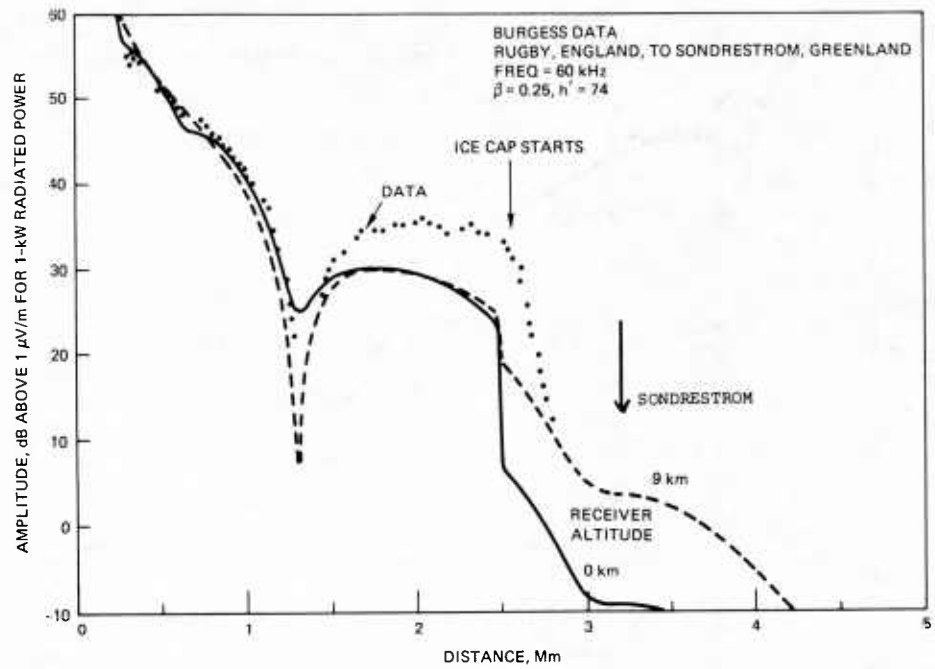
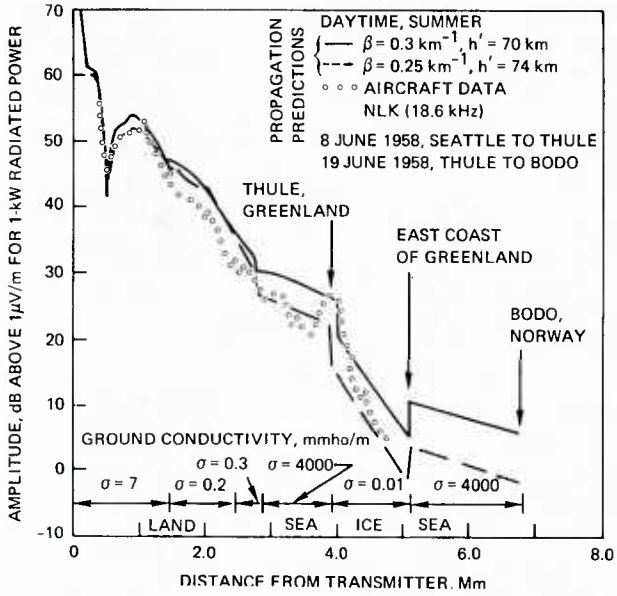
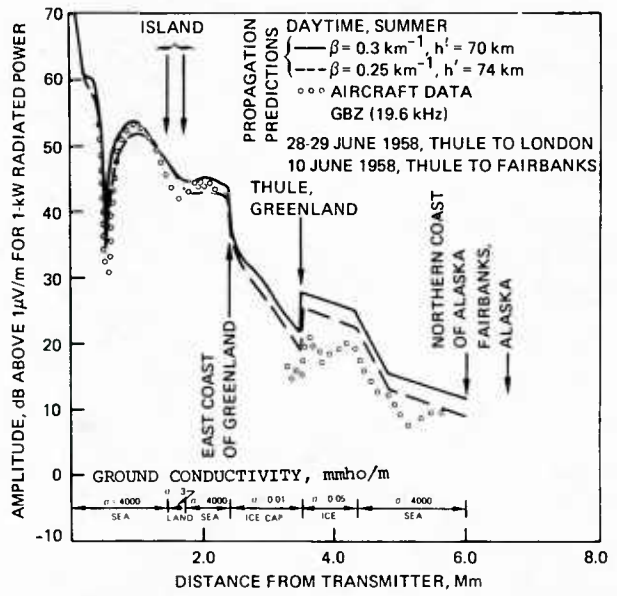


Fig. 49: High-Latitude Propagation
Propagation parameters are the same as Fig. 48 except that only the $\beta = 0.25$ km^{-1}, $h' = 74$ km profile is shown. Field strengths are computed for receiver altitudes (RALT) of 0 and 9 km.



Flight Path: Seattle, Washington across Canada and over the Greenland ice cap
Ground conductivity: Various values of land, sea, and ice along the path
Transmitter: NLK, Seattle, Washington (18.6 kHz)
Date: June 1958
Electron Density Profiles: $\beta = 0.3\text{km}^{-1}$, $h' = 72$ km with the base of the ionosphere at 30 km and, $\beta = 0.25\text{km}^{-1}$, $h' = 74$ km with the base of the ionosphere at 40 km
Method of Computation: WKB approximation

Fig. 50: High-Latitude Propagation



Flight Path: Thule, Greenland across the ice cap to London, England and and Thule to Fairbanks, Alaska
Ground conductivity: Various values of land, sea, and ice along the path
Transmitter: NLK, Seattle, Washington (18.6 kHz)
Date: June 1958
Electron Density Profiles: $\beta = 0.3\text{km}^{-1}$, $h' = 72$ km with the base of the ionosphere at 30 km and, $\beta = 0.25\text{km}^{-1}$, $h' = 74$ km with the base of the ionosphere at 40 km
Method of Computation: WKB approximation

Fig. 51: High-Latitude Propagation

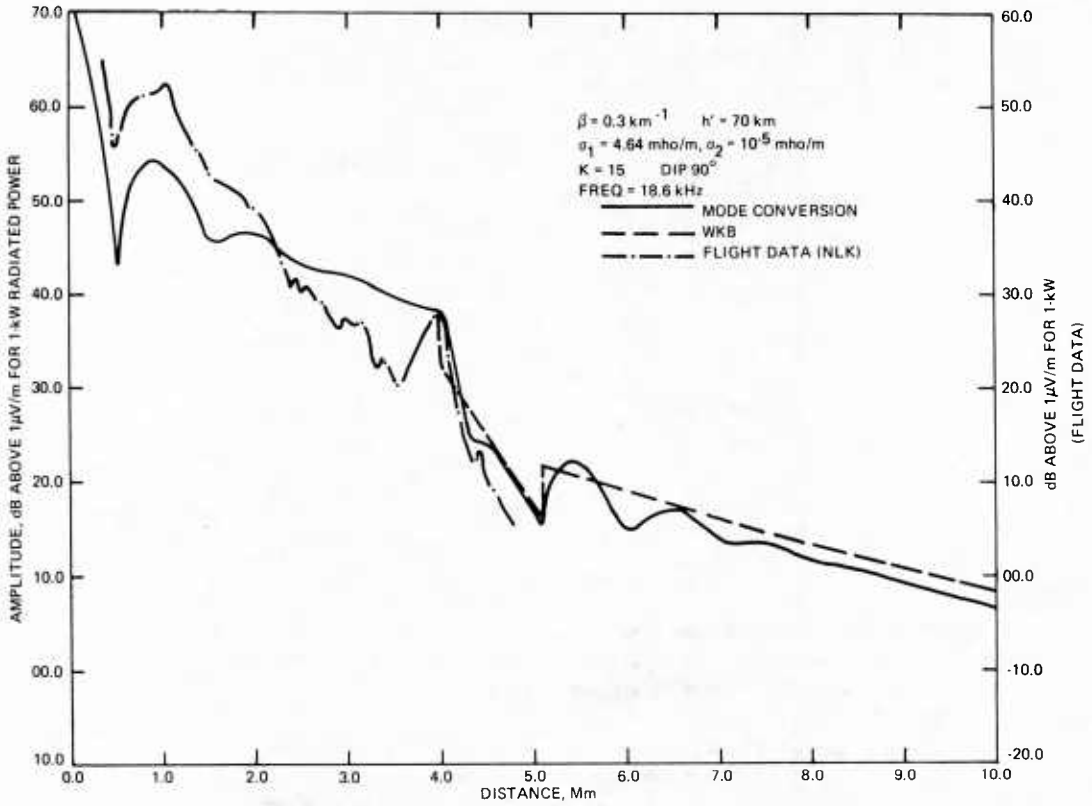


Fig. 52: High-Latitude Propagation

The propagation parameters are the same as those of Fig. 50 except that the earth's conductivities are taken as only sea water, σ = 4.64 mhos/meter and Greenland ice, $\sigma = 10^{-5}$ mhos/meter. The computations are made using the mode-conversion procedures.

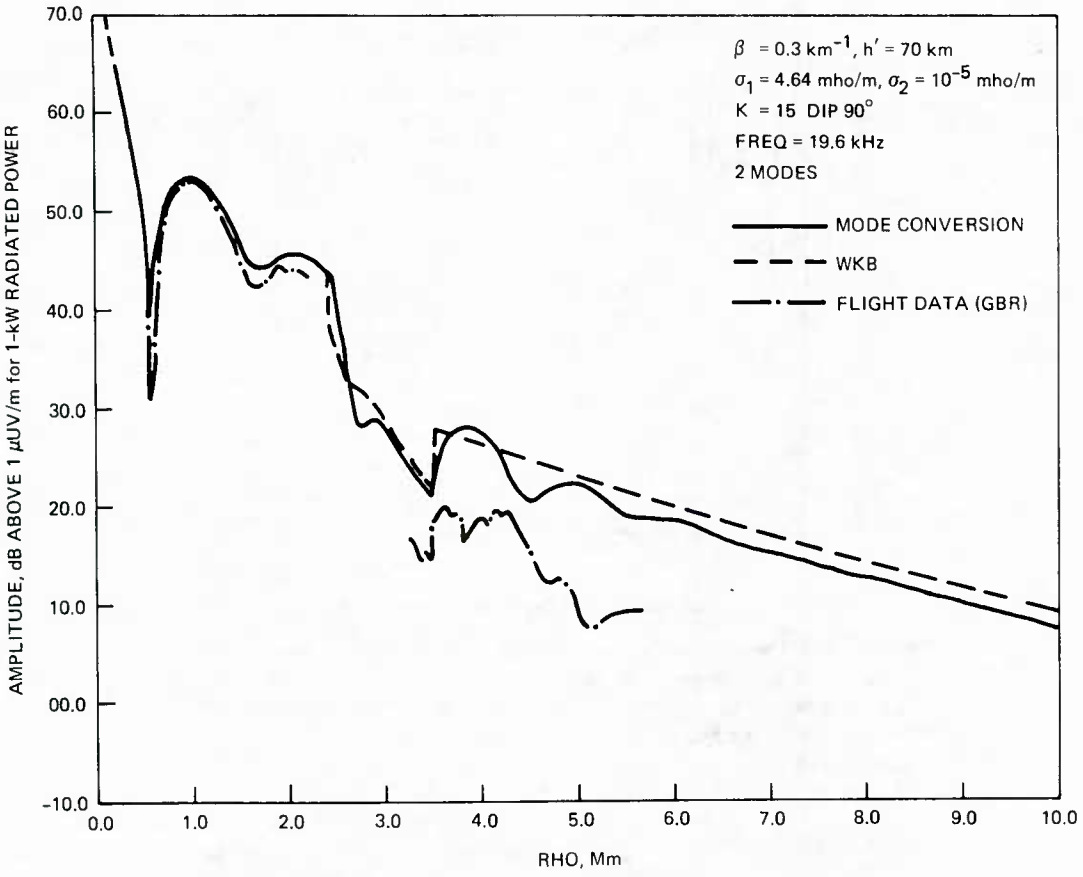


Fig. 53: High-Latitude Propagation

The propagation parameters are the same as those of Fig. 51 except that the earth's conductivities are taken as only sea water, (σ = 4.64 mhos/meter) and Greenland ice, ($\sigma = 10^{-5}$ mhos/meter). The computations are made using the mode-conversion procedures.

In this expression the sum of spherical harmonic Legendre polynomial P_q (cos Θ) functions, which utilize integral values of q, are re-expressed as an integral over complex contour in the variable ν using a Legendre function P_ν [cos (π − Θ)] which utilizes a complex value of the subscript ν. The function P_ν [cos (π − Θ)] behaves like a propagating and exponentially decaying wave in the Θ direction; i.e. along the surface of the earth. In the waveguide mode theory the integral in Eq. 31 is evaluated by the method of residues at residue points ν_i. Each value of ν_i is referred to as a waveguide mode root and is a solution of Eq. 32.

$$D_\nu = 0 = 1 - R_g^{(\nu)} R_i^{(\nu)} \frac{h_\nu^{(1)}(ka) h_\nu^{(2)}(kc)}{h_\nu^{(2)}(ka) h_\nu^{(1)}(kc)} \tag{32}$$

The wave-hop formulation for the propagation of long radio waves originated at a very early time after Marconi's original transatlantic transmissions. An early paper by Hollingworth (Ref. 42) is particularly noteworthy in that Hollingworth experimentally determined that the field strength versus distance of a long wavelength station truly behaves in an undulatory manner, thereby exhibiting the effect of groundwave and skywave interference at ranges governed by geometry and wavelength. The hop approach was adopted by most workers in the field during the thirties and forties—as exemplified by Bremmer's textbook (Ref. 43).

Wait (Refs. 40 and 44) opened up a new line for theoretical advance when he pointed out that the Watson integral expression for the fields (Eq. 31), which leads to the waveguide mode formulation, can also be expressed in terms of a geometric series in the reflection coefficients. The first term in the geometric series corresponds to the ground wave. The second term to a one-hop skywave. The third term describes a two-hop skywave, etc. This concept was extended by L. Berry, J.R. Johler, and others (Refs. 45-49) to produce a method for expressing the fields utilizing the anisotropic ionospheric reflection coefficients in a realistic way. These methods were developed during the 60's and culminated in computer programs produced by Berry, Herman, and Gonzales, and by Lewis in the early 70's. The wave hop method has been shown by Jones et al. (Ref. 50) and Morfitt and Halley (Ref. 51) to compare well with the waveguide mode methods of calculation.

The wavehop programs have been directly applied to making field strength prediction maps. The wavehop method possesses the great virtue that after inserting the parameters of a propagation path, the results are produced directly after a straightforward calculation that requires no human intervention to complete. Results are automatically mapped to provide signal to noise coverage charts (Figs. 54 through 57).

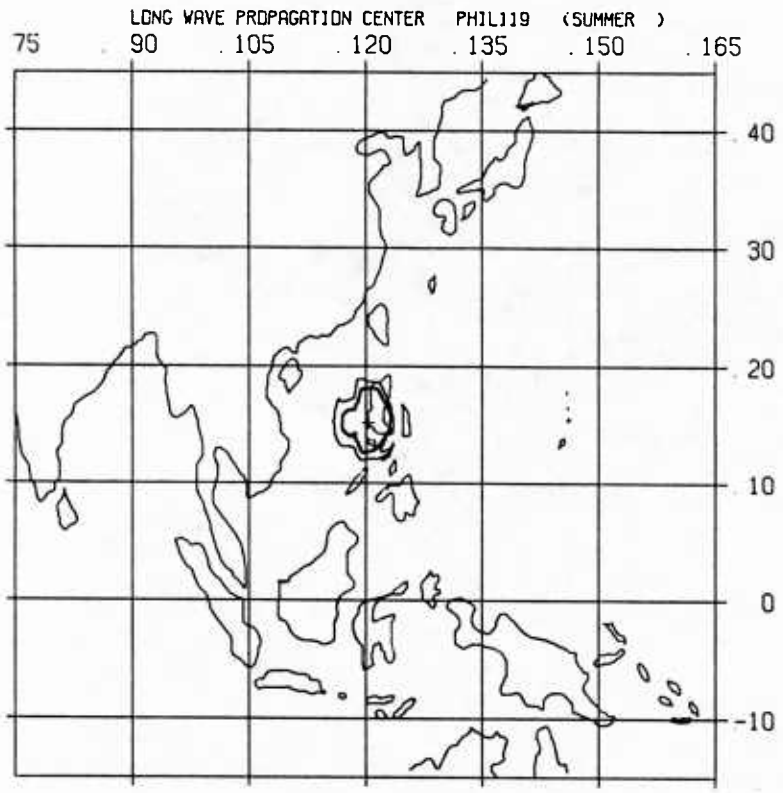


Fig. 54: Coverage chart showing signal to noise contours (3dB, heavy line; OdB light line) for a Naval Broadcast Station at Tarlac, Philippines on 119.85 kHz for 99% time availability for all hours of the month of July. Assumed transmitter power is 2.5 kW, and noise bandwidth is 1 kHz.

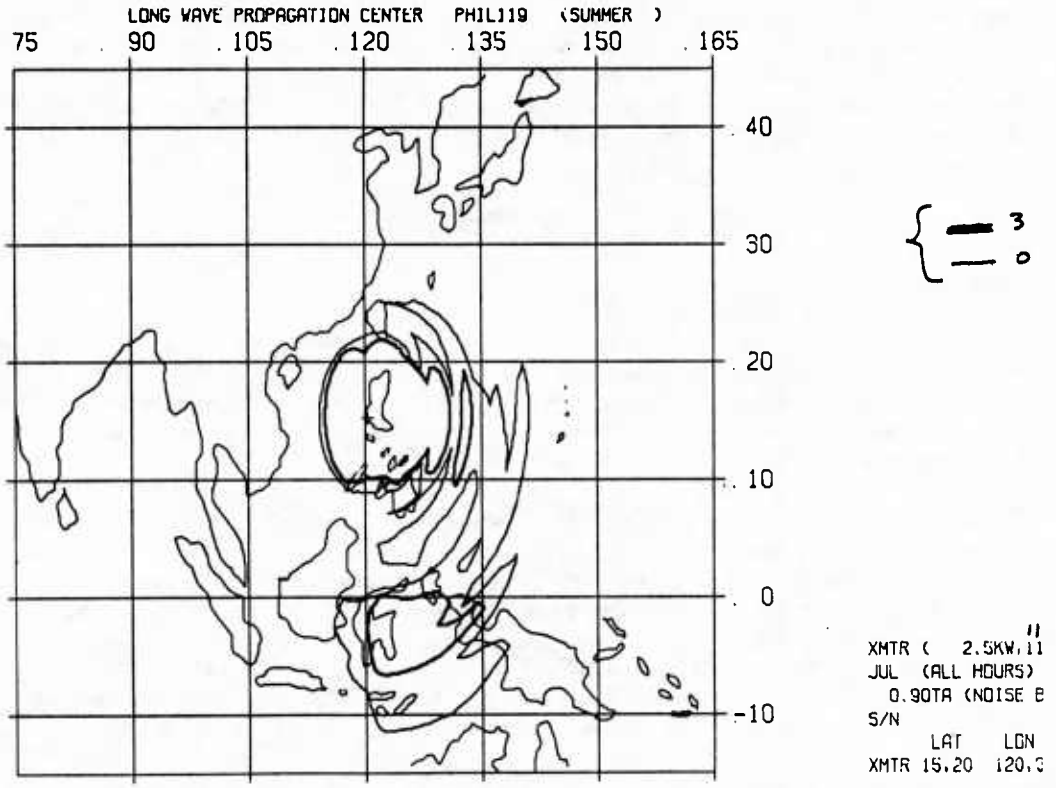


Fig. 55: Coverage chart for some conditions as previous figure except time availability is reduced to 90%

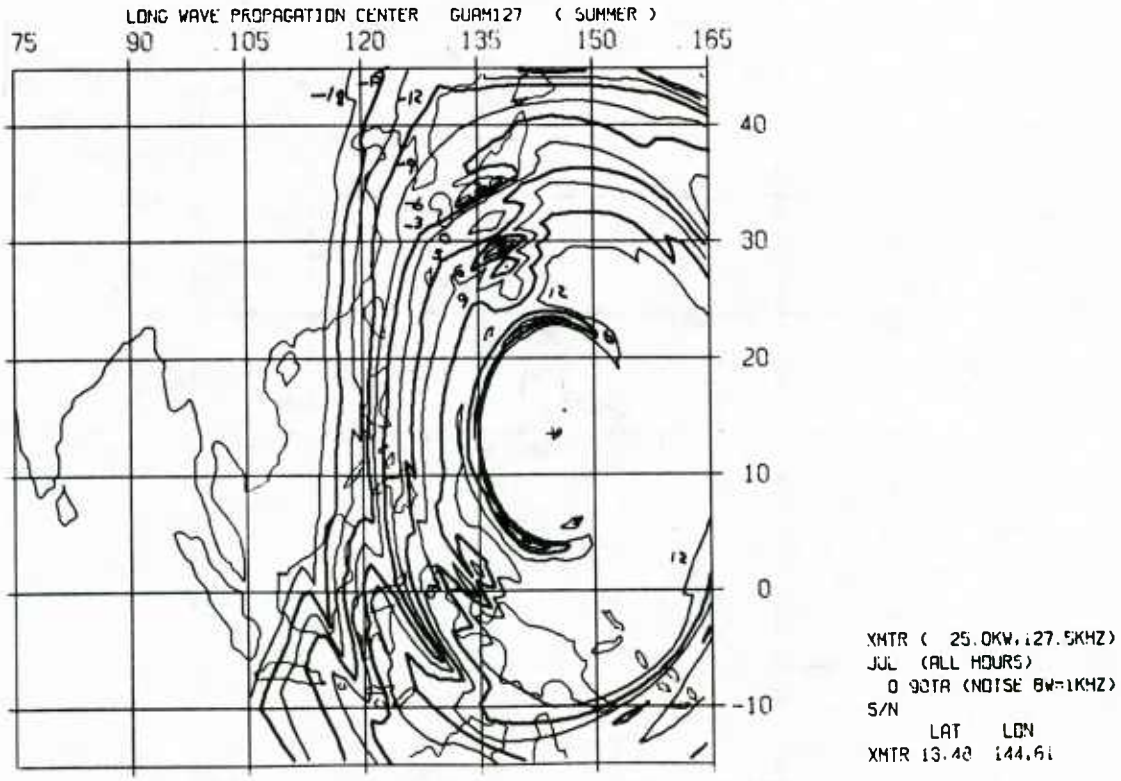


Fig. 56: Coverage chart showing signal to noise contours for a broadcast station at Guam on 127.5 kHz for 90% time availability for all hours of July. Assumed transmitter power is 23 kW, and noise bandwidth is 1 khz.

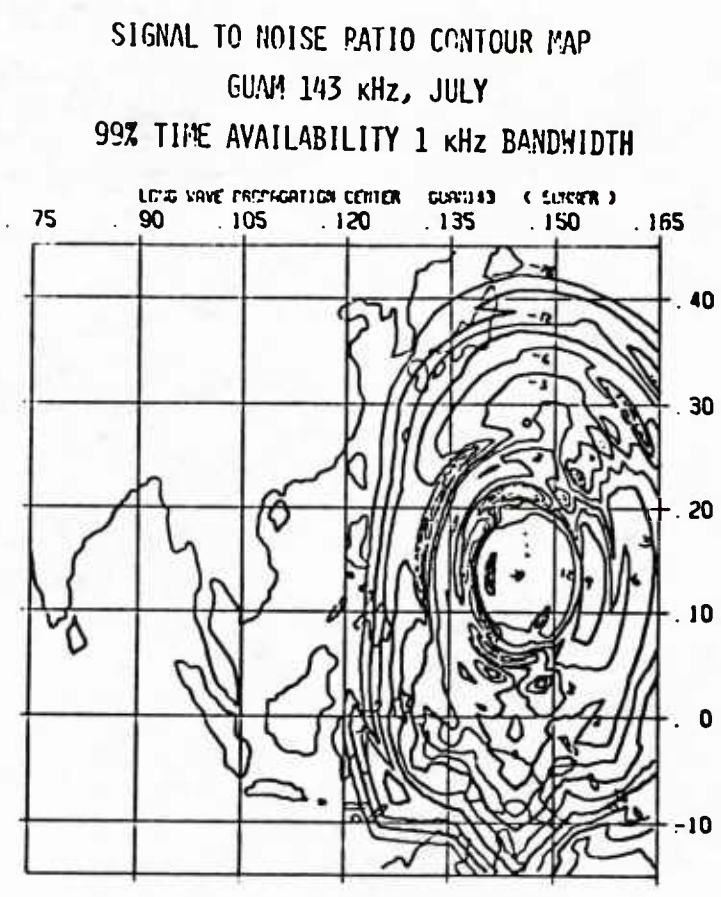


Fig. 57: Coverage chart for same conditions as previous figure except time availability is increased to 99%

3.4.1. Recent Developments in Wave Hop

Recently we have been implementing some of the initial ideas of Johler to make the wavehop program amenable to the treatment of inhomogeneous paths in a realistic way. To do this, one must calculate ionospheric reflection coefficients along the propagation path at a variety of angles of incidence and utilize these values in evaluating the fields via interpolation techniques. Thus, greater realism should be introduced into the programs at the cost of increased complexity and bookkeeping problems. Some calculations utilizing this technique are compared with the original Berry and Herman method in Figs. 58 and 59. We see that for long daytime paths the two programs give similar predictions, but for passage through the day-night terminator there are obvious differences.

Because both the multimode and multihop methods have complimentary advantages when applied to longwave propagation predictions, both methods should be applied to problems whenever either one is in doubt.

3.5 Future Needs for Coverage Predictions

Future progress in the accurate prediction of long wave radio fields will be strongly tied to our ability to monitor and model the lower ionospheric D-region and to accurately map ground conductivity of the earth as a function of frequency. A medium-sized campaign of aircraft flights and long term ground based receivers could reduce our uncertainties in many areas.

The possibility for making a world-wide D-region sensing and monitoring system is very appealing. Numerous systems can produce D-region ionospheric information, and special systems can be readily designed and implemented to serve this purpose. For example, the auroral photographs currently available from the DMSP satellite can be correlated with actual propagation calculations and measurements to get a model of the aurora-dominated situation. Data from DMSP type sensors working in various spectral bands could detect regions of high electron precipitation to provide real-time feedback to communications traffic managers to permit efficient mode, transmitter location and frequency selection. Numerous inexpensive ground systems can be implemented to provide auxiliary data. A prime example of a ground based D-region ionosphere sensor would be the large Arecibo Thomson Scatter Radar. It gives local electron density profile information from 60 km up to 1000 km. Performing a series of experiments with this facility tied in with OTH HF radar, partial reflection sensors, riometers, and microwave and airglow monitors should greatly enhance our knowledge of the D-region and of its properties relevant to long wavelength propagation prediction. A second series of experiments are the auroral zone using the University of Tromso's Partial Reflection Experiments and EISCAT Thompson Scatter Radar facility at Tromso, Norway, would give additional validity to the polar ionospheric models and help with the validation of a polar imaging system. The data

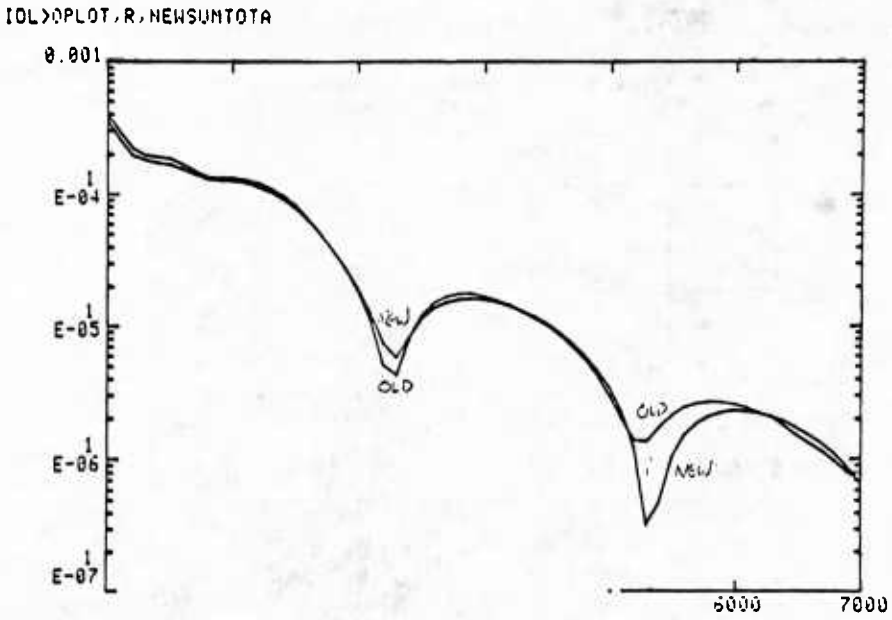


Fig. 58: Field strength vs distance predictions for Naval LF Broadcast Station Greece on 59 kHz for July, 3.5 kw radiated power, due north at 0000 UT. Solar zenith angle at mid-points (3500 km) is 62°. Path is quite homogeneous daytime, so old and new methods of prediction give similar results.

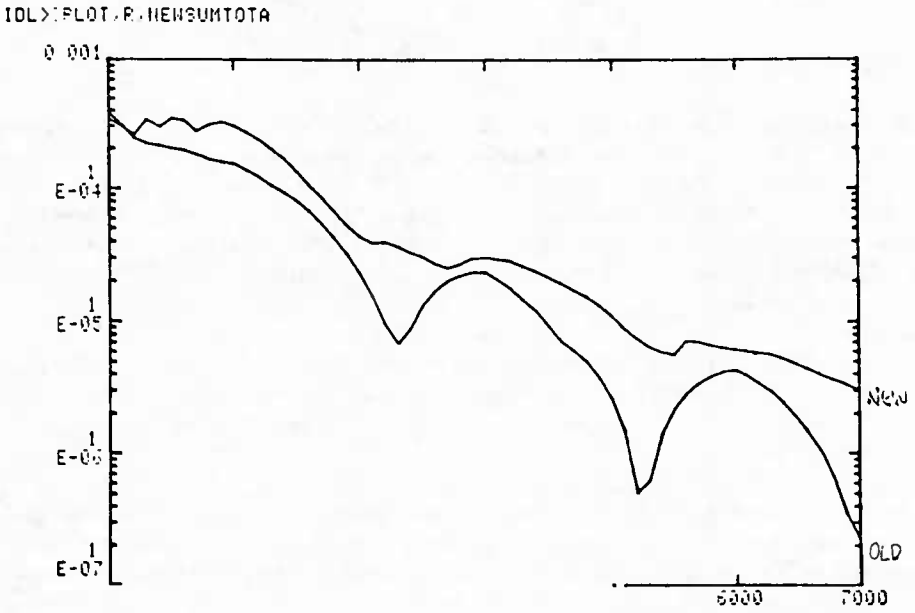


Fig. 59: Field strength vs distance predictions for Naval LF Broadcast Station Greece on 59 kHz for July, 3.6 kw radiated power, due north at 0000 UT. Solar zenith angle at mid-point (3500 km) is 87° or daytime condition; at 2500 km the solar zenith angle is 96° or nighttime condition. The old method interprets the path as all daylight; the new method has reduced attenuation for the nighttime portion of the path and appropriate daytime attenuation for the daytime portion.

from the HILAT satellite discussed by Huffman et al (Ref. 52) are another example of valuable satellite produced information.

3.6 Dispersion and Multipath

Because wide bandwidth or short pulse-length modulation can be used for VLF communication systems, it is important to consider how multimode propagation and dispersion will tend to distort short signal pulses. References 53 and 54 presented various methods for treating transient signals in a dispersive channel.

Reference 55 considered VLF pulse stretching from the point of view of the waveguide-mode theory of VLF propagation to show the predictions that this theory makes about short-pulse transmission.

When short pulses are propagated in a waveguide, the received pulse shape is altered by both multimode (or equivalently multipath) propagation and by dispersion of the pulses propagating by individual modes. Multimode distortion arises when a pulse leaves a transmitting antenna and travels down the waveguide in several modes. The different modes have different phase and group velocities, so the pulses will arrive at the receiver displaced from one another in time. Depending on how much displacement has occurred, the pulse can look like one long pulse or like a series of two or more pulses. Such multimode distortion is identical, in effect, to the classical multipath distortion that always arises when a signal can take more than one path to a receiver.

Dispersive distortion is independent of the multimode problem. A short pulse is composed of a spread of frequencies around a carrier frequency and thus occupies a certain bandwidth Δf, which is roughly given by the reciprocal of the length of the pulse τ_0; i.e., $\Delta f = 1/\tau_0$. Since the group velocities of the components on different frequencies are not the same, the components arrive at the receiver displaced in time. This leads to pulse smearing due to propagation even on a single mode.

It is important to remember that multimode distortion and dispersive distortion are distinct and independent phenomena. Thus, propagation channels that have a great deal of multimode distortion but little dispersion, and vice versa, could exist.

The equations that describe pulse stretching due to multimode propagation in the waveguide are straightforward. The index of refraction for the nth mode n_n is given by the ratio of the speed of light to the phase velocity (we use the letter n both for the index of refraction and as a mode index).

$$n_n = c/v_{ph,n} \tag{33}$$

the group velocity $v_{g,n}$ of a signal on the nth mode is obtained from the index of refraction and its derivative as follows:

$$v_{g,n}/c = (n_n + f\, dn_n/df)^{-1} \tag{34}$$

The time delay ΔT_{mn} between pulses arriving on different modes (the mth and nth modes) is simply given by:

$$\Delta T_{mn} = a\Theta \left(\frac{1}{v_{g,m}} - \frac{1}{v_{g,n}} \right) \tag{35}$$

Here $a\Theta$ $(=d)$ is the distance between the transmitter and receiver. The equations that describe the dispersion spread are also obtained very straightforwardly (Ref. 56). The difference in arrival times $\Delta\tau_{\text{disp}}$ between the components on the different frequencies on a single mode are calculated for a frequency difference Δf:

$$\Delta\tau_{\text{disp}} = a\,\Theta \left(\frac{1}{v_g(f)} - \frac{1}{v_g(f+\Delta f)} \right) \tag{36}$$

Collecting terms and introducing the initial pulse length $\tau_0 = 1/\Delta f$, we obtain

$$\Delta\tau_{\text{disp}} = \frac{a\,\Theta\,\Delta f}{v_g^2} \left(\frac{v_g(f+\Delta f) - v_g(f)}{\Delta f} \right) = \frac{a\Theta}{\tau_0 v_g^2} \frac{dv_g}{df} \tag{37}$$

The fractional dispersive spreading is given by

$$\frac{\Delta\tau_{\text{disp}}}{\tau_0} = \frac{a\Theta}{\tau_0^2 v_g^2} \frac{dv_g}{df} \tag{38}$$

There is a further type of dispersive distortion that cannot be treated by these simple concepts. This type of spreading occurs because of the frequency variation of the excitation factors and of the attenuation coefficients on the various modes. The amplitudes of different frequency components of a pulse will be excited differently and attenuated differently. A pulse

that was initially composed of frequencies over an initial effective bandwidth will appear to be shifted in frequency or to occupy a smaller bandwidth because of the greater absorption of the frequency components at one side or the other of the occupied band. This type of distortion could be very significant, but it appears that it can only be treated by considering various pulse shapes and performing Fourier transforms. The procedure for numerically studying this type of dispersion is well known. A pulse shape $v_t(t)$ at the transmitting antenna is given by some modulation function A(t) times a basic sine wave exp $i\omega_0 t$ on the frequency ω_0:

$$v_t(t) = A(t)\exp i\omega_0 t \tag{39}$$

This pulse can be Fourier analyzed:

$$v_t(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} A(t)\exp i\omega_0 t \exp -i\omega t \, dt \tag{40}$$

The received pulse $v_{r,n}(\omega)$ on the nth mode will be given by:

$$v_{r,n}(\omega) = v_t(\omega) \frac{e^{-ika\Theta}}{a(\Theta \sin\Theta)^{1/2}} V_n \tag{41}$$

where V_n is the nth term in the summation of (2); i.e.

$$V_n = \sin^{3/2}\theta_n \, \Lambda_n \exp(i\pi/4 + ik_o \, d\sin\theta_n) \tag{42}$$

The received pulse must then be transformed back to the time domain to obtain the pulse shape as a function of time:

$$v_{r,n}(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} e^{i\omega t} v_{r,n}(\omega) d\omega \tag{43}$$

If we desire the total received pulse shape, we merely sum the pulses on the various modes:

$$v_r(t) = \sum_{n=1}^{\infty} v_{r,n}(t) \tag{44}$$

We could just as well obtain $v_r(t)$ by performing the summation over the V_n before the integration over ω. We would then have obtained:

$$v_r(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} e^{i\omega t} \frac{v_t(\omega)e^{ika\Theta}}{a(\Theta\sin\Theta)^{1/2}} \left[\sum_{n=1}^{\infty} V_n\right] d\omega \tag{45}$$

The equivalence of (44) and (45) indicates that it makes no difference whether we consider the pulses as carried by different modes or whether we just consider the existence of the complicated complex transfer function:

$$H(\omega) = e^{-ika\Theta} \left[\sum_{n=1}^{\infty} V_n\right] / a(\Theta \sin\Theta)^{1/2} \tag{46}$$

It is easier to think of pulses traveling by different modes or hops, so we prefer to use this separate mode approach rather than the transfer function approach.

3.6.1 Dispersion Calculation

One may use the isotropic exponential ionosphere model with a $\beta = 0.5$ and heights of 70 and 90 km to represent daytime and nighttime ionospheric conditions. From data given in Ref. 23, one may calculate the index of refraction (Fig. 60) assuming an infinite ground conductivity. The derivative of the index of refraction can be numerically determined (Fig. 61); and thus the group velocity as a function of frequency has been obtained (Fig. 62). Using these group velocities, the delays between pulses on various modes were calculated. The multimode interference will normally have the greatest observable effect when two or more modes are approximately equal in strength. The region of frequency-range space in which two or more modes are within 4 db of each other was determined by studying the graphs of signal strength versus distance given by Ref. 24, which are based on the parameters given by Ref. 23. The group delays, the time of arrival between pulses on the first and second modes, and between pulses on the second and third modes were then computed. These delay times are shown in Tables 3 and 4. In these tables, two time delays are given for most frequencies. The upper time delay is between the first and second-mode pulses; the lower time delay is between the second- and third-mode pulses. For some frequencies, Ref. 23 gives the phase velocity for only two modes; and for these cases only one delay time is presented. Delay time values are given only for the region for which two or more modes have amplitudes within 4 db of each other. Delay time values at other ranges can be easily calculated by proportionality. In the regions given, two pulses would likely be separated

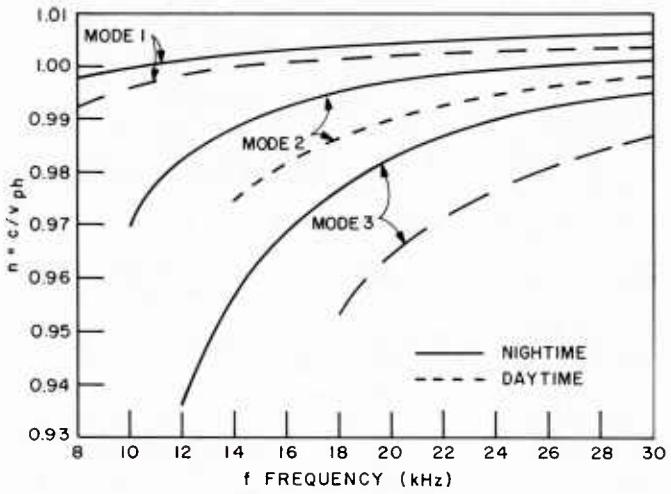


Fig. 60: Index of refraction vs frequency calculated by *Wait and Spies* [1964] for three modes ($\beta = 0.5$, $h = 70$ km, and 90 km). Solid line, nighttime; dashed line, daytime.

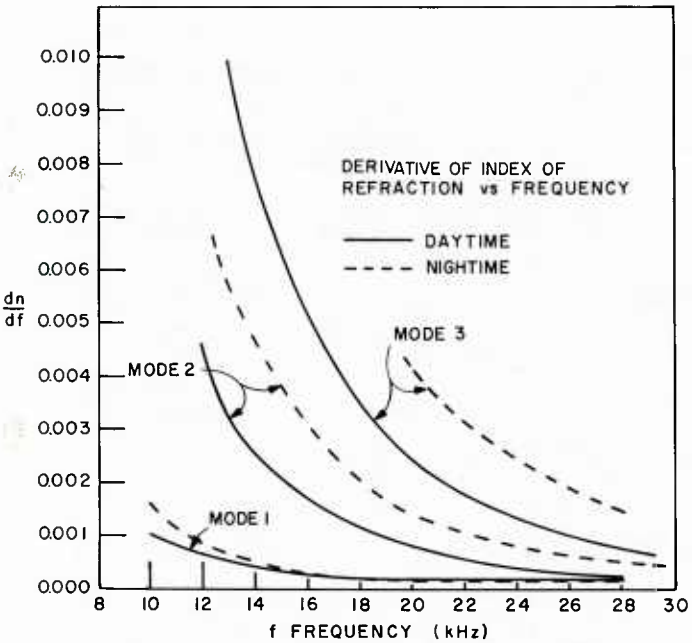


Fig. 61: Derivative of the index of refraction vs frequency. Solid line, daytime; dashed line, nighttime.

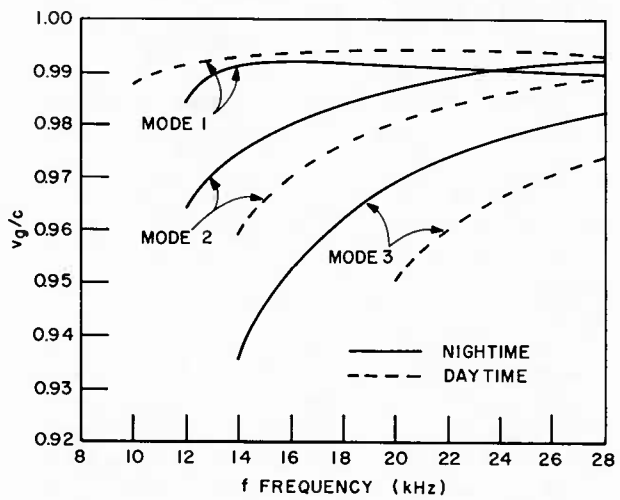


Fig. 62: Group velocity as a function of frequency. Solid line, nighttime; dashed line, daytime.

Table 3 — Daytime multipath time delays in milliseconds
(β = 0.5,h = 70km)

Frequency, kHz	Range, Mm					
	0.5	1.0	1.5	2.0	4.0	6.0
16	0.039					
18	0.031	0.061				
20	0.024	0.047	0.071			
	0.052	0.104	0.156			
22	0.018	0.036	0.054	0.073		
	0.042	0.084	0.126	0.168		
24	0.014		0.042	0.056		
	0.036		0.107	0.142		
26	0.009			0.038	0.076	
	0.031			0.123	0.246	
28	0.008				0.066	0.101
	0.025				0.202	0.302

Where two values of delay are given, the upper value is the delay between pulses on the first- and second-order modes, and the lower is the delay between pulses on the second- and third-order modes.

Table 4 — Nighttime multipath time delays in milliseconds
(β = 0.5, h = 90 km)

Frequency kHz	Range, Mm									
	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0	12.0	14.0
12	0.048	0.097								
14	0.030	0.061	0.091	0.121						
	0.065	0.130	0.195	0.261						
16	0.020		0.061	0.082	0.164					
	0.057		0.152	0.203	0.406					
18	0.0132				0.106	0.158				
	0.0383				0.312	0.467				
20	0.0076					0.091	0.121	0.151	0.182	
	0.0315					0.378	0.404	0.530	0.756	
22	0.0034									0.095
	0.0262									0.734
24	8×10^{-10}	1.6×10^{-9}								
	0.022	0.045								
26	2.5×10^{-4}	5×10^{-5}	7.5×10^{-5}							
	0.019	0.038	0.057							
28	4.1×10^{-5}	0.83×10^{-4}	1.24×10^{-4}	1.6×10^{-4}						
	0.017	0.034	0.051	0.068						

Where two values of delay are given, the upper value is the delay between pulses on the first- and second-order modes, and the lower is the delay between pulses on the second- and third-order modes.

by the time delay given on the upper line, or, if the two dominant pulses are arranged so as to cancel out to a great extent, the pulse remnant of the cancellation of the first two pulses would be followed by a pulse traveling on another mode.

The dispersive time delay fraction $\Delta\tau/\tau_0$ from (38) was also calculated for an initial pulse length of 2 msec received at 6000 km from the transmitter. The results are given in Figs. 63 and 64. We see that the spreading is small, typically only a few per cent.

3.6.2 Results of Dispersion Calculations

One observes that the time delays between the second- and third-order mode pulses are usually greater than the delay between the first- and second-order mode pulses. During the daytime the greatest delay between the first- and second-mode pulses in the interference region is approximately 0.1 msec at 28 kHz and 6000-km range. The greatest delay between the second- and third-mode pulse is approximately 0.3 msec at the same frequency and range. At nighttime the interference region is shifted to lower frequency. The greatest time delay between pulses on the first and second modes is approximately 0.18 msec at a frequency of 20 kHz and 12-Mm range. The time delay between the second- and third-mode pulses is approximately 0.76 msec at that frequency and range.

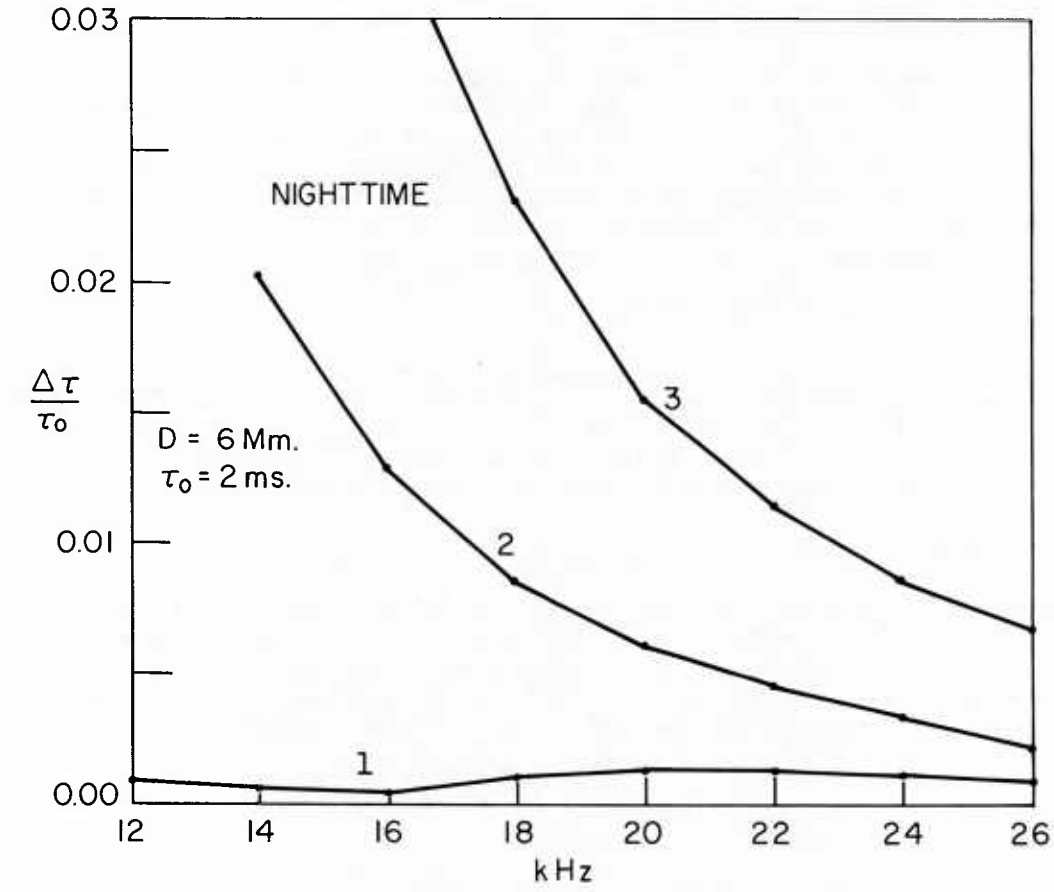


Fig. 63: Fractional time delay dispersion vs frequency for three modes for a distance $D = a\theta$ of 6 Mm for nighttime ($\beta = 0.5$, $h = 90$ km)

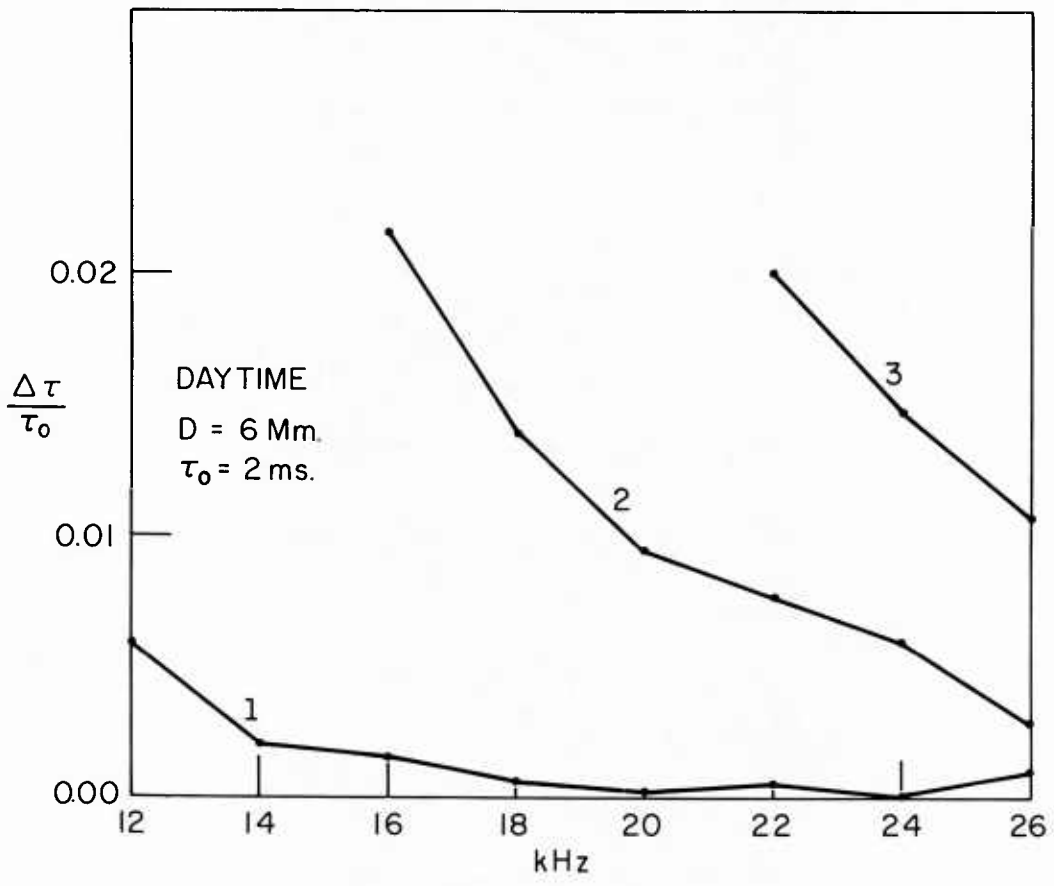


Fig. 64: Fractional time delay dispersion vs frequency for three modes for a distance $D = a\theta$ of 6 Mm for daytime ($\beta = 0.5$, $h = 70$ km)

3.6.3 Dispersion Conclusions

Theoretical estimates were obtained of the severity of multimode and dispersive distortion in the VLF propagation channel. It appears that multimode propagation will cause greater pulse distortion than will dispersion. Even at long distance at night, the multimode propagation introduces appreciable pulse distortion on some frequencies for pulse lengths shorter than ~1 or 2 msec, since the maximum delay time is of the order of 0.7 msec. During the daytime, however, the delays are smaller, only about 0.3 msec, and the pulse stretching would be smaller. From the present calculation, it would appear that a communicator could choose an optimum frequency for transmission to avoid the regions of frequency-range space given in Tables 3 and 4 and thereby avoid both problems of multimode reception time delays and of modal interference nulls.

4.0 MODULATION TECHNIQUES FOR LONGWAVE SYSTEMS

For most of its history, the longwave communication channel was dominated by on-off keyed (OOK) systems transmitting Morse code. Frequency-shift keyed (FSK) systems were introduced to provide higher speed data transfer. The performance of these systems are summarized in Ref. 2. In even more recent times minimum shift keying (MSK) systems have been introduced to provide still more efficient use of the limited frequency spans of each longwave communications channel. The following mathematical development is intended to provide an understanding of the newer MSK systems.

4.1 Elementary Signal Theory

Figure 65 shows schematically the outlines of a long wavelength communication system. Figure 66 shows a conceptual version of a receiver and demodulation portion of Fig. 65. The operation of the integrator portion of the equipment in Fig. 66 is to sample the received waveforms $h(t)$ and to multiply them by the desired waveform $h_1(t)$ and $h_2(t)$ and to integrate over a time equal to a bit-keying-element period T_B, centered about the time t_l the midpoint of the l th expected bit interval. The transmitted waveform is constructed as a linear superposition of $h_1(t)$ and $h_2(t)$ according to

$$h(t) = X_l\, h_1(t) + Y_l\, h_2(t). \tag{47}$$

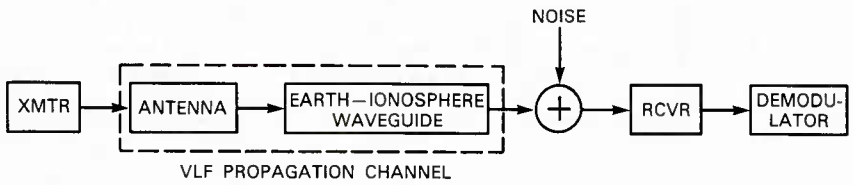


Fig. 65: Model of the communication system

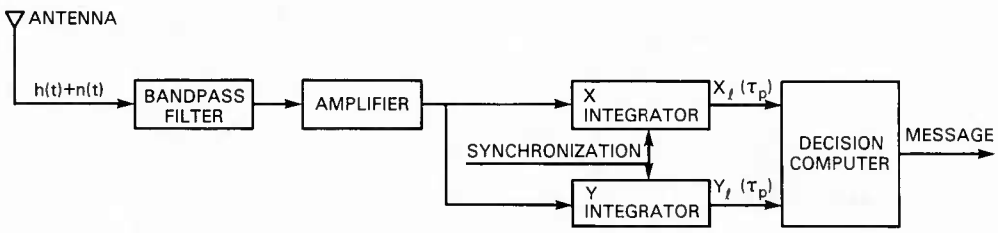


Fig. 66: Model of the receiver-demodulator

The outputs of the integrators will yield the values of X_l and Y_l which contain the transmitted information,

$$X_l = (T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_1(t) h(t)\, dt \tag{48}$$

and

$$Y_l = (T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_2(t) h(t)\, dt, \tag{49}$$

as long as $h_1(t)$ and $h_2(t)$ are orthonormal to each other:

$$(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_i(t)\, h_j(t)\, dt = \delta_{ij}. \tag{50}$$

Here δ_{ij} is the Kronecker delta function, equal to zero for $i \neq j$ and equal to 1 for $i = j$.

4.1.1 Correlation of Orthogonal Waveforms

A set of orthogonal wideband waveforms $h_1(t)$ and $h_2(t)$ can be generated by using a baseband waveform $h_B(t)$ to modulate both a sinusoidal and cosinusoidal wave:

$$h_1(t) = h_B(t) \cos \omega_0 t, \tag{51}$$

$$h_2(t) = h_B(t) \sin \omega_0 t, \tag{52}$$

$$h(t) = h_B(t)(X_l \cos \omega_0 t + Y_l \sin \omega_0 t). \tag{53}$$

The values of X_l and Y_l give the strength of the cosinusoidal and sinusoidal components. A correlation receiver can then be made to extract the value of X_l and Y_l:

$$X_l = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_B(t)(X_l \cos \omega_0 t + Y_l \sin \omega_0 t)\, h_B(t) \cos \omega_0 t \, dt \tag{54}$$

and

$$Y_l = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_B(t)\, (X_l \cos \omega_0 t + Y_l \sin \omega_0 t)\, h_B(t) \sin \omega_0 t \, dt. \tag{55}$$

A propagation delay time of τ_p modifies the correlator outputs to give new ones which we call $X_l(\tau_p)$ and $Y_l(\tau_p)$:

$$X_l(\tau_p) = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_B(t + \tau_p)\, [X_l \cos \omega_0(t + \tau_p) + Y_l \sin \omega_0(t + \tau_p)] \times h_B(t) \cos \omega_0 t \, dt \tag{56}$$

and

$$Y_l(\tau_p) = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + t_B/2} h_B(t + \tau_p)([X_l \cos \omega_0(t + \tau_p) + Y_l \sin \omega_0(t + \tau_p)] \times h_B(t) \sin \omega_0 t \, dt. \tag{57}$$

We find that the vector $\vec{R}_l(\tau_p) = (X_l(\tau_p), Y_l(\tau_p))$ is approximately the phase-rotated version of $\vec{R}_l = (X_l, Y_l)$. We have

$$X_l(\tau_p) = T_B^{-1}(X_l \cos \omega_0 \tau_p + Y_l \sin \omega_0 \tau_p) \int_{t_l - T_B/2}^{t_l + T_B/2} h_B(t)\, h_B(t + \tau_p)\, dt \tag{58}$$

and

$$Y_l(\tau_p) = T_B^{-1}(-X_l \sin \omega_0 \tau_p + Y_l \cos \omega_0 \tau_p) \int_{t_l - T_B/2}^{t_l + t_B/2} h_B(t) h_B(t + \tau_p)\, dt. \tag{59}$$

Thus we can write this equation in vector form if $\vec{R}_l = (X_l, Y_l)$:

$$\vec{R}_l(\tau_p) = T_B^{-1} \left[\int_{t_l - T_B/2}^{t_l + T_B/2} h_B(t) h_B(t + \tau_p) dt \right] \begin{bmatrix} \cos \omega_0 \tau_p & \sin \omega_0 \tau_p \\ -\sin \omega_0 \tau_p & \cos \omega_0 \tau_p \end{bmatrix} \vec{R}_l. \tag{60}$$

The magnitude of the vector $\vec{R}_l(\tau_p) = [X_l(\tau_p)^2 + Y_l(\tau_p)^2]^{1/2}$ is given by

$$|\vec{R}_l(\tau_p)| = T_B^{-1} |\vec{R}_l| \int_{t_l - T_B/2}^{t_l + T_B/2} h_B(t) h_B(t + \tau_p) dt. \tag{61}$$

The phase of the vector $\vec{R}_l(\tau_p)$ is given by

$$\phi_l(\tau_p) = arg(X_l(\tau_p) + iY_l(\tau_p)) = arg(X_l + iY_l) + \omega_0 \tau_p. \tag{62}$$

Thus, the magnitude of the $\vec{R}_l(\tau_p)$ vector varies slowly with τ_p, and the phase $\phi_l(\tau_p)$ changes rapidly.

4.1.2 Correlation of Other Orthogonal Waveforms

A more complicated set of orthogonal waveforms can be generated using two orthogonal baseband waveforms $h_{B1}(t)$ and $h_{B2}(t)$ such that for $i, j = 1, 2$

$$(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{Bi}(t) h_{Bj}(t)\, dt = \delta_{ij}, \tag{63}$$

where δ_{ij} is the Kronecker delta function. The transmitted waveform $h(t)$ is given by

$$h(t) = X_l h_{B1}(t) \cos \omega_0 t + Y_l h_{B2}(t) \sin \omega_0 t. \tag{64}$$

To extract the value of X_l and Y_l from this waveform, we need a correlator receiver to do the following integrations:

$$X_l = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} [X_l h_{B1}(t) \cos \omega_0 t + Y_l h_{B2}(t) \sin \omega_0 t][h_{B1}(t) + h_{B2}(t)] \cos \omega_0 t \, dt \tag{65}$$

and

$$Y_l = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} [X_l h_{B1}(t) \cos \omega_0 t + Y_l h_{B2}(t) \sin \omega_0 t][h_{B1}(t) + h_{B2}(t)] \sin \omega_0 t \, dt. \tag{66}$$

When the time of arrival is changed to $t = t + \tau_p$, we have

$$X_l(\tau_p) = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} [X_l h_{B1}(t + \tau_p) \cos \omega_0(t + \tau_p) + Y_l h_{B2}(t + \tau_p) \sin \omega_0(t + \tau_p)] \times [h_{B1}(t) + h_{B2}(t)] \cos \omega_0 t \, dt \tag{67}$$

and

$$Y_l(\tau_p) = 2(T_B)^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} [X_l h_{B1}(t + \tau_p) \cos \omega_0(t + \tau_p) + Y_l h_{B2}(t + \tau_p) \sin \omega_0(t + \tau_p)] \times [h_{B1}(t) + h_{B2}(t)] \sin \omega_0 t \, dt. \tag{68}$$

This gives

$$X_l(\tau_p) = X_l \cos \omega_0 \tau_p T_B^{-1} \int h_{B1}(t + \tau_p)[h_{B1}(t) + h_{B2}(t)] \, dt + Y_l \sin \omega_0 \tau_p T_B^{-1} \int h_{B2}(t + \tau_p)[h_{B1}(t) + h_{B2}(t)] \, dt \tag{69}$$

and

$$Y_l(\tau_p) = -X_l \sin \omega_0 \tau_p T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B1}(t + \tau_p)[h_{B1}(t) + h_{B2}(t)] \, dt + Y_l \cos \omega_0 \tau_p T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B2}(t + \tau_p)[h_{B1}(t) + h_{B2}(t)] \, dt. \tag{70}$$

If the two baseband modulation functions are sufficiently orthogonal that

$$T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B1}(t + \tau_p) h_{B2}(t) \, dt \approx 0 \tag{71}$$

and sufficiently similar that

$$T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B1}(t + \tau_p) h_{B1}(t) \, dt = T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B2}(t + \tau_p) \, h_{B2}(t) \, dt, \tag{72}$$

we again obtain

$$X_l(\tau_p) = (X_l \cos \omega_0 \tau_p + Y_l \sin \omega_0 \tau_p) T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B1}(t + \tau_p) h_{B1}(t) \, dt \tag{73}$$

and

$$Y_l(\tau_p) = (-X_l \sin \omega_0 \tau_p + Y_l \cos \omega_0 \tau_p) T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B1}(t + \tau_p) h_{B1}(t) \, dt. \tag{74}$$

When the more complicated correlators of Eqs. (65) and (66) are used, and the orthogonal waveforms satisfy the conditions of Eqs. (71) and (72), the type of waveform of Eq. (64) gives a behavior similar to that of Eq. (53).

4.2 Modulation Methods

The values of $\vec{R}_l$ can be modified at the transmitter to initiate the communication of information. In amplitude modulation the magnitude of $\vec{R}_l$ is varied. In phase-shift keying the phase angle ϕ_l is modulated. Thus, in general, information may be transmitted by changing the transmitted vector $\vec{R}_l$ in some prescribed way. The stability of the received vector $\vec{R}_l(\tau_p)$ in amplitude and phase is important in the performance of communication systems. In the realistic communication channel the values of $\vec{R}_l(\tau_p)$ will also depend on the noise present. In this study we neglect noise and simply study the effects of propagation on $\vec{R}_l(\tau_p)$.

4.3 MSK Signals

In the field of VLF communications, minimum shift keying (MSK) is a commonly used type of spread-spectrum modulation. MSK is mathematically described in terms of a transmitted waveform $s(t)$, a baseband modulation $u(t)$, and a chip waveform $p(t)$, as follows:

$$s(t) = \mathrm{Re}\left[u(t) e^{i\omega_0 t}\right], \tag{75}$$

$$u(t) = \sum_{n=0}^{\frac{N}{2}-1} \left[X_l I_n \, p\,(t - nT) - i\, Y_l \, Q_n \, p\,(t - T/2 - nT)\right], \tag{76}$$

$$p(t) = \begin{cases} \sqrt{\frac{2}{T}} \left[\cos\left(\frac{\pi t}{T}\right)\right] \\ 0, \quad |t| > T/2. \end{cases}, \quad -T/2 \le t \le T/2, \tag{77}$$

This is a specific version of the signal described in general by Eq. (64), in which the orthogonal baseband waveforms $h_{B1}(t)$ and $h_{B2}(t)$ are given by

$$h_{B1}(t) = \sqrt{T} \sum_{n=0}^{\frac{N}{2}-1} I_n p(t - nT) \tag{78}$$

and

$$h_{B2}(t) = \sqrt{T} \sum_{n=0}^{\frac{N}{2}-1} Q_n p(t - T/2 - nT). \tag{79}$$

In these above equations I_n and Q_n are pseudorandomly chosen integers of value +1 or -1. It is assumed that both the transmitter and receiver possess synchronized pseudorandom-number-sequence generators that provide the I_n and Q_n sequences. The frequency ω_0 is the carrier frequency of the broadcast. The time interval T is the duration of a chip. Since two chips are transmitted during each chip duration, the chip length τ_c is $T/2$. The MSK modulation method gives rise to a constant-amplitude waveform with continuous phase as long as X_l and Y_l are equal to ± 1. The bandwidth of the transmitted signal is inversely related to the chip period. The value of the waveform $s(t)$ depends on one value of I_n and one value of Q_n simultaneously except at the instant when a transition is being made from one value of I_n or Q_n to the next. At the transition time the amplitude of the waveform undergoing transition is zero, because the $p(t)$ function goes smoothly to zero at the beginning and end of its range. At that instant the value of $s(t)$ depends only on the other integer Q_n or I_n. Let us consider the autocorrelation function $B(\tau)$ for the MSK waveform $s(t)$:

$$B(\tau) = \lim_{T' \to \infty} \frac{1}{T'} \int_0^{T'} s(t)s(t + \tau)\, dt. \tag{80}$$

Use of $s(t)$ from Eq. (75) will make $B(\tau)$ nonzero as long as $|\tau| < T$. When $|\tau| > T$, the contributions to $B(\tau)$ will all be from terms having the products $Q_n I_m$, $Q_n Q_m$, and $I_n I_m$ with $n \neq m$. These terms are as often +1 as they are -1 for an orthogonal or uncorrelated set of $\{Q_n\}$ and $\{I_n\}$; so a summation based on these values should remain small. Thus a large value for the autocorrelation function occurs only for $|\tau| < T$, when I_n^2 and Q_n^2 terms contribute. Since $I_n^2 = Q_n^2 = 1$, these terms are easy to evaluate. The correlation vector $\vec{R}_l(\tau_p)$ is closely related to the autocorrelation function $B(\tau)$. At very low frequencies the electromagnetic waves may be considered to propagate by distinct waveguide modes or hops. The time delays between dominant modes and hops are small—less than a millisecond (Ref. 55)—producing the coherent overlap of one chip with its time-delayed image. At some receiver locations the interference of the two waves with each other results in "selective" fading. Such fading is called "selective" because it occurs only for frequencies for which the interference condition is satisfied. At other nearby receiver locations or frequencies the time delays will be such as to produce two pulses which reenforce each other. The propagation channel can be simulated by a multipath model in which $h_T(t)$ is the transmitted wave and the received wave is

$$h_R(t) = \sum_{i=1}^{N} A_i\, h_T\, (t - \tau_i), \tag{81}$$

where τ_i is the propagation time for the ith mode or hop and A_i is the amplitude for the ith mode or hop. In the case of earth-ionosphere waveguide propagation, sometimes the waves have distinctly different group and phase velocities. Then, we must consider both the phase and group delay times $\tau_p^{(i)}$ and $\tau_g^{(i)}$. The phase delay of the baseband modulation is governed by $\tau_p^{(i)}$, and the group delay is governed by $\tau_g^{(i)}$.

4.4 The MSK Correlation Receiver

One can readily evaluate the performance of a correlation receiver for MSK modulation and coherent detection for signal reception on a *single-path* propagation channel. The correlation vector $\vec{R}_l(\tau_p)$ obtained when one uses MSK modulation is explicitly calculated by substituting the expression for $h_{B1}(t)$ from Eq. (78) into Eqs. (73) and (74). The integral $I_B(\tau_p)$, defined as

$$\begin{aligned} I_B(\tau_p) &= T_B^{-1} \int_{t_l - T_B/2}^{t_l + T_B/2} h_{B1}(t + \tau_p) h_{B1}(t)\, dt \\ &= T_B^{-1} \sum_{\substack{n=0 \\ n'=0}}^{\frac{N}{2}-1} I_n I_{n'} \int_{t_l - T_B/2}^{t_l + T_B/2} p(t - nT) p(t + \tau_p - n'T)\, dt, \end{aligned} \tag{82}$$

is readily evaluated. The limits of integration in this equation extend far beyond the region of variable t, for which the integrand is nonzero. We can change the integration variable to

$$t' = t - nT. \tag{83}$$

The limits of integration are thereby changed to $t_l + T_B/2 - nT$ and $t_l - T_B/2 - nT$:

$$I_B(\tau_p) = T_B^{-1} \sum_{\substack{n=0 \\ n'=0}}^{\frac{N}{2}-1} I_n I_{n'} \int_{t_l - T_B/2 - nT}^{t_l + T_B/2 - nT} p(t') p(t' + nT + \tau_p - n'T)\, dt. \tag{84}$$

Since $p(t') = 0$, for $|t'| > T/2$, we need integrate only between $-T/2 \leq t \leq T/2$. We obtain

$$I_B(\tau_p) = T_B^{-1} \sum_{\substack{n=0 \\ n'=0}}^{\frac{N}{2}-1} I_n I_{n'} \int_{-T/2}^{+T/2} p(t')p[t' + \tau_p + (n - n')T]\, dt. \tag{85}$$

From this formulation we see that the integral depends only on the difference between n and n' and not on the individual values of n and n' separately. Let us define

$$C_k(\tau_p) = \int_{-T/2}^{T/2} p(t')p(t' + \tau_p + kT)\, dt. \tag{86}$$

Since $k = n - n'$; Eq. (85) becomes

$$\begin{aligned} I_B(\tau_p) &= T_B^{-1} \sum_{\substack{n=0 \\ n'=0}} I_n I_{n'} C_{n-n'}(\tau_p) \\ &= T_B^{-1} \sum_{n=0}^{\frac{N}{2}-1} \sum_{k=n}^{n+1-\frac{N}{2}} I_n I_{n-k} C_k(\tau_p) \\ &= T_B^{-1} \sum_{n=0}^{\frac{N}{2}-1} \sum_{k=-\left(\frac{N}{2}-1\right)}^{\frac{N}{2}-1} I_n I_{n-k} C_k(\tau_p) U_{n-k}, \end{aligned} \tag{87}$$

where

$$\begin{aligned} U_{n-k} &= 1, \text{ if } 0 \leq n - k \leq \frac{N}{2} - 1, \\ &= 0, \text{ if } n - k < 0 \text{ or } n - k > \frac{N}{2} - 1. \end{aligned}$$

We then interchange the order of summation in Eq. (87) to get

$$I_B(\tau_p) = T_B^{-1} \sum_{k=-\left[\frac{N}{2}-1\right]}^{\frac{N}{2}-1} C_k(\tau_p) \sum_{n=0}^{\frac{N}{2}-1} I_n I_{n-k} U_{n-k}. \tag{88}$$

Because of the orthogonality of the set $\{I_n\}$ with any sequentially shifted copy of itself, we can write

$$\sum_{n=0}^{\frac{N}{2}-1} I_n I_{n-k} U_{n-k} = \frac{N}{2} \delta_{k,0}. \tag{89}$$

This assumes that the binary sequence $\{I_n\}$ has been generated by a method which has "good" autocorrelation properties for the pseudo-noise-coding application. We obtain

$$I_B(\tau_p) = \frac{NT_B^{-1}}{2} C_0(\tau_p). \tag{90}$$

To calculate $C_0(\tau_p)$, we substitute into Eq. (86)

$$C_0(\tau_p) = \int_{-T/2}^{T/2} p(t')p(t' + \tau_p)\, dt. \tag{91}$$

The value of $C_0(\tau_p)$ is symmetric with respect to τ_p and can be obtained by using elementary integration methods:

$$\begin{aligned} C_0(\tau_p) &= \frac{4}{T} \int_{-\frac{T}{2}}^{T/2-\tau_p} p(t')p(t' + \tau_p)\, dt, \qquad \text{if } \tau_p \geq 0, \\ &= \frac{4}{T} \int_{-\frac{T}{2}}^{T/2-\tau_p} \cos\left(\frac{\pi t'}{T}\right) \cos\left[\pi\left(\frac{t' + \tau_p}{T}\right)\right] dt' \\ &= \frac{4}{T}\left[\left(1 - \frac{|\tau_p|}{T}\right)\cos\frac{\pi\tau_p}{T} + \frac{1}{\pi}\sin\frac{|\tau_p|}{T}\right]. \end{aligned} \tag{92}$$

Equations. (73) and (74) become

$$X_l(\tau_p) = (X_l \cos\omega_0\tau_p + Y_l \sin\omega_0\tau_p) NT_B^{-1} C_0(\tau_p) \tag{93a}$$

and

$$Y_l(\tau_p) = (-X_l \sin\omega_0\tau_p + Y_l \cos\omega_0\tau_p) NT_B^{-1} C_0(\tau_p). \tag{93b}$$

In matrix form we have

$$\vec{R}_l(\tau_p) = NT_B^{-1}T^{-1} \begin{pmatrix} \cos\omega_0\tau_p & \sin\omega_0\tau_p \\ -\sin\omega_0\tau_p & \cos\omega_0\tau_p \end{pmatrix} \vec{R}_l\, C_0(\tau_p). \tag{93c}$$

A receiver can be adaptively synchronized to the incoming signal by introducing a variable delay τ_R into the stream of baseband keying elements to generate a family of ideal baseband keying elements h_{B1} and h_{B2}. Mathematically this simply means the introduction of the variable $t + \tau_R$ into the expression for h_{B1} in Eqs. (68) and (69) in place of t. The selection of the synchronization time τ_R can greatly affect receiver performance when the actual propagation delay time τ_p is unknown (or equivalently the clocks that generate the chip streams are not in perfect initial synchronism). Therefore we rewrite Eq. (93c) to exhibit the dependence of system performance on τ_R. We obtain the following expression for $\vec{R}_l\ (\tau_p, \tau_r)$:

$$\vec{R}_l(\tau_p, \tau_R) = NT_B^{-1} \begin{bmatrix} \cos \omega_0 \tau_p & \sin \omega_0 \tau_p \\ -\sin \omega_0 \tau_p & \cos \omega_0 \tau_p \end{bmatrix} \vec{R}_l\ C_0(\tau_p - \tau_R). \tag{96}$$

The magnitude of the components of $\vec{R}_l(\tau_p, \tau_R)$ are proportional to the in-phase and quadrature components of the received radio-frequency carrier wave because of the multipliers $\cos \omega_0 \tau_p$ and $\sin \omega_0 \tau_p$ in the matrix. However, the magnitude of $\vec{R}_l(\tau_p, \tau_R)$ depends on the propagation delay time τ_p and on the receiver time τ_R by the factor $\tau_p = \tau_R$. If the receiver is badly out of synchronization, this factor can be zero or very small. If the receiver is employing phase-shift keying, any change in the propagation time τ_p, even after the receiver is perfectly synchronized on the chip ($0 \approx |\tau_p - \tau_R| << T$), will induce a corresponding phase change in the phase of the vector $\vec{R}_l = (\tau_p, \tau_R)$; that is, the doppler phase variation of the radio-frequency carrier induces an equal doppler shift on the phase angle ϕ_R. These types of doppler-signal phase variations are often seen when an mobile transmitter or receiver is used. In such a circumstance we cannot assume $\tau_p(t)$ will be constant at each instant within a bit period to permit the factorization shown in Eqs. (67) and (68). However, the other manipulations that we performed on the expression for $\vec{R}$ are all valid up to this point. To obtain the resultant correlation vector $\vec{R}_l(\tau_p, \tau_R)$ if the delay time is slowly changing within a bit period, we must average the instantaneous values of $\vec{R}_l[\tau_p(t), \tau_p]$ over an entire bit period:

$$\vec{R}_l(\tau_p, \tau_R) = NT_B^{-1}T^{-1} \int_{t_l - T_{B/2}}^{t_l + T_{B/2}} \begin{bmatrix} \cos \omega_0 \tau_p(t) & \sin \omega_0 \tau_p(t) \\ -\sin \omega_0 \tau_p(t) & \cos \omega_0 \tau_p(t) \end{bmatrix} \vec{R}_l\ C_0[\tau_p(t) - \tau_R]. \tag{97}$$

From this expression one can approximately evaluate the effect on the $\vec{R}_l(\tau_p(t), \tau_R)$ vector for various conditions of $\tau_p(t)$ and synchronization conditions for a single path propagation.

We have reviewed the elementary cases of a stationary and moving transmitter having a single path of propagation, because this is the simplest case. Linear superposition holds for the waves and correlation procedures; so a resultant from a case having several propagation paths can be obtained by addition of terms from single-path cases.

4.5 Two Propagation Modes and a Stationary Transmitter

When a stationary transmitter produces a signal at a receiver that comes by two propagation paths, there are two propagation delay times τ_{p1} and τ_{p2} and two (real-valued) amplitudes of the two modes A_1 and A_2. Alternatively, we may say that each mode is described by a complex amplitude, with the magnitude of each complex amplitude being the real amplitude and the phase being related to the propagation time. The voltage impressed on the receiving antenna by the received signal represents a waveform that is the sum of two time-delayed replicas of the initial waveform. Let the subscript i denote the mode index. Then we will get the following expression for the total correlation vector $\vec{R}^T(\tau_R)$:

$$\begin{aligned} \vec{R}_l^T(\tau_R) &= \sum_{i=1}^{2} \vec{R}_l(\tau_{p,i}, \tau_R) \\ &= NT_b^{-1}T^{-1} \sum_{i=1}^{2} A_i \begin{bmatrix} \cos \omega_0 \tau_{p,i} & \sin \omega_0 \tau_{p,i} \\ -\sin \omega_0 \tau_{p,i} & \cos \omega_0 \tau_{p,i} \end{bmatrix} \begin{bmatrix} X_l \\ Y_l \end{bmatrix} C_0(\tau_{p,i} - \tau_R). \end{aligned} \tag{98}$$

The behavior of $\vec{R}_l^T(\tau_R)$ as a function of the receiver chip synchronization time τ_R can be quite different from the behavior when there is only a single path. A splitting of the correlation peak can occur. Indeed, instead of the correlation being best when $\tau_{p,i} = \tau_R$, it could be bad at this time. Numerical examples given later exhibit this circumstance.

5.0 LONGWAVE FIELD-STRENGTH VARIATIONS FROM AN AIRBORNE ANTENNA (Based on Ref. 57)

Consider the fields generated by an inclined, elevated transmitting antenna as the antenna moves in a circular path (orbit). This antenna is an idealization of a trailing-wire antenna carried by an aircraft circling at a constant altitude.

As the antenna goes around a circular path, the VLF waveguide modes generated by the vertical and horizontal components of this inclined antenna cause periodic variations in field strength. The magnitude of these orbit-produced periodic variations is a function of the distance from the transmitter. The sense of the variation (i.e., increasing or decreasing at a given time) can be opposite at points that are relatively close together on the same radial from the transmitter. Either one or two maximum values of received signal strength per orbit can be observed. Consider the case in which the antenna undergoes a periodic change of verticality caused by a gradient in wind velocity at the altitudes through which the antenna passes (or droops). In this case, the magnitude of the periodic variation in received field strength depends on the direction of the receiver from the transmitter.

5.1 Theory

Ref. 16 gives the equations by which the electric and magnetic fields can be calculated from a vertical and horizontal antenna in the waveguide mode theory. We first consider only expressions for the vertical electric field strength, using the earth-ionosphere waveguide model in which there is no conversion of horizontal electric fields to vertical fields at the ionospheric level.

$$E_r^{ve}(r, \theta) = i(\mu_0 / \epsilon_0)^{1/2} I \, ds^{ve} \sum_n S_n^{1.5} F_n \tag{99}$$

$$E_r^{he}(r, \theta) = i(\mu_0 / \epsilon_0)^{1/2} I \, ds^{he} \cos\phi \sum_n \Delta_n (z_s) S_n^{0.5} F_n \tag{100}$$

where

$$F_n = \frac{1}{(D\lambda h)^{1/2}} \left[\frac{D / a}{\sin(D / a)} \right]^{1/2} G_n(z) G_n(z_s) \Lambda_n \cdot \exp(i\pi / 4 + ik_0 D S_n) \tag{101}$$

Here

$E_r^{ve}(r, \theta)$, $E_r^{he}(r, \theta)$ indicate the vertical electric field strength at the point r, θ received from a vertical or a horizontal electric dipole transmitting antenna, respectively (see Fig. 67).

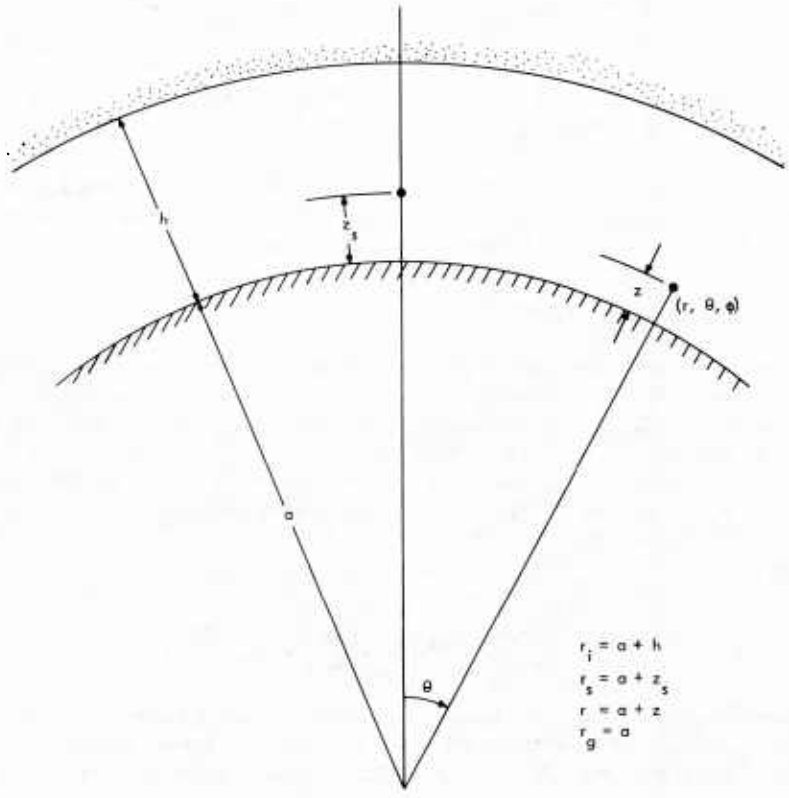


Fig. 67: Diagram of sender and receiver locations

$(\mu_0/\epsilon_0)^{1/2}$ is the impedance of free space.

I is the transmitting antenna current.

ds^{he}, ds^{ve} indicate the length of the horizontal or vertical transmitting antenna.

S_n is the sine of the complex eigenangle for the nth TM mode.

$\Delta_n(z_s)$ is the impedance of the nth TM mode at the height z_s of the transmitter.

ϕ is the angle between the great-circle path to the receiver and the end-fire direction of the horizontal antenna.

λ is the free-space wavelength of the electromagnetic wave.

D is the great-circle distance between the transmitter and the receiver.

a is the radius of the earth.

$G_n(z)$ is the height-gain function of the nth mode.

Λ_n is the excitation factor of the nth mode.

$k_o = 2\pi/\lambda$.

The total field $E_r(r,\theta)$ from an inclined antenna is simply the sum of the fields from the horizontal and vertical components of the inclined antenna when the inclined antenna's length ds is broken up into its horizontal and vertical components ds^{he} and ds^{ve}; and they are inserted in (1) and (2):

$$E_r(r, \theta) = E_r^{ve}(r, \theta) + E_r^{he}(r, \theta) \tag{102}$$

To simulate mathematically the effect of the motion of an antenna trailing behind a circling aircraft, we construct a Cartesian coordinate system with its origin on the earth's surface below the center of the circle. The z axis is normal to the earth's surface and the y axis points in the direction of the receiver. In these (x, y, z) coordinates, we locate the antenna originally at the point (r_o,O,z_s), where r_o is the radius of the orbit and z_s is the height of the antenna above the ground. The receiver is located at the point (O, d, O), i.e., on the y axis on the surface of the earth a great-circle distance d from the origin. A diagram of this arrangement is shown in Fig. 68. The antenna is a short straight-line element centered at (r_o,O,z_s) and inclined at the angle γ from the normal. (Therefore, $ds^{he} = ds \sin\gamma$ and $ds^{ve} = ds \cos\gamma$.) The horizontal projection of the antenna points along the y axis. We define ζ as the angle between the x axis and the projection in the x-y plane of the line from the origin of the coordinate system to the center of the antenna. Since $d >> r$, we use the approximation $\phi \approx \zeta$, which states that, when the horizontal projection of the antenna is parallel to the y axis, this horizontal projection points directly at the receiver. There is a small error in this approximation, but in this calculation it is at most of the order of $r_o/d \approx 0.01$ radian. The great circle path D from the antenna to the receiver is calculated according to

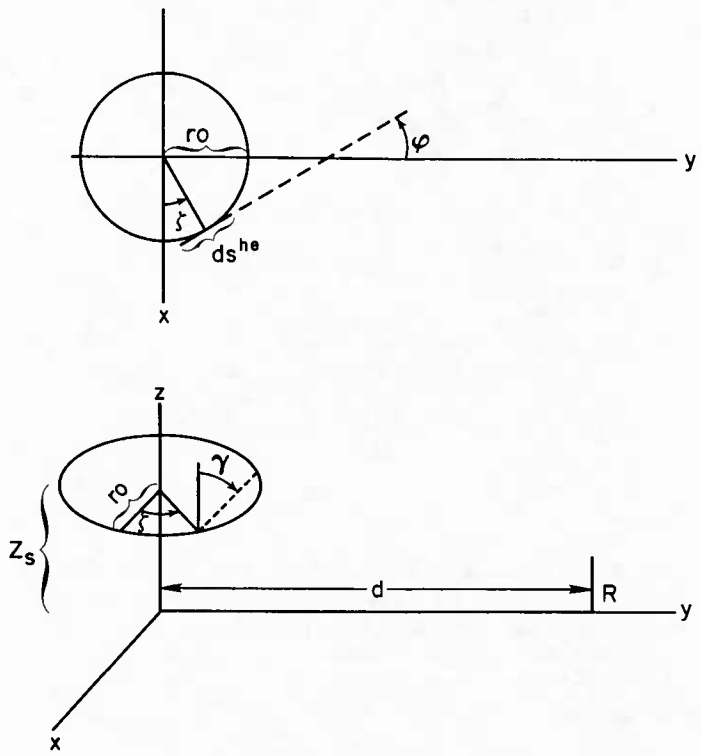


Fig. 68: Diagram defining various angles. Upper diagram shows orbit from above.

$$D = (d^2 + r_o^2)^{1/2} \tag{103}$$

which is exact for a flat earth and is a good approximation for the spherical earth. With these values we then calculate $E_r(r,\theta)$ at the receiver. Next, we wish to carry the antenna smoothly about a circle. We choose some incremental angle $\Delta\zeta$ and move the antenna to the new coordinates (x_s', y_s', z_s') from the old coordinates (x_s, y_s, z_s) which were (r_o, O, z_s) according to the transformation

$$x_s' = x_s \cos\Delta\zeta - y_s \sin\Delta\zeta \tag{104a}$$

$$y_s' = x_s \sin\Delta\zeta + y_s \cos\Delta\zeta \tag{104b}$$

$$z_s' = z_s \tag{104c}$$

which is a rotation of the antenna location by an angle $\Delta\zeta$ about the z axis. The previous angles ζ and ϕ, which were zero, are now changed to ζ' and ϕ' angles according to

$$\phi' = \phi + \Delta\zeta = \zeta' = \zeta + \Delta\zeta \tag{105}$$

A new great-circle distance D' between the transmitting antenna and the receiver is calculated according to

$$D' = ((x_s')^2 + (d - y_s')^2)^{1/2} \tag{106}$$

When we consider the case of a static inclined antenna that is being carried in a circle, the angle γ of the antenna from the vertical is unchanged. Because of the changes in the antenna direction and in the winds at different altitudes, we must consider that the antenna might undergo a periodic change in inclination as it traverses a circle. We can take this yo-yo motion into account by letting the angle γ oscillate about some average angle γ_0 as the antenna goes about an orbit by setting.

$$\gamma = \gamma_0 + \Delta\gamma \sin(\zeta + \zeta_0) \tag{107}$$

where ζ_0 is a phase angle that relates the orbital variation of the end-fire angle ϕ with the variation of the inclination angle γ. Static conditions exist when $\Delta\gamma = 0$. The vertical electric field for the new antenna location is calculated according to (99) through (102). Then the antenna is moved through a further $\Delta\zeta$ rotation, and the process is repeated. From the results of these calculations, the variations of amplitude and phase are determined as the antenna proceeds about a circle.

5.2 Calculations and Results

For our sample calculations, we chose the mode parameters of Galejs for 20-kHz transmissions. The numerical values that we used are given in Table 5.

Table 5 — Constants Used in the Calculations

$\Lambda_1 = 0.4$	$G_1(z) = 1$
$\Lambda_2 = 1.2$	$G_2(z) = 1$
$S_1 = 1.002 + 0.000689i$	$\Delta_1(z_s) = -0.02i$
$S_2 = 0.993 + 0.00138i$	$\Delta_2(z_s) = 0.05i$
$z = 0$ km	$r_0 = 4$ km
$G_1(z_s) = 0.965 + 0.259i$	$z_s = 10$ km
$G_2(z_s) = 0.95$	Radiated power = 1 kw

After each complete rotation, we moved the receiver 100 km farther from the transmitter over a distance range of 500 to 5500 km. The topmost curve of Fig. 69 shows the vertical electric field strength obtained by considering the first- and second-order modes from a vertical antenna as a function of distance. The fourth curve from the top of Fig. 69 shows the maximum field strength as a function of distance from an elevated horizontal antenna. Note that the modal interference pattern for the horizontal antenna is quite different from that of the vertical antenna. Since an isotropic model is considered here, these results do not include the effects of polarization changes on reflection from the ionosphere.

The second and third curves from the top of Fig. 69 show field strength as a function of distance for antennas inclined with $\gamma = 45°$ and $\gamma = 80°$, respectively, for $\Delta\gamma = 0°$. The error bars represent the maximum and minimum values obtained as the antenna follows an orbit. The lower portion of Fig. 69 shows the relative phase of the fields from the vertical and horizontal antennas.

Figure 70 gives in more detail the variation of amplitude and phase for an inclined antenna ($\gamma = 80°$) at several ranges. Figure 71 shows the field strength variations with distance for the values of $\gamma_0 = 60°$, $\Delta\gamma = 11°$, and $\zeta_0 = 0°$.

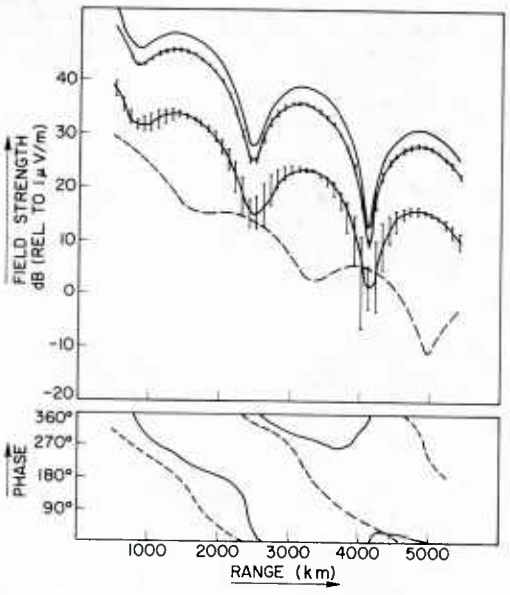


Fig. 69: Amplitude and relative phase plots of fields generated by vertical, inclined, and horizontal antennas.

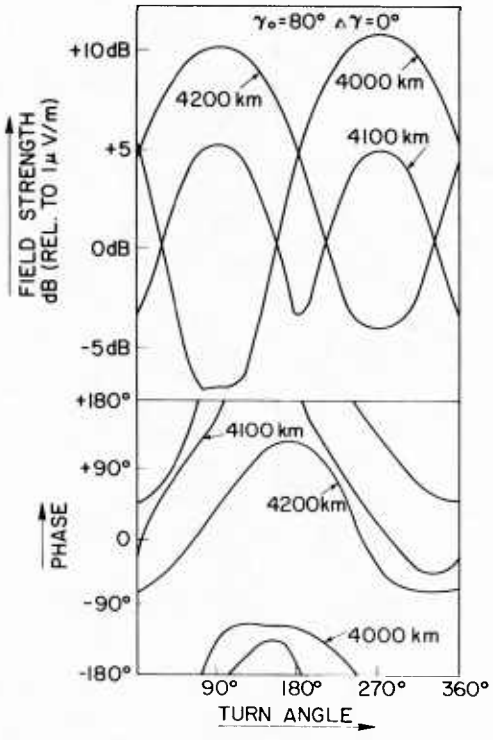


Fig. 70: Amplitude and phase plots of fields as a function of the turn angle ζ of the antenna

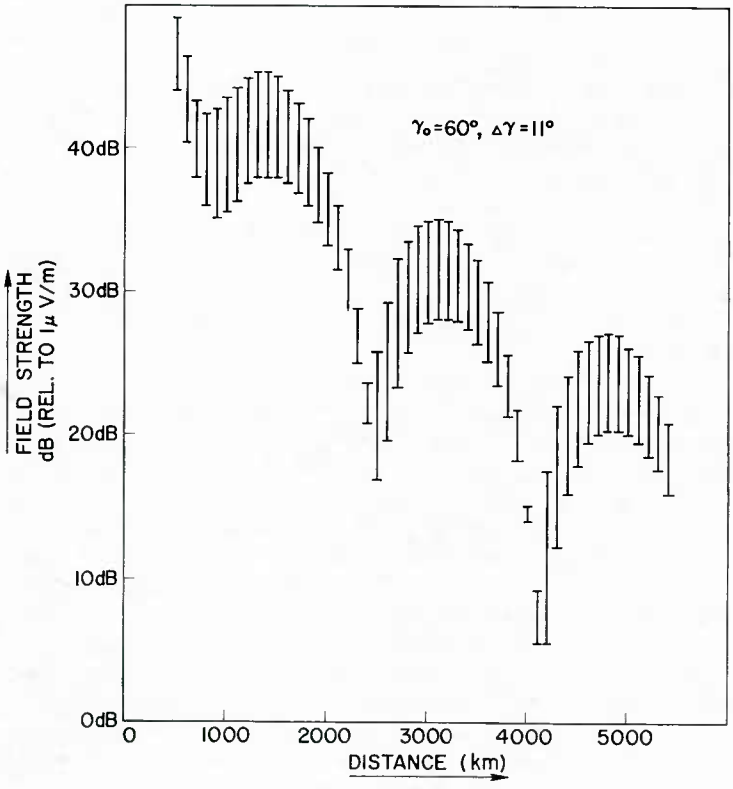


Fig. 71: Amplitude plot of field generated by an orbiting, yo-yoing antenna

5.3 Discussion and Results

Figure 69 indicates that the amplitude of the orbit-produced, periodic variations of the received field strength is a function of the distance from the transmitter. In Fig. 70 we see a complete reversal of the sense of the oscillation between 4000 and 4200 km. At 4100 km the field strength has two relative maxima per orbit. This behavior seems very puzzling, but it has a natural explanation. Figure 72 shows a complex phasor representation of the sum contained in (102). The total vertical electric field strength is the phasor sum of the field strengths generated by the vertical and horizontal components of the antenna, E_r^V and E_r^H. (These are the same as E_r^{ve} and E_r^{he}, respectively.)

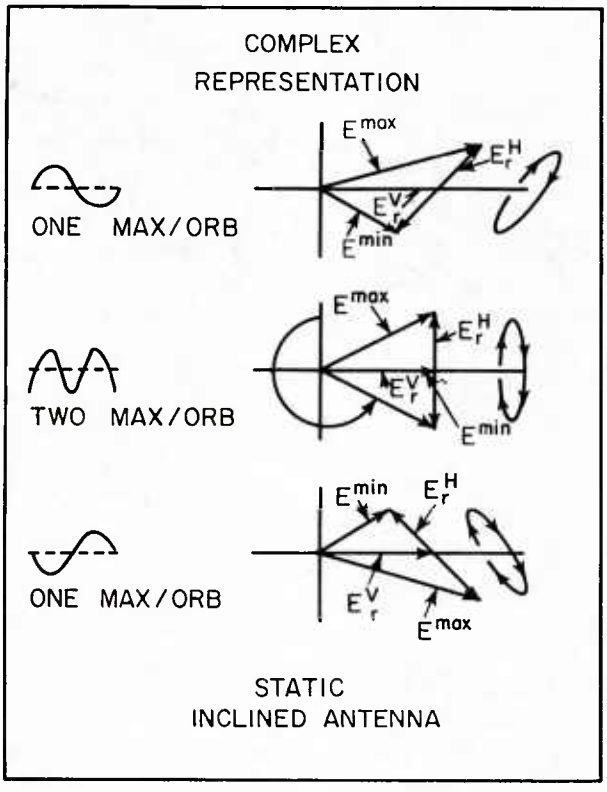


Fig. 72: Diagram showing how the number of maximum field strength values per orbit may change with the distance from the transmitter.

These phasors may, in general, have any angle with respect to one another, and this angle varies with distance. The E_r^V phasor represents an omni-directional radiation pattern that does not change with orbit (as long as γ is fixed). The E_r^H phasor is not omnidirectional, however, and varies as the cos ϕ (see expression 100); therefore, it is modulated by a factor that changes from $+1$ to -1 with the orbit. In Fig. 72 this is represented by $\pm E_r^H$. The resultant field strength is represented by a phasor, one end of which is fixed at the origin; the other end slides along the E_r^H phasor. The sense of the variation of the resultant phasor is indicated by the curved trace on the right-hand side of the figure. The field strength, therefore, goes first through a maximum and then to a minimum as the antenna orbits. The middle diagram of Fig. 72 shows the E_r^H phasor at 90° to the E_r^V phasor; this relation causes two relative maxima in field strength per orbit. When the angle between phasors rotates farther, as shown on the bottom of Fig. 72, the case of a single maximum per orbit is reestablished, but now the variation is in the opposite sense to that shown in the top diagram of Fig. 72.

From Fig. 69 we see a gradual change in the size of the periodic variations with distance. We also see that quite different modal interference patterns are set up by horizontal and vertical antennas. (The difference is caused by the presence of the impedance $\Delta_n(z_s)$ in (100) and by its absence in (99)). The ratio of E_r^H to E_r^V changes with distance, and the nulls of the two interference patterns occur at different ranges. This effect is diagramed in Fig. 73. The lower portion shows two horizontally generated phasors from two modes that combine to give a resultant phasor. Two different E_2^H phasors are shown to illustrate the relative rotation of the horizontally generated phasors of the first and second modes as the distance from the transmitter changes. In the upper portion, two vertically generated phasors also combine to give a resultant. Here also, two different E_2^V phasors are shown to illustrate the relative rotation of the vertically generated phasors of the first and second modes as the distance from the transmitter changes. The vertically generated resultant is then added to the horizontally generated resultant. The horizontally generated resultant is modulated by a cos ϕ term; therefore, the final resultant phasor may have one or two maxima per orbit, as we mentioned previously. As illustrated in the upper diagram of Fig. 73, the relative rotation of the mode phasors can cause the condition to change from a single maximum per orbit at one distance to two relative maxima per orbit at another distance. With the help of diagrams such as Figs. 72 and 73, the behavior shown is Figs. 69 and 70 can be understood.

When the angle γ changes with the orbit angle, as in (107), we find an additional periodic variation that enhances or reduces the received field strength variation caused by purely orbital motion. This is illustrated in Fig. 71, where a large, abrupt change in the field strength variation is predicted between 4000 and 4200 km and between 2400 and 2500 km. For these cases, the purely orbital variation changes its sense of direction, whereas the yo-yo variation does not change appreciable in size or sense. On the near side of the null at 2400 km, the two fluctuations are out of phase and tend to cancel each other (1.3-db oscillation). On the far side of the null at 2500 km, however, the fluctuations are in phase and reinforce to give a 12-db oscillation. If we had chosen ζ_0 to be 180° instead of 0°, the situation would be reversed; thus, the variations are dependent on bearing as well as on range. The reference direction for determining the bearing dependence is provided by the wind-shear direction.

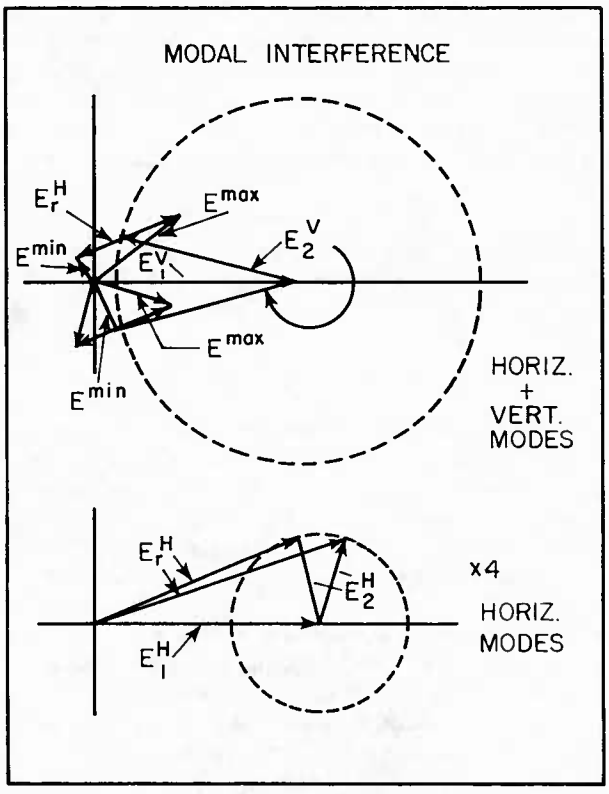


Fig. 73: Diagram showing how changes in modal interference with range can change the received amplitude pattern

For the cases that have been considered, we have assumed a homogeneous isotropic ionosphere, i.e., no earth's magnetic-field effects. For this isotropic condition, the vertical electric field E_r comes from the transverse magnetic (TM) modes. The transverse electric (TE) modes from the horizontal component are generated with a sin ϕ variation, but they do not carry any E_r field. When the earth's magnetic field is taken into account, however, we find that the TE modes become quasi-TE and have an appreciable E_r field that is due to polarization conversion at the ionosphere. This will create an additional complication, which is illustrated in Fig. 74. Here, the E_r^{ve} phasor is fixed in magnitude, but the E_r^{he} (TM) and E_r^{he} (TE) phasors are modulated by cos ϕ and sin ϕ, respectively, as the antenna goes around an orbit. These three phasors must be added together to give the resultant observed field:

$$E_r = E_r^{ve} + E_r^{he}(\text{TM}) + E_r^{he}(\text{TE}) \tag{108}$$

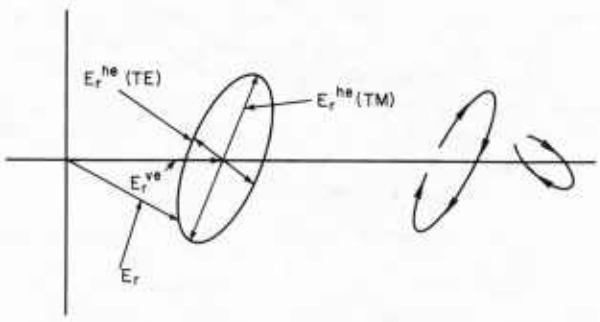


Fig. 74: Diagram showing the additional complication that the quasi-TE modes may make in the received field variation

Actually, there will be both E_r^{ve} (TE) and E_r^{ve} (TM) fields, but they are combined to form a constant E_r^{ve} field. The E_r^{he} (TE) phasor has a sinusoidal variation. The variation is shifted by just 90° in ϕ angle from the E_r^{he} (TM) phasor. Instead of the end of the resultant phasor E_r following the line traced out by the E_r^{he} (TM) phasor, it will now follow around an ellipse as ϕ goes from 0° to 360°, as shown in Fig. 74. As E_r^{he} (TM) goes through zero, E_r^{he} (TE) reaches its maximum, and vice versa. It is possible, of course, for the ellipse to become either a straight line or a circle, depending on the relative magnitudes and directions of E_r^{he} (TM) and E_r^{he} (TE) on the phasor plane.

For an orbiting horizontal antenna, the homogeneous, isotropic ionosphere model predicts that the field strength will be zero for $\phi = 90°$. However, the introduction of the earth's magnetic field changes this prediction; the contribution from the quasi-TE modes has a maximum at $\phi = 90°$. Since the quasi-TE and quasi-TM modes generate different modal interference patterns, there should exist some combination of range, frequency, and bearing at which the ellipse of Fig. 74 is a circle. In this case, a horizontal orbiting antenna would cause no amplitude variation at the receiver. This special case is not likely to be experienced often, but this possibility shows the diversity of the results that may be obtained from an orbiting antenna.

We see that the variation of the signal received from an orbiting antenna is quite complicated, even though the antenna itself is a simple short dipole. The magnitude of the variation is a function of the range and bearing of the receiver. Large changes in the size of the variations can occur with small changes in range.

6.0 ANTENNA MOTION EFFECTS ON MODULATION SYSTEMS (Based on Ref. 58)

It is easy to show the effects of rotation by an inclined elevated antenna given the previous development of the mathematics. The horizontal component of the inclined antenna generates both TE and TM modes; and the horizontally generated fields have phases and amplitudes different from those generated by the vertical antenna components. The waveguide modal interference pattern for an elevated horizontal antenna differs from that of a vertical antenna. This effect is illustrated in Fig. 69, which shows the fields generated by variously inclined antennas at 20 kHz. The amplitudes of the horizontally generated fields vary as the cosine of the angle between the horizontal antenna component and the direction of propagation.

In simulating the effect of this variation on the receiver output vector, we use Eq. (98) for the received correlation vectors, and we use four amplitudes $\{A_l\}$, four phase time delays $\{\tau_l^p\}$, and four group time delays $\{\tau_l^g\}$. Two of the amplitudes represent two vertically generated mode fields and vary as $\sin\gamma$. The other two represent horizontally generated TM mode fields and vary as the product $\cos\zeta\cos\gamma$. The resultant expression for the correlator output is the same as Eq. (98) except that in this case the correlation output is the summation of the outputs for the four individual modes.

Denoting the $X_l^{(i)}(\tau_R)$ from the vertically electrically generated modes as $X_l^{(i)ve}(\tau_R)$ and the $X_l^{(i)}(\tau_R)$ from the horizontally electrically generated modes as $X_l^{(i)he}(\tau_R)$, we can write

$$X_l^T(\tau_R) = \cos\gamma \sum_{i=1}^{N'} X_l^{(i)ve}(\tau_R) + \cos\zeta \sin\gamma \sum_{i=1}^{N'} X_l^{(i)he}(\tau_R). \tag{109a}$$

Parallel definitions can be made for $Y_l^T(\tau_R)$, $Y_l^{(i)ve}(\tau_R)$ and $Y_l^{(i)he}(\tau_R)$, and we obtain

$$Y_l^T(\tau_R) = \cos\gamma \sum_{i=1}^{N'} Y_l^{(i)ve}(\tau_R) + \cos\zeta \sin\gamma \sum_{i=1}^{N'} Y_l^{(i)he}(\tau_R). \tag{109b}$$

The vector $\vec{R}_l^T(\tau_R)$ is a sum of two vectors $\vec{R}_l^{he}(\tau_R)$ and $\vec{R}_l^{ve}(\tau_R)$:

$$\vec{R}_l^T(\tau_R) = \cos\gamma\, \vec{R}_l^{ve}(\tau_R) + \cos\zeta \sin\gamma\, \vec{R}_l^{he}(\tau_R). \tag{110}$$

The vectors $\vec{R}_l^{ve}(\tau_R)$ and $\vec{R}_l^{he}(\tau_R)$ can be parallel, perpendicular, or opposed to each other, depending on circumstances and the value of τ_R. The resultant $\vec{R}_l^T(\tau_R)$ vector can change quite dramatically during a complete orbit. For example, if $\vec{R}_l^{ve}(\tau_R)$ and $\vec{R}_l^{he}(\tau_R)$ are of the same magnitude and phase, there would be one large oscillation of $|\vec{R}_l^T(\tau_R)|$ per orbit (provided $\gamma = 45°$). If $\vec{R}_l^{ve}(\tau_R)$ and $\vec{R}_l^{he}(\tau_R)$ are perpendicular to each other, there would be two oscillations of $|\vec{R}^T(\tau_R)|$ with one cycle ζ. There is a mathematical isomorphism between the phasor $E_r(D)$ and the vector $\vec{R}_l^T(\tau_R)$, since the composition of various phasors which form the complex phasor sum $E_r(D)$ is isomorphic to the composition of various correlation vectors which form the two-dimensional correlation vector sum $\vec{R}_l^T(\tau_R)$. When both TE and TM waveguide mode propagation from an airborne transmitting facility are taken into account, we obtain the following relation.

$$\vec{R}_l^T(\tau_R) = \vec{R}_l^{ve}(\tau_R)\cos\gamma + \sin\gamma[\cos\zeta \vec{R}_l^{he}(TM,\tau_R) + \sin\zeta \vec{R}_l^{he}(TE,\tau_R)]. \tag{111}$$

The total vector $\vec{R}_l^T(\tau_R)$ is the sum of independently propagated vertical endfire- and broadside-generated components. These components are themselves sums of components which are propagated by different waveguide modes.

6.1 Diagrammatic Representations

It is instructive to view diagrams that illustrate the behavior of the various $\vec{R}_l$ vectors as τ_R and ζ vary. The following example illustrates what can happen in a particularly severe case of modal interference. Figure 75 depicts the variation of a single-mode correlation output $\vec{R}_l^{(1)ve}(\tau_R)$ from a stationary vertical electric transmitter. Figure 76 shows an output function $\vec{R}_l^{(2)ve}(\tau_R)$ for a signal coming slightly later on a second-order mode and 180° out of phase with the first. Figure 77 gives the resultant $\vec{R}_l^{ve}(\tau_R)$, which is the sum of the correlation outputs of the first and second modes. Figure 78 shows the phase of the vector $\vec{R}_l^{ve}(\tau_R)$ as a function of τ_R.

Figure 79 shows the resultant $\vec{R}_l^T(\tau_R) = \cos\gamma\, \vec{R}_l^{ve}(\tau_R) + \sin\gamma\cos\zeta \vec{R}_l^{he}(\tau_R)$, where ζ is the angle between the direction of propagation and the endfire direction of the antenna. The angle γ is the inclination angle of the antenna; we take $\gamma = 45°$

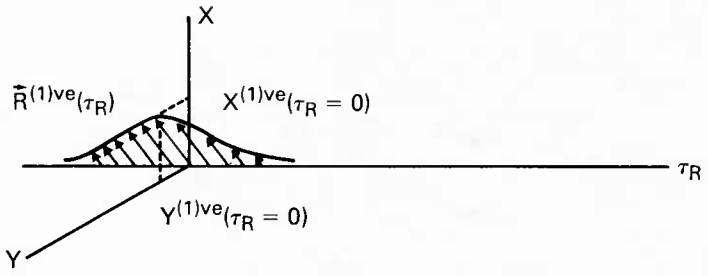


Fig. 75: Graph of $\vec{R}_l^{(1)ve}(\tau_R)$ vs τ_R for a stationary vertical electric-dipole transmitter

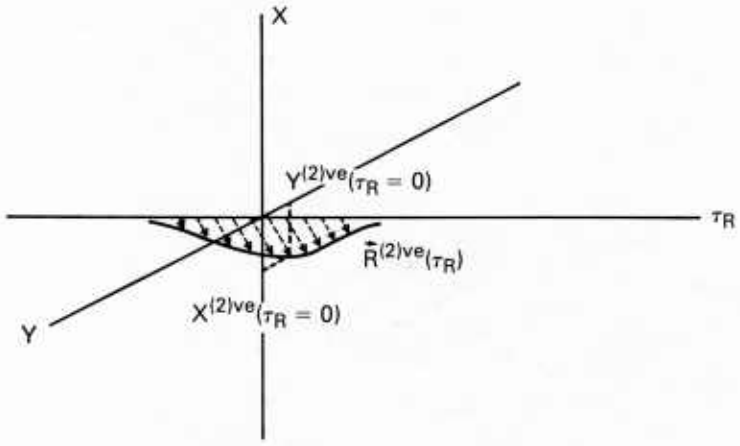


Fig. 76: Graph of $R_l^{(2)ve}(\tau)$ vs τ_R for a stationary vertical electric dipole transmitter. The phase of this wave is 180° opposite to that illustrated in Fig. 75. The maximum value of $|R_l^{(2)ve}(\tau_R)|$ occurs at slightly greater τ_R values than does the maximum value of $|R_l^{(1)ve}(\tau_R)|$ shown in Fig. 75.

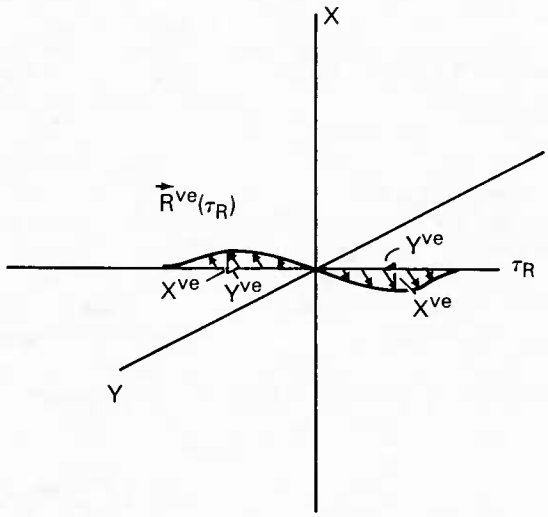


Fig. 77: Graph of $\vec{R}_l^{ve}(\tau_R) = \vec{R}_l^{(1)ve}(\tau_R) + \vec{R}_l^{(2)ve}(\tau_R)$, showing an abrupt change of phase at $\tau_R = 0$ and a large shift in the value of τ_R for which $|\vec{R}_l^{ve}(\tau_R)|$ is greatest

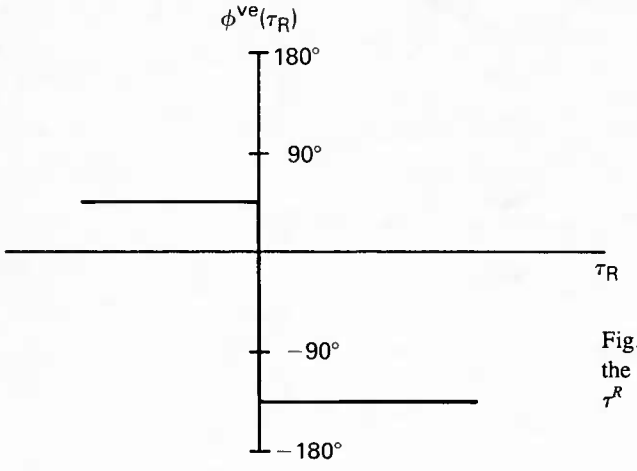


Fig. 78: Graph of $\phi_l^{ve}(\tau_R)$, the angle of the $\vec{R}_l^{ve}(\tau_R)$ vector with the x axis in Fig. 77, showing an abrupt shift of phase at $\tau^R = 0$

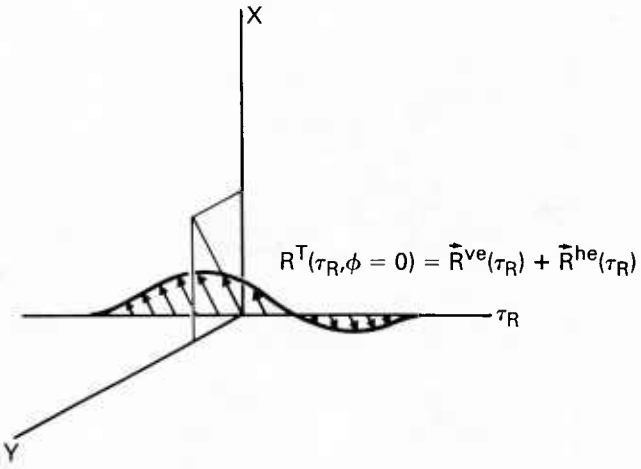


Fig. 79: Graph of $\vec{R}_l^T(\tau_R)$ for an endfire case ($\cos \zeta = +1$), illustrating the change in the value of τ_R for which $|\vec{R}_l^T(\tau_R)| = 0$ and the variation in the size of the peak values of $|\vec{R}_l^T(\tau_R)|$ on opposite sides of $\tau_R = 0$. Here a relatively small value of $|\vec{R}_l^{ve}(\tau_R)|$ is assumed.

and $\zeta = 0$ for this example. Figure 80 shows the same resultant except for the condition $\zeta = 180°$. The conditions for $\zeta = 90°$ and $\zeta = 270°$ are both shown by the pattern of Fig. 77, since $\vec{R}_l^T(\tau_R)$ is composed solely of the vertically generated modes in these cases. These figures give a picture of $\vec{R}_l^T(\tau_R)$ at the four cardinal points of an orbit.

In Fig. 80 four sample synchronization times, designated τ_{R1}, τ_{R2}, τ_{R3}, and τ_{R4}, represent possible correlation times that might occur within a receiver. We look at the variation of the correlation vector magnitude $|\vec{R}_l^T(\tau_{Ri})|$ and phase $\phi_l^T(\tau_{Ri})$ during a complete orbit for these times. Figures 81 and 82 show the phase and magnitude of $\vec{R}_l^T(\tau_{R1})$. The phase is constant in this case, because $\vec{R}_l^{ve}(\tau_{R1})$ is greater than and parallel to $\vec{R}_l^{he}(\tau_{R1})$. The magnitude varies in a mild way. In Figs. 83 and 84 showing the phase and amplitude for the time τ_{R2}, the phase makes two quick successive 180° changes in the neighborhood of $\zeta = 180°$, and the amplitude goes through two successive nulls. Figures 85 and 86 give $\vec{R}_l^T(0)$. This vector varies cosinusoidally, because there is no contribution from the vertical electric antenna component. Figures 87 and 88 show the behavior of $\vec{R}_l^T(\tau_{R3})$, which is similar to the behavior of $\vec{R}_l^T(\tau_{R2})$ except that the nulls occur near $\zeta = 0°$ rather than $\zeta = 180°$. Figures 89 and 90 show the behavior of $\vec{R}_l^T(\tau_{R4})$, which is similar to $\vec{R}_l^T(\tau_{R1})$ except for a shift in orbit angle ζ of 180°.

Finally, Fig. 91 shows an estimate of τ_{Ropt}, the optimum value of receiver correlation time for different values of orbit angle ζ. This graph shows that τ_{R1} and τ_{R4} are preferable when $|\vec{R}_l(\tau_{R1})|$ and $|\vec{R}_l(\tau_{R4})|$ are largest. The phase is constant, and the amplitude varies gently. Relatively mild amplitude behavior occurs for this correlation time because $|\vec{R}_l^{ve}(\tau_{R1})| > |\vec{R}_l^{he}(\tau_{R1})|$. The phase is constant because $\vec{R}_l^{ve}(\tau_{R1})$ is parallel to $\vec{R}_l^{he}(\tau_{R1})$. Although $\vec{R}_l^{he}(\tau_{R1})$ can change the magnitude of $\vec{R}_l^T(\tau_{R1})$, it does not change the phase.

In the preceding example, there was minimum correlation near the middle of the correlation envelope, and the correlation maximum was greatly displaced from the maximum point of the correlation of either individual multipath signal. Thus, the rather small difference in arrival times $\tau_{p1} - \tau_{p2}$ changes the optimum chip synchronization time τ_R by a major amount. This effect is fairly easy to explain. The condition described by our equations is a severe case of multipath interference.

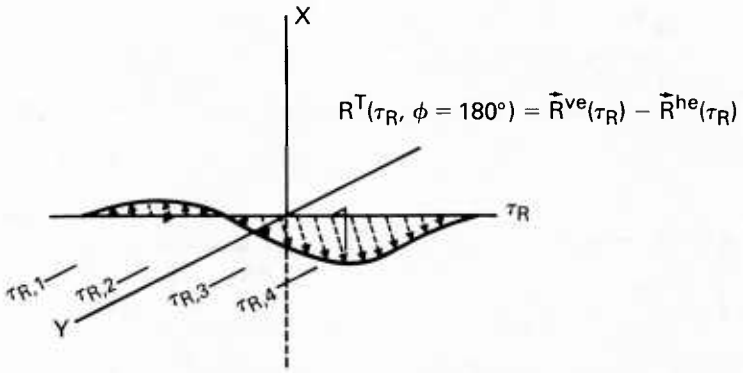


Fig. 80: Graph of $\vec{R}_l^T(\tau_R)$ for the opposite endfire case ($\cos \zeta = -1$) to the case in Fig. 79. Here the largest value of $| \vec{R}_l^T(\tau_R) |$ occurs for $\tau_R > 0$ whereas, in Fig. 79 it was for $\tau_R < 0$. Synchronization times (τ_R) to be considered in subsequent figures are shown.

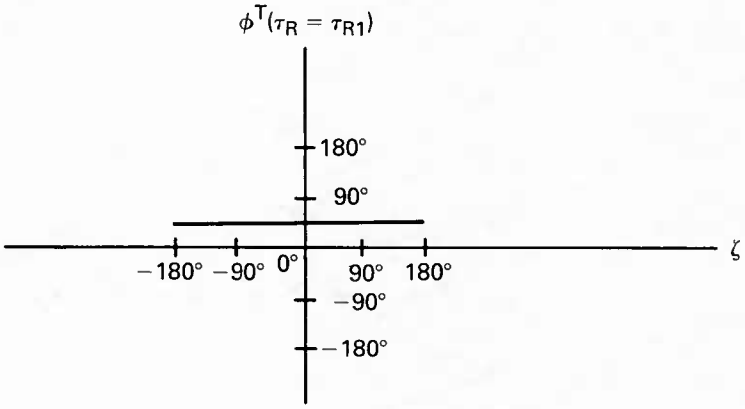


Fig. 81: Graph of $\phi_l^T(\tau_{R1})$ as the angle of turn ζ of an inclined antenna changes from $-180°$ to $+180°$ for the synchronization time τ_{R1}. This phase is constant because the contribution of the vertical antenna component always exceeds that of the horizontal component for this synchronization time.

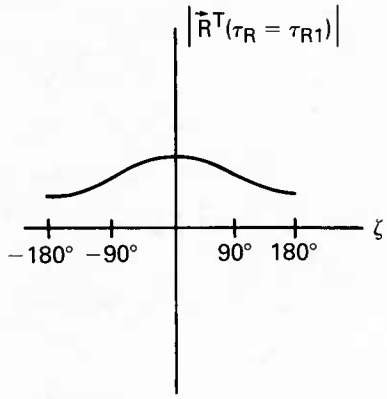


Fig. 82: Graph of $| \vec{R}_l^T(\tau_{R1}) |$ as ζ varies from $-180°$ to $+180°$. The relatively gentle variation of $| \vec{R}_l^T(\tau_{R1}) |$ occurs because $| \vec{R}_l^{ve}(\tau_{R1}) | > | \vec{R}_l^{he}(\tau_{R1}) |$.

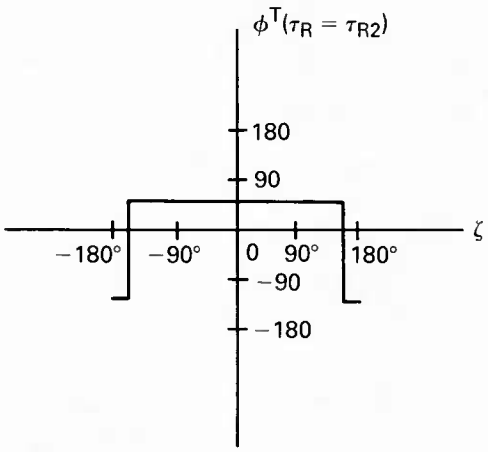


Fig. 83: Graph of $\phi_l^T(\tau_{R2})$ vs orbit angle ζ. For this lock-on time there are two rapid and large phase changes near $\zeta = 180°$.

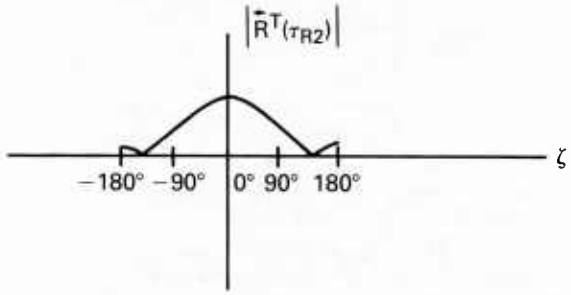


Fig. 84: Graph of $| \vec{R}_l^T(\tau_{R2}) |$ vs orbit angle ζ, showing how this function goes to zero when $\phi_l^T(\tau_{R2})$ is making the abrupt phase transition

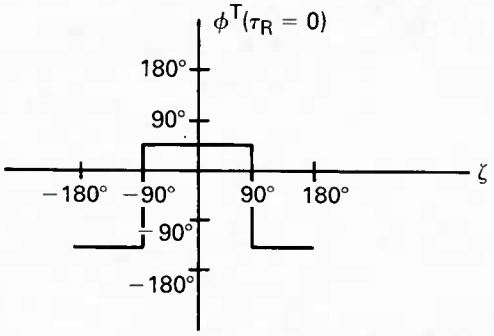


Fig. 85: Graph of $\phi_l^T(\tau_R = 0)$ vs angle ζ. There is essentially no contribution from the vertical component for this value of τ_R, so the phase is controlled by the horizontal component of the transmitting antenna.

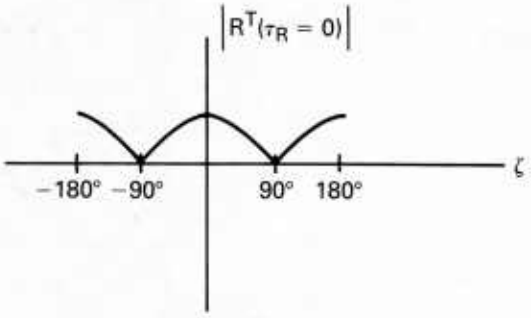


Fig. 86: Graph of $|\ \vec{R}_l^T(\tau_R = 0)\ |$ vs orbit angle ζ. The cosinusoidal variation is caused by the cos ζ modulation of the horizontal component. The vertical component contributes essentially zero.

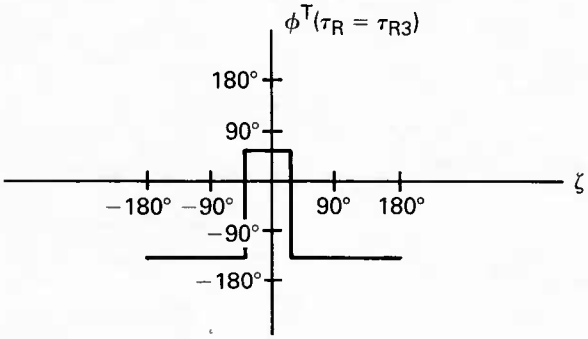


Fig. 87: Graph of $\phi_l^T(\tau_{R3})$ vs orbit angle ζ. For this synchronization time the abrupt changes in phase occur around the $\zeta = 0$ point rather than around the $\zeta = 180°$ for the τ_{R2} case illustrated in Fig. 83.

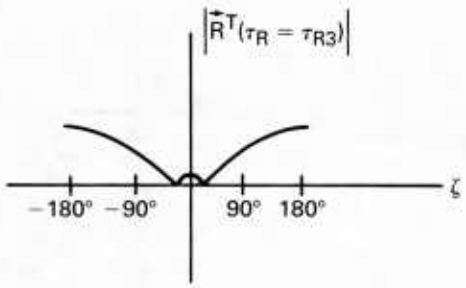


Fig. 88: Graph of $|\ \vec{R}_l^T(\tau_{R3})\ |$ vs orbit angle ζ. The smallest values of $|\ \vec{R}_l^T(\tau_{R2})\ |$ occur for orbit angles near zero degrees rather than near 180° as was the case for $|\ \vec{R}_l^T(\tau_{R2})\ |$ as shown in Fig. 78.

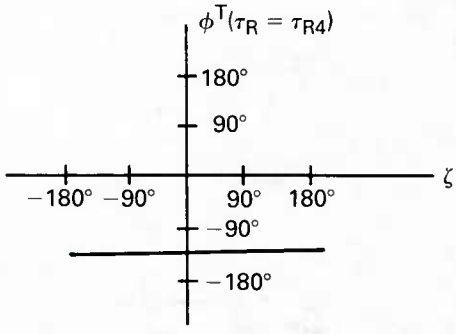


Fig. 89: Graph of $\phi_l^T(\tau_{R4})$ vs orbit angle ζ. for this receiver time the vertically generated correlation vector dominates; that is $|\ \vec{R}_l^{ve}(\tau_{R4})\ | > |\ \vec{R}_l^{he}(\tau_{R4})\ |$. No change of phase can occur, since $\vec{R}_{l_{ve}}(\tau_{R4})$ and $\vec{R}_l^{he}(\tau_{R4})$ are parallel.

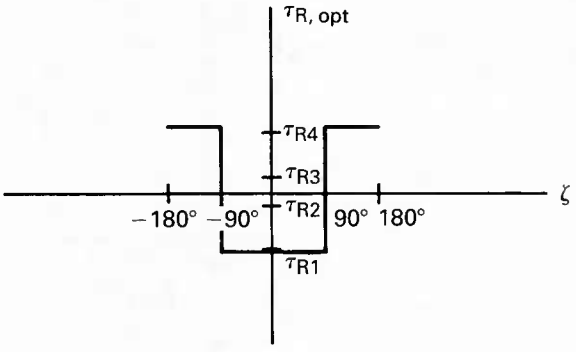


Fig. 90: Graph of $|\ \vec{R}_l^T(\tau_{R4})\ |$ vs orbit angle ζ. For this receiver time the maximum correlation vector occurs for $\zeta = 180°$, whereas for τ_{R1} the maximum correlation-vector magnitude occurred at $\tau = 0°$.

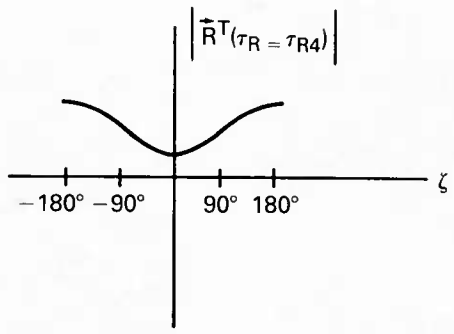


Fig. 91: Graph showing the values of τ_R that should be chosen for various orbit angles ζ for the case illustrated in Figs. 75 through 90

Two signals come to the receiver antenna offset in time by $\tau_{p1} - \tau_{p2}$. The first and second signals travel independently of each other, so when the first signal pulse arrives, the signal begins to build up. Then the second signal arrives out of phase and cancels out the first signal for a long interval of time while both are strong. Then the first signal goes to zero because its waveform has passed by, while the second signal remains to give an excitation of the antenna again. Thus the strongest portions of each signaling element in this case of interference occur at the beginning and the end of each signaling element. In the middle of the element, the two signals arriving from different paths interfere, and there is a reduction of the field present.

When a receiver tries to synchronize with the signal at its strongest point, it must search for synchronization either at the beginning of the chip or at the end but not in the middle. Thus, if a system were presynchronized with an optimum τ_R for either of the individual mode delays or for an average of the two mode delays, it would actually be poorly synchronized for optimum operation! For better operation three correlators might be used—one adjusted at the central expected time of arrival of the wave, and one at each end of the chip. If the two waveforms reinforce each other to give the maximum in the middle, the central correlation peak will be large and the two side peaks will be small. If however, an interference condition is present, the middle correlator will be weak and the two side correlators will be strong.

6.2 Numerical Calculations

6.2.1 Procedures

We will calculate some values of $\vec{R}_l^T$ for cases that are inspired by realistic mode-theory calculations of the phase and amplitude of VLF waves. For the calculated fields shown in Fig. 6, the waveguide mode constants are given in Table 5.

These constants were used in the following equations to generate the field -strength values in Fig. 69:

$$E_r^{ve}(r,\theta) = i(\mu_0/\epsilon_0)^{1/2} I\, ds^{ve} \sum_n S_n^{1.5} F_n \tag{112}$$

and

$$E_r^{he}(r,\theta) = i(\mu_0/\epsilon)^{1/2} I\, ds^{he} \cos\phi \sum_n \Delta_n(z_s) S_n^{0.5} F_n. \tag{113}$$

where

$$F_n = \frac{1}{(D\lambda h)^{1/2}} \left[\frac{D/a}{\sin(D/a)}\right]^{1/2} G_n(z) G_n(z_s) \Lambda_n \exp(i\pi/4 + ik_0 DS_n). \tag{114}$$

These equations are simplified versions, since they contain only TM-mode constants of an isotropic ionosphere. We calculate $\vec{R}_l^{(i)ve}$ for each mode according to

$$X_l^{(i)ve}(\tau_R,\zeta) = N\,|A_i^{ve}|\,\cos[\omega_0 \tau_i^{p,ve}(\zeta)]\; C\,(\tau_i^{g,ve}(\zeta) - \tau_R) \tag{115}$$

and

$$Y_l^{(i)ve}(\tau_R,\zeta) = N\,|A_i^{ve}|\,\sin[\omega_0 \tau_i^{p,ve}(\zeta)]\; C\,(\tau_i^{g,ve}(\zeta) - \tau_R), \tag{116}$$

where

N is a normalizing constant,

A_i^{ve} is amplitude of the ith TM mode generated by the vertical component of the transmitting antenna,

ω_0 is the angular frequency of the wave,

$\tau_i^{p,ve}(\zeta)$ is the delay time of the phase of the wave when the aircraft is at angle ζ of its orbit for the ith waveguide mode, and

$\tau_i^{g,ve}(\zeta)$ is the group delay time of the wave when the aircraft is at angle ζ of its orbit for the ith waveguide mode.

The amplitude is given by

$$A_i^{ve} = S_i^{1.5} G_i(z)\, G_i(z_s) \Lambda_i \exp(i\pi/r + i\, k_0\, DS_i). \tag{117}$$

The phase delay time $\tau_i^{p,ve}(\zeta)$ is obtained by evaluating the phase of A_i:

$$\omega_0\, \tau_i^{p,ve}(\zeta) = \arg(A_i^{ve}). \tag{118}$$

The group delay time $\tau_i^{g,ve}(\zeta)$ is obtained by evaluating the derivative of the phase of A_i:

$$\tau_i^{g,ve}(\zeta) = \frac{\partial}{\partial\omega}[\arg(A_i^{ve})]. \tag{119}$$

The path length D from the transmitting antenna to the receiver location is a function of the orbit angle ζ and the orbit radius r according to the approximation

$$D = D_0 + r \sin\zeta, \tag{120}$$

where D_0 is the distance from the receiver to the center of the orbit. To calculate $\vec{R}_l^{ve}(\zeta,\tau_R)$, we performed summations as indicated in Eqs. (108) through (110).

We calculate the horizontally electrically generated fields in an analogous manner:

$$X_l^{(i)he}(\tau_R,\zeta) = N\,|A_i^{he}|\,\cos[\omega_0\tau_i^{p,he}(\zeta)]\; C_0\,(\tau_i^{g,he}(\zeta) - \tau_R) \tag{121}$$

and

$$Y_l^{(i)he}(\tau_R,\zeta) = N\,|A_i^{he}|\,\sin[\omega_0\tau_i^{p,he}(\zeta)]\; C_0\,(\tau_i^{g,he}(\zeta) - \tau_R). \tag{122}$$

The amplitude of the i th mode is given as follows:

$$A_i^{he} = G_i(z)G_i(z_s)\Lambda_i \exp(i\pi/4 + ik_0\, D\; S_i)\; S_i^{.5}\; \Delta_i(z_s) = A_i^{ve}\; S_i\; \Delta_i(z_0). \tag{123}$$

The phase and group delay times are given in the same manner as for the vertically generated modes:

$$\omega_0\, \tau_i^{p,he}(\zeta) = \arg(A_i^{he}) \tag{124}$$

and

$$\tau_i^{g,he}(\zeta) = \frac{\partial}{\partial\omega}[\arg(A_i^{he})]. \tag{125}$$

Tables 6 and 7 give the numerical values used in making the correlation calculations.

6.2.2 Results

Figures 92 through 97 show $\vec{R}_l^T(\tau_R)$ and $\phi_l^T(\tau_R)$ contours for γ values of 90°, 85°, 80°, 60°, 45°, and 0° with the propagation parameters of Table 6. Figure 98 shows $\vec{R}^{(1)ve}(\tau_R)$, the correlation vector of the first-order mode. Figures 99 through 104 show the correlation vector for propagation parameters of Table 7 for γ values of 90°, 85°, 80°, 60°, 45°, and 0°. The "average" chip duration for each of these calculations was $\tau_c = 1$ ms. The total length of each chip is $T = 2\tau_c = 2$ ms. The radius of the orbit is 3 km; the distance from the center of the orbit to the receiver is 4000 km.

6.2.3 Discussion

Figure 92 shows the correlator performance for the case of a completely horizontal antenna. For this case $|\vec{R}_l(\tau_R)|$ goes to zero at $\zeta = 90°$ and 270°, and there are sudden 180° phase flips at these angles. The best correlation time appears to be at $\tau_R = 13.5$ ms, which is the middle of the correlation peak.

Figure 93 shows the correlator performance for the case of an antenna with inclination $\gamma = 85°$ from the vertical. For this case it is difficult to say if there is a single best receiver synchronization time τ_R. During most of the orbit it appears that the time $\tau_R = 13.5$ ms would be the best. However, this value of τ_R appears to give a value of $|\vec{R}_l^T(\tau_R)|$ that is severely reduced at $\zeta = 90°$ and $\zeta = 270°$. For this case a system of selection of three values of τ_R separated by 1-ms intervals (12.5, 13.5, and 14.5 ms) would give a more stable average performance in the evaluation of $\vec{R}_l^T(\tau_R)$. The discontinuity in Fig. 93d at $\tau_R = 13.8$ ms, signified by the heavy vertical line, is due to a 360° ambiguity which has no real significance except at the nulls ($\zeta = 120°$ and $\zeta = 240°$), where $|\vec{R}_l^T(\tau_R)| \to 0$. Comparison of Figs. 93c and 93d illustrate this feature clearly.

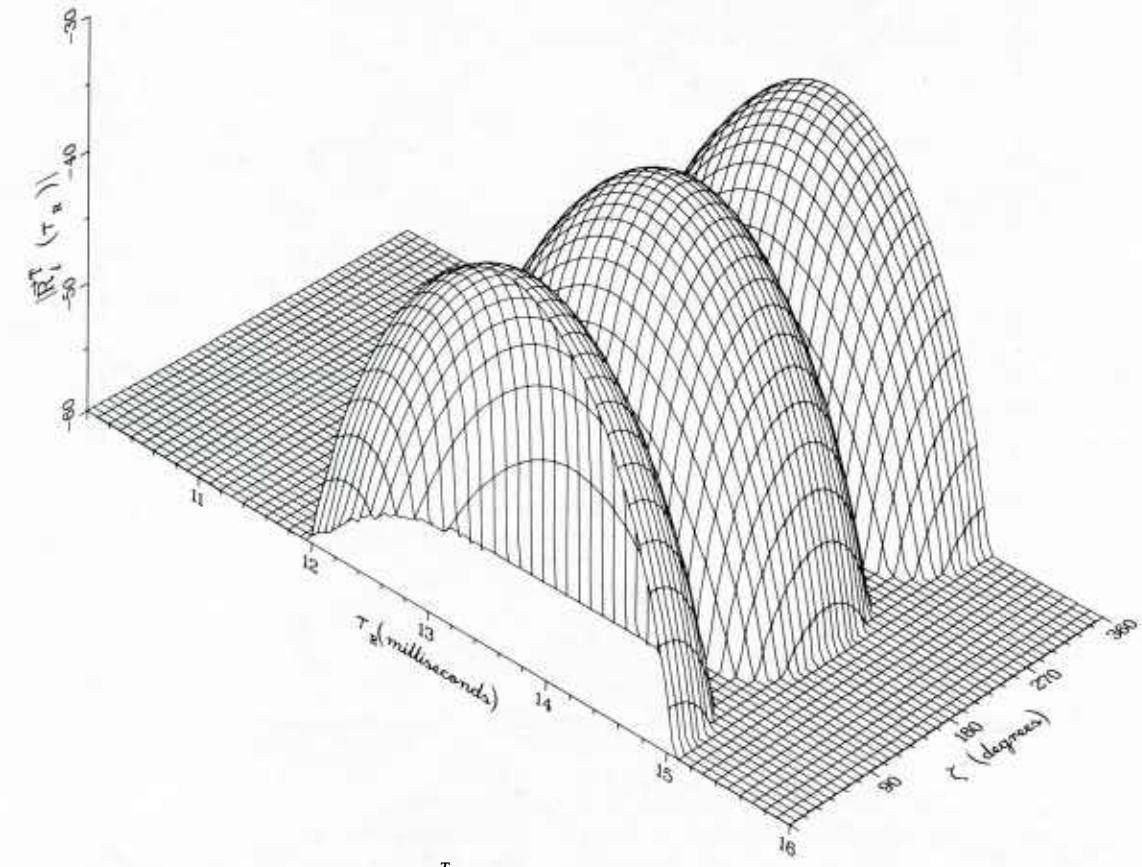


Fig. 92a: Amplitude behavior of $\bar{R}_l^T(\tau_R)$ for $\zeta = 90°$, calculated with the parameters of Table 6

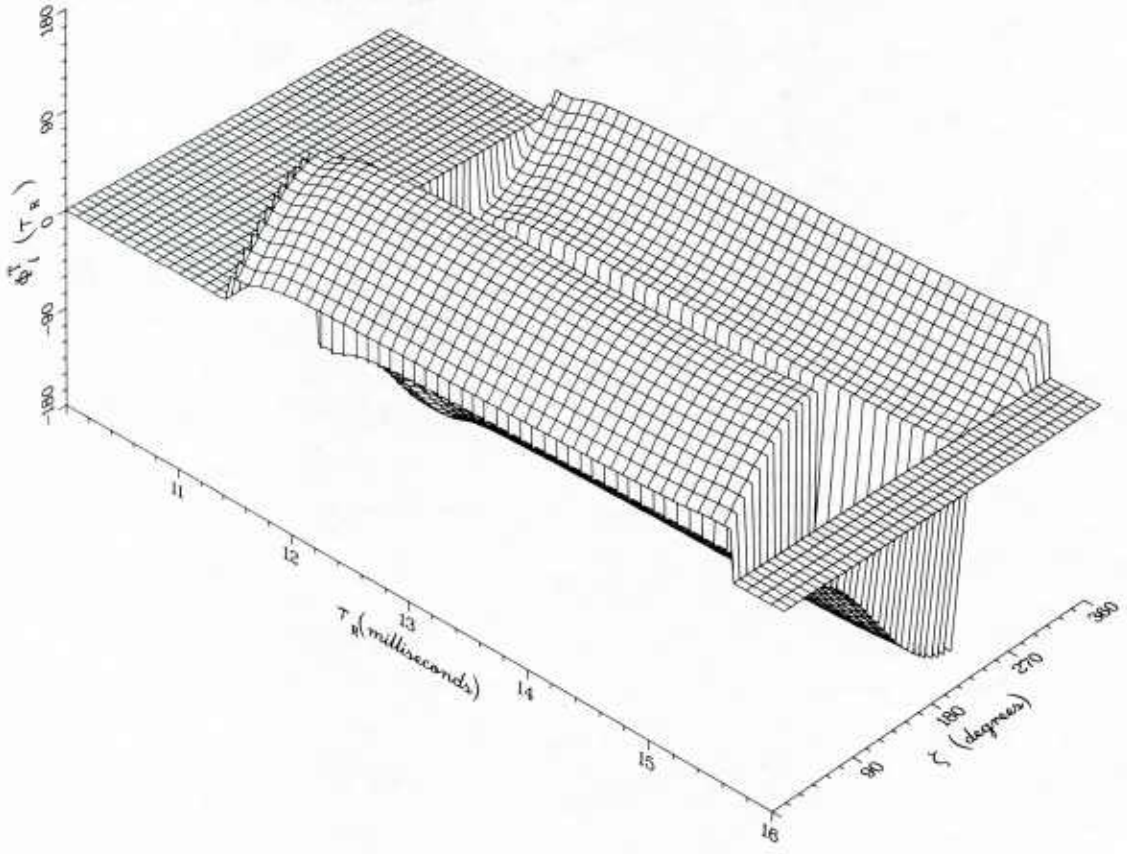


Fig. 92b: Phase behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 90°$ calculate with the parameters of Table 6

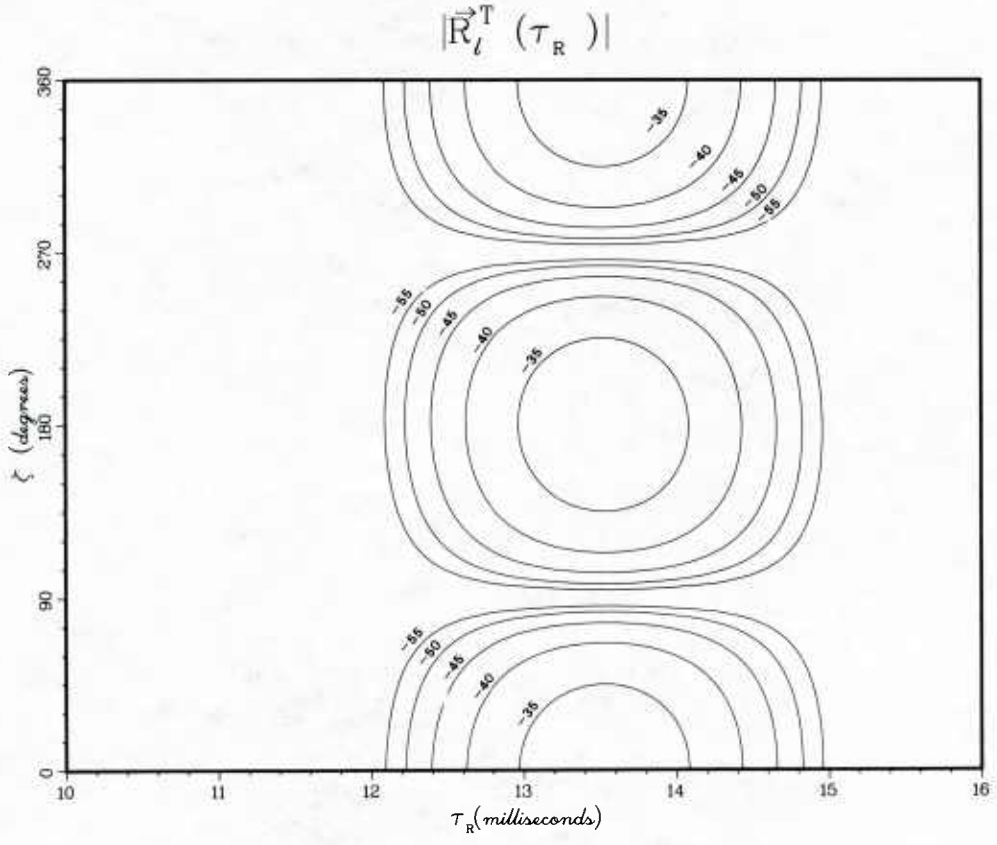


Fig. 92c: Amplitude contours corresponding to Fig. 92a

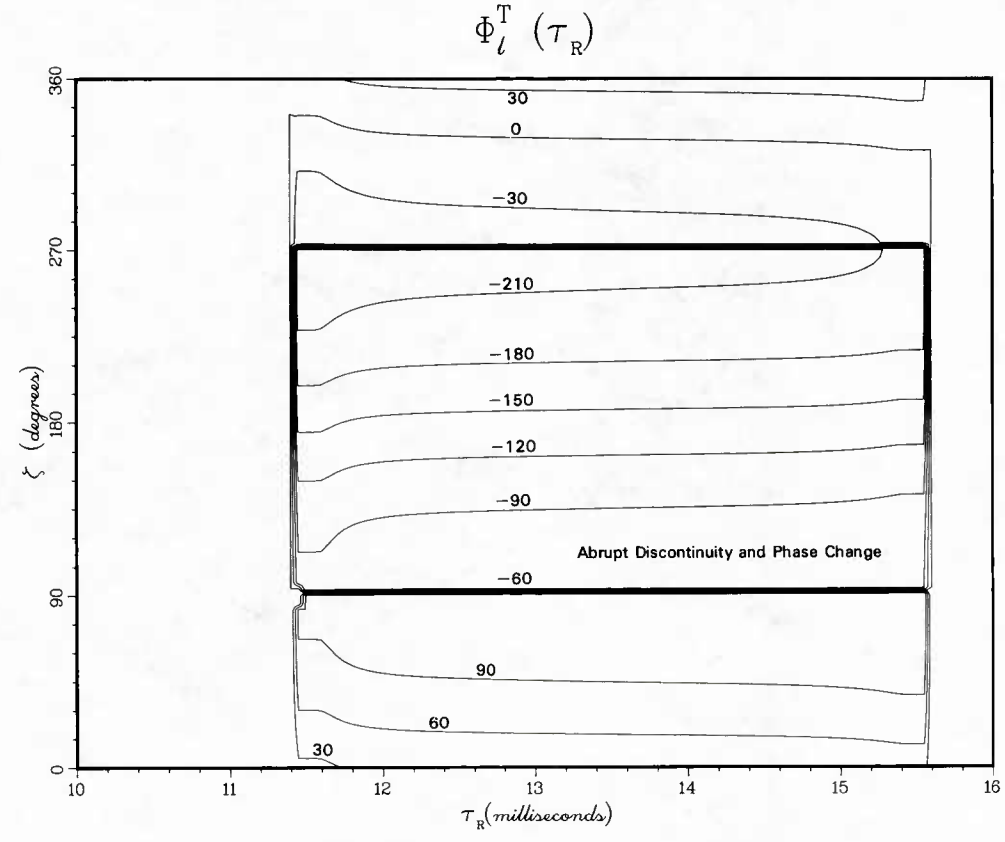


Fig. 92d: Phase behavior corresponding to Fig. 92b

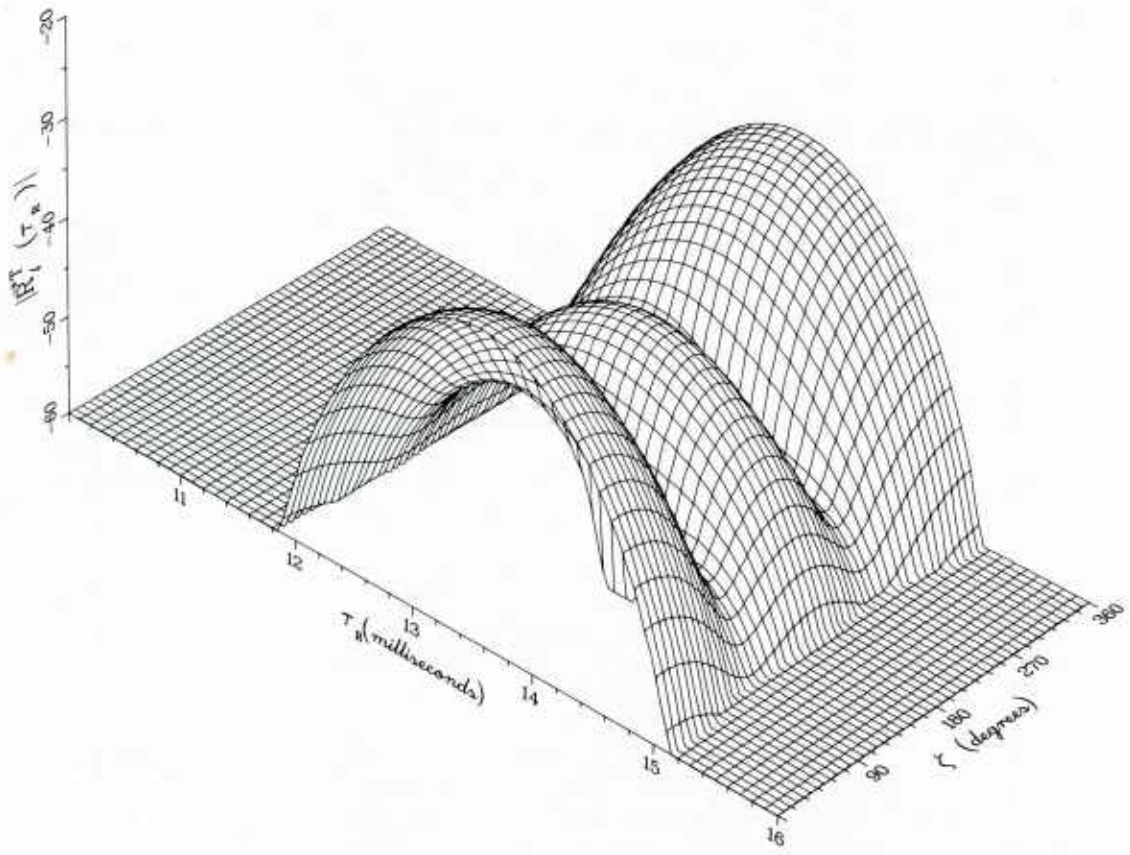


Fig. 93a: Amplitude behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 85°$, calculated with the parameters of Table 6

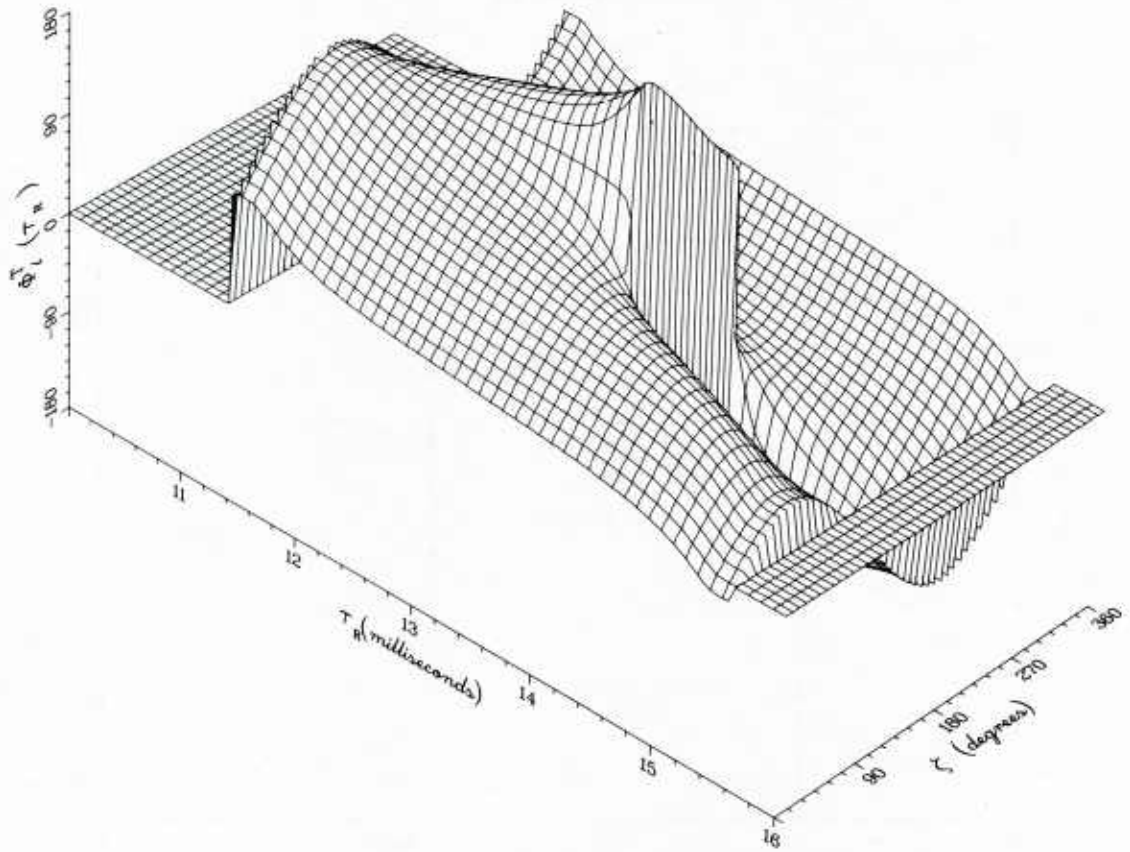


Fig. 93b: Phase behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 85°$, calculated with the parameters of Table 6

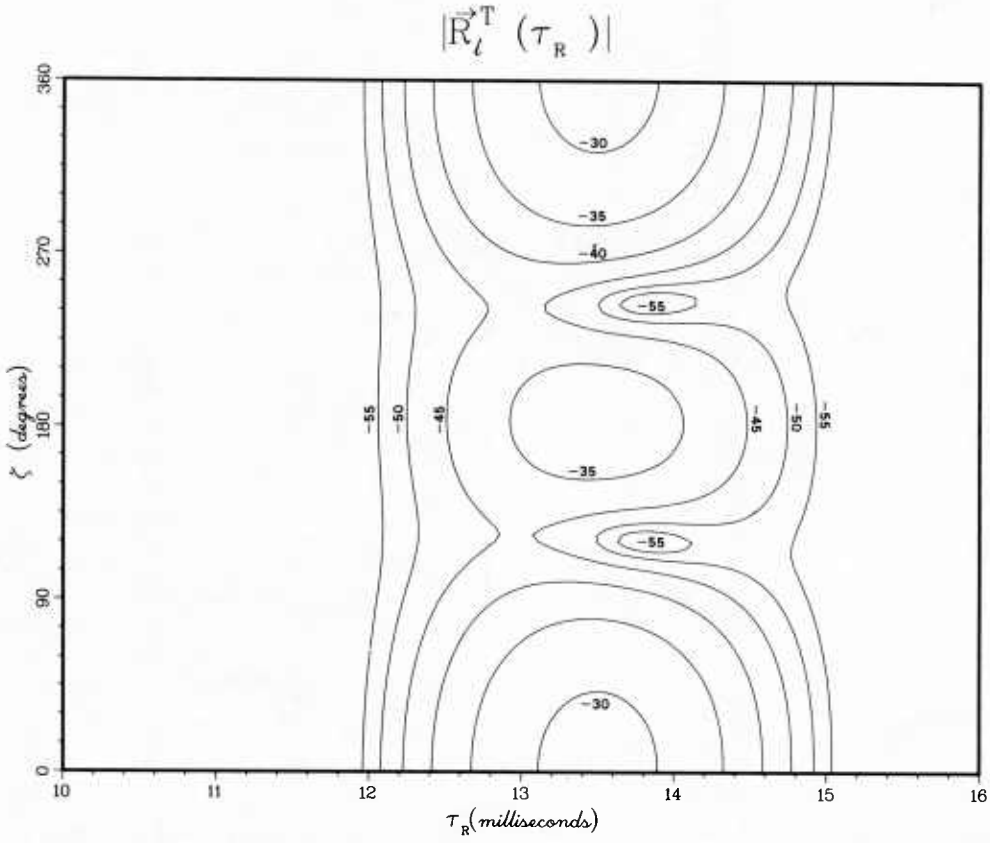


Fig. 93c: Amplitude contours corresponding to Fig. 93a

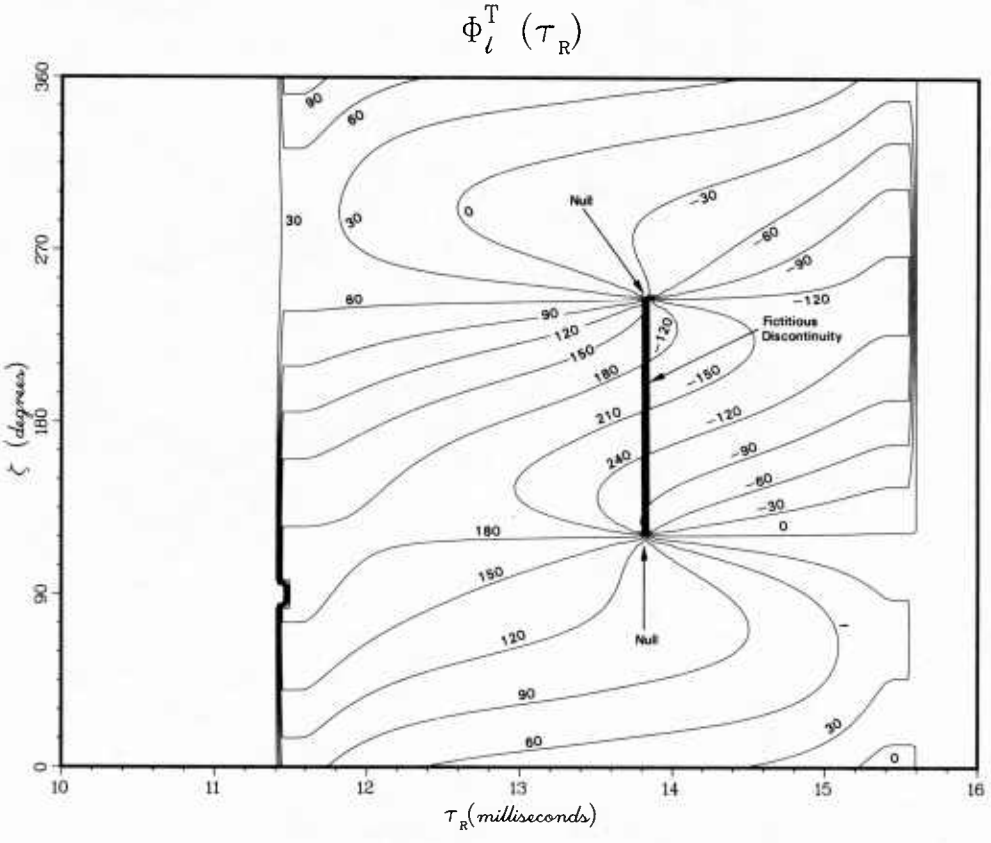


Fig. 93d: Phase behavior corresponding to Fig. 93b

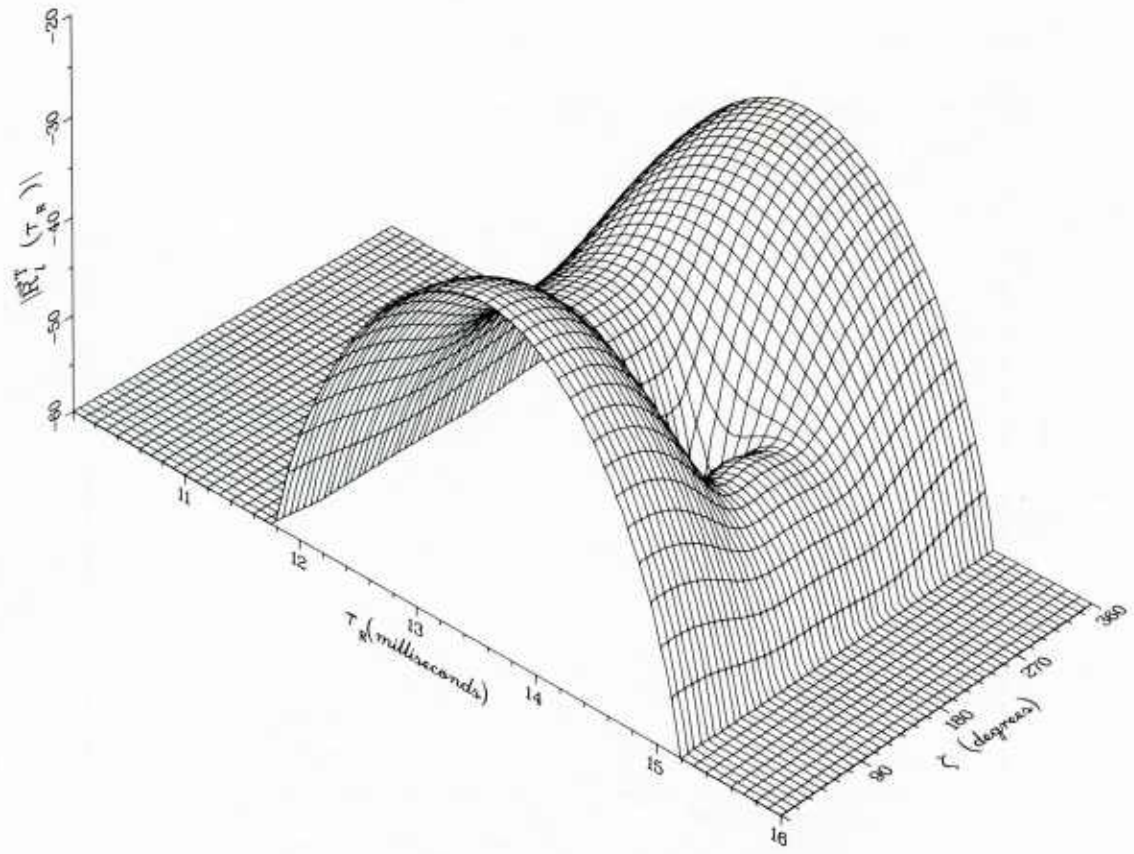


Fig. 94a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 80°$, calculated with the parameters of Table 6

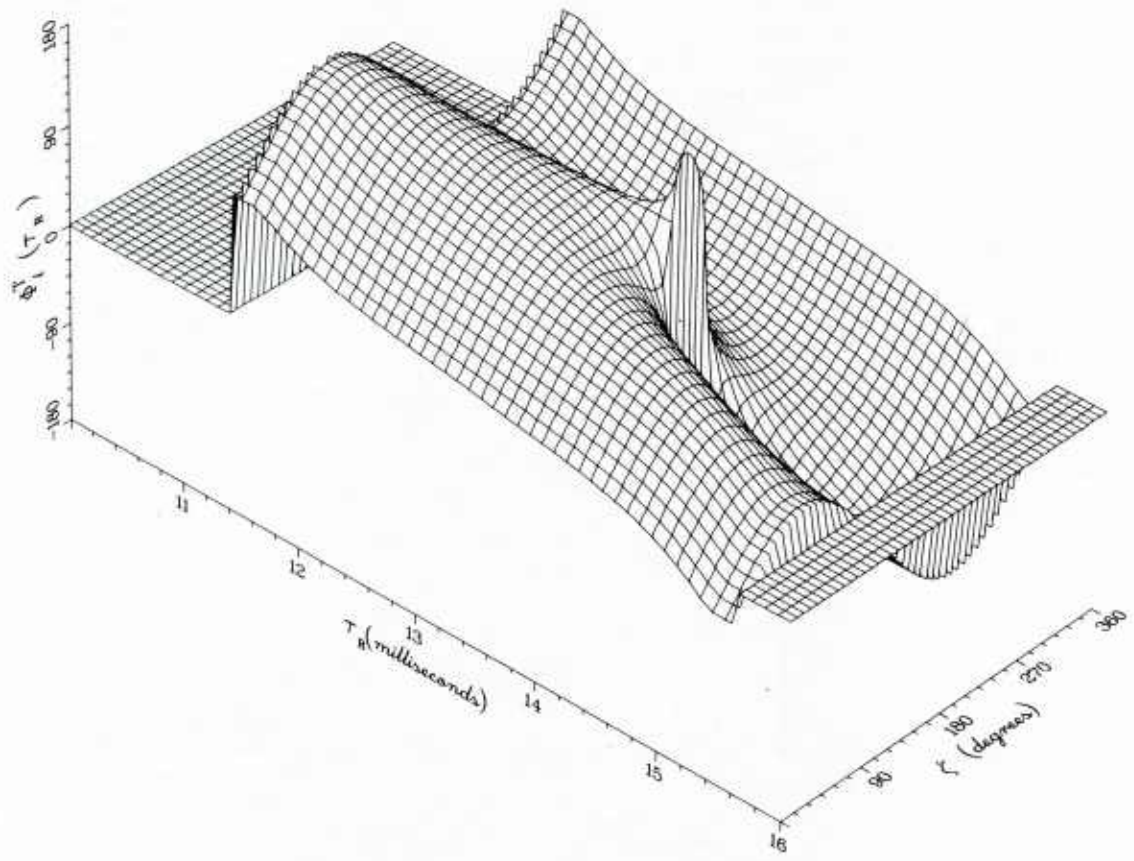


Fig. 94b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 80°$, calculated with the parameters of Table 6

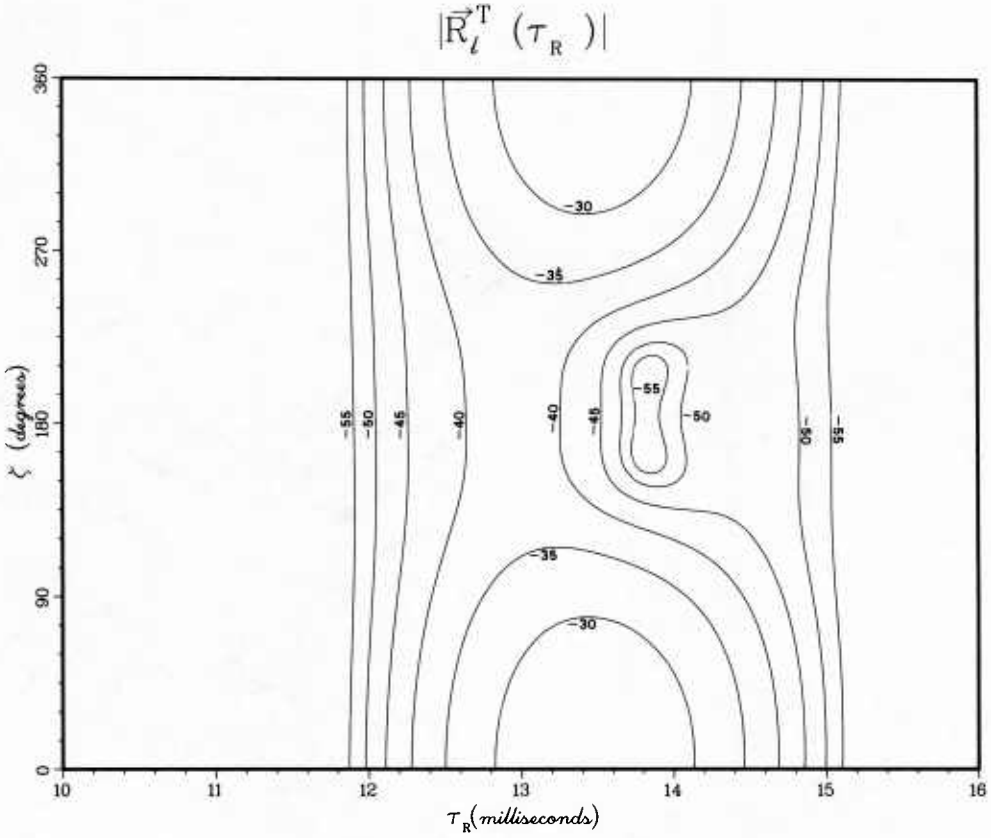


Fig. 94c: Amplitude contours corresponding to Fig. 94a

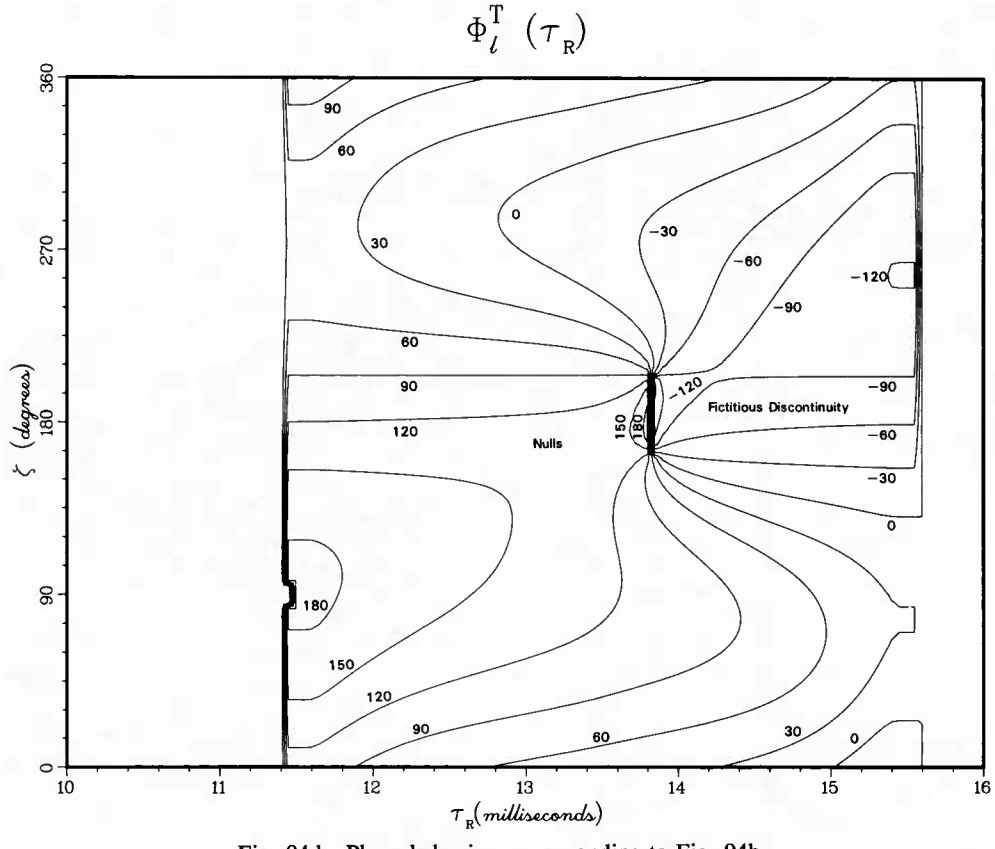


Fig. 94d: Phase behavior corresponding to Fig. 94b

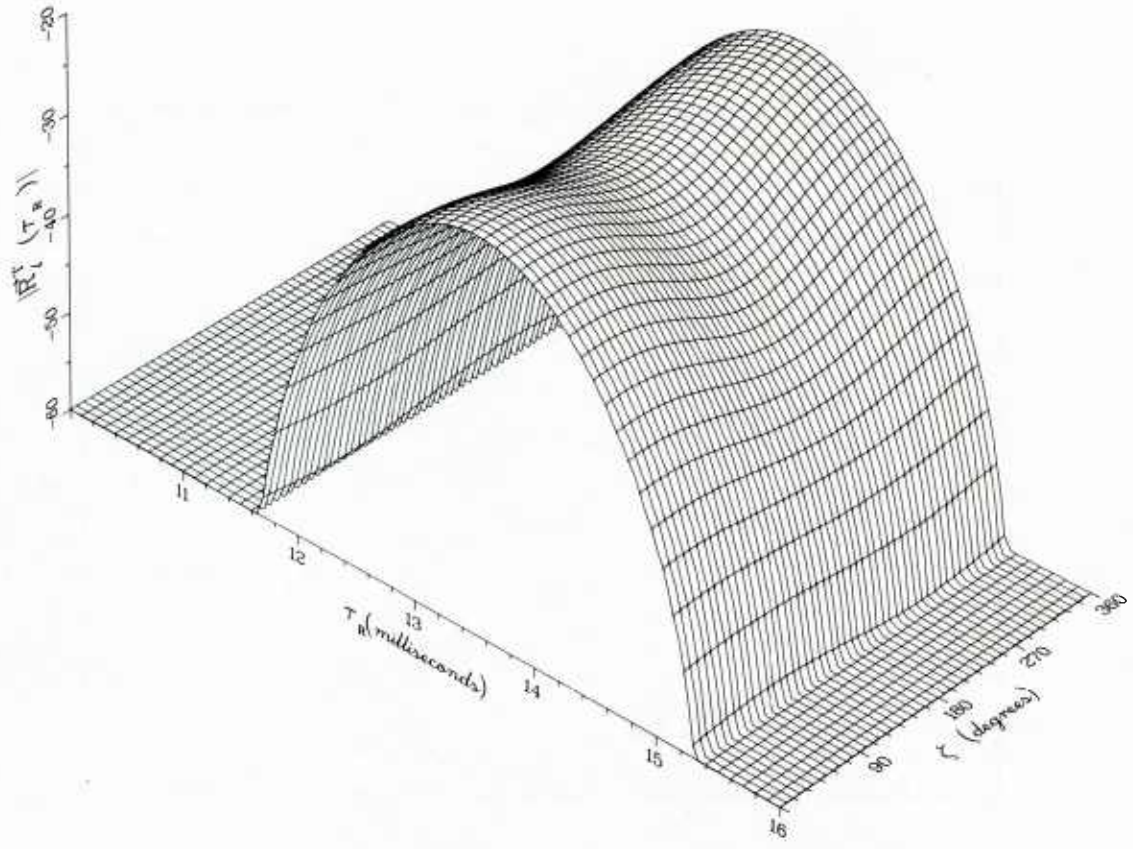


Fig. 95a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 60°$, calculated with the parameters of Table 6

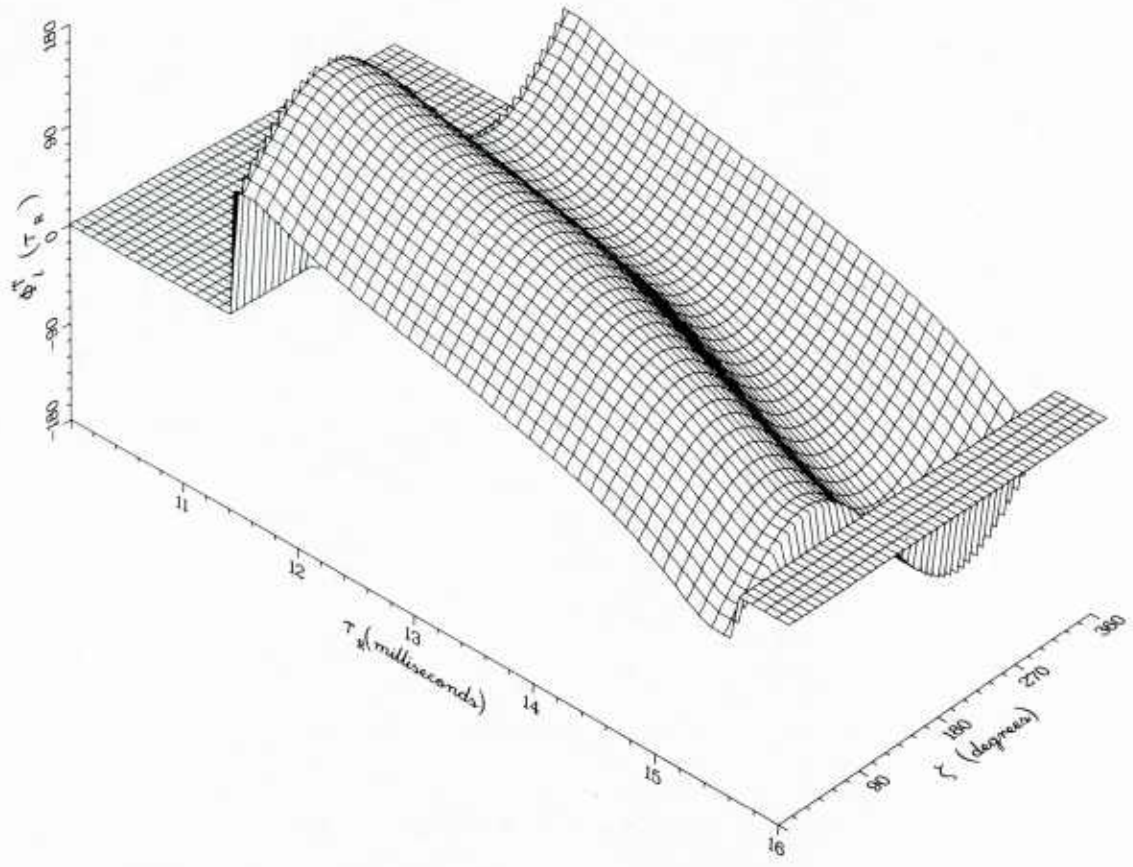


Fig. 95b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 60°$, calculated with the parameters of Table 6

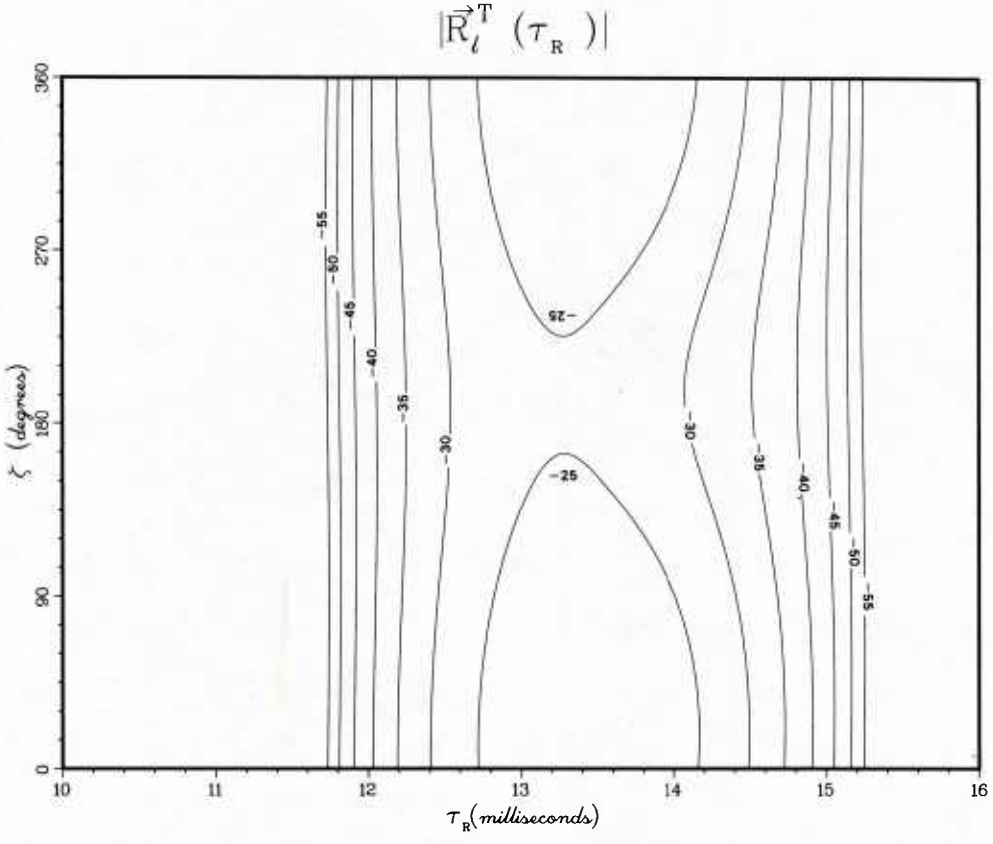


Fig. 95c: Amplitude contours corresponding to Fig. 95a

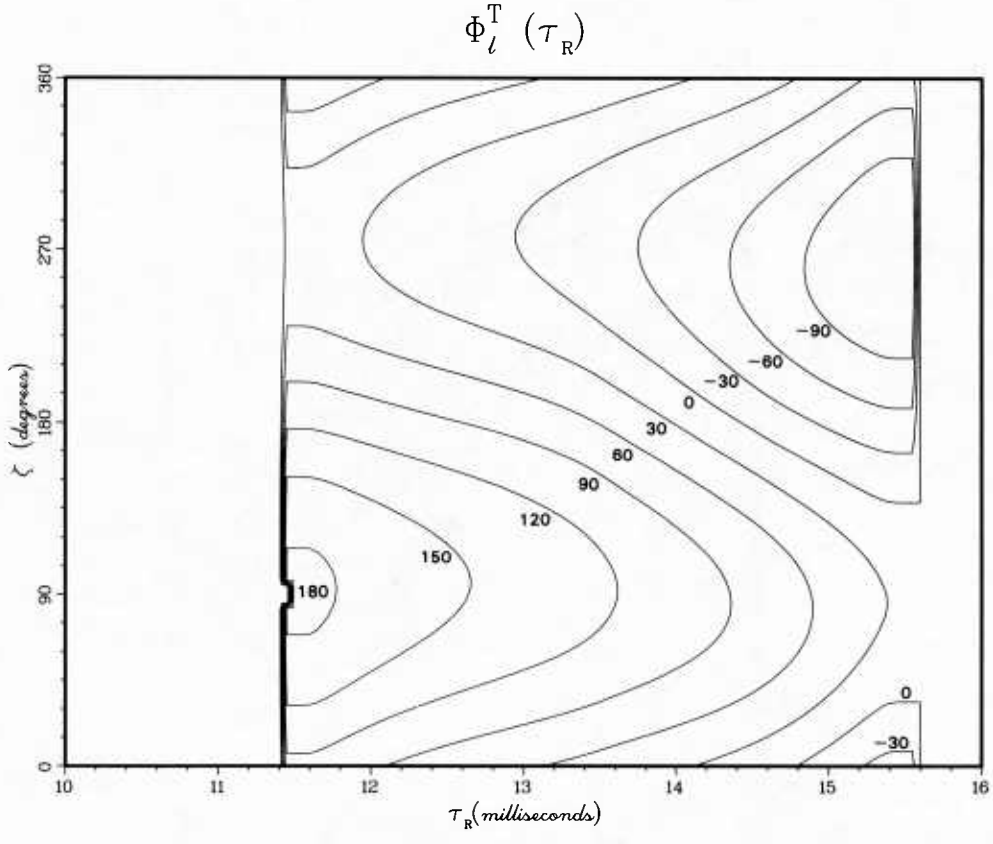


Fig. 95d: Phase behavior corresponding to Fig. 95b

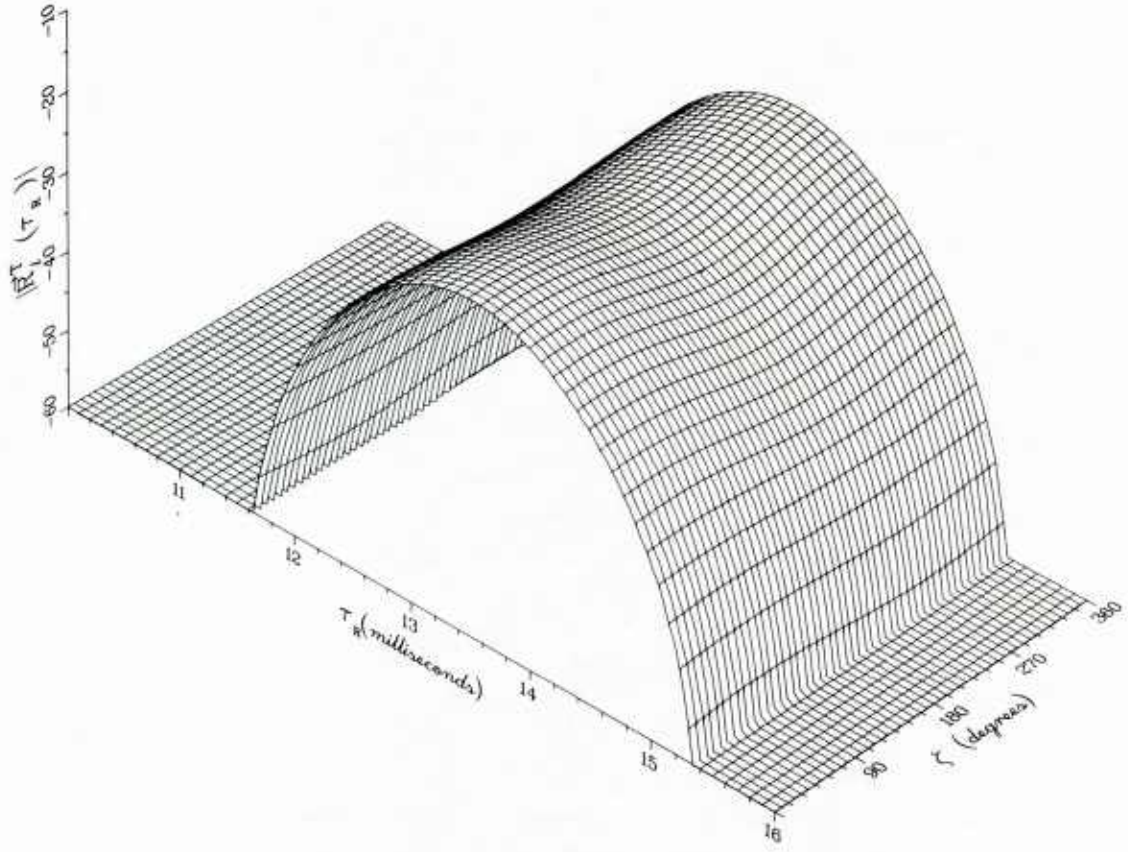


Fig. 96a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 45°$, calculated with the parameters of Table 6

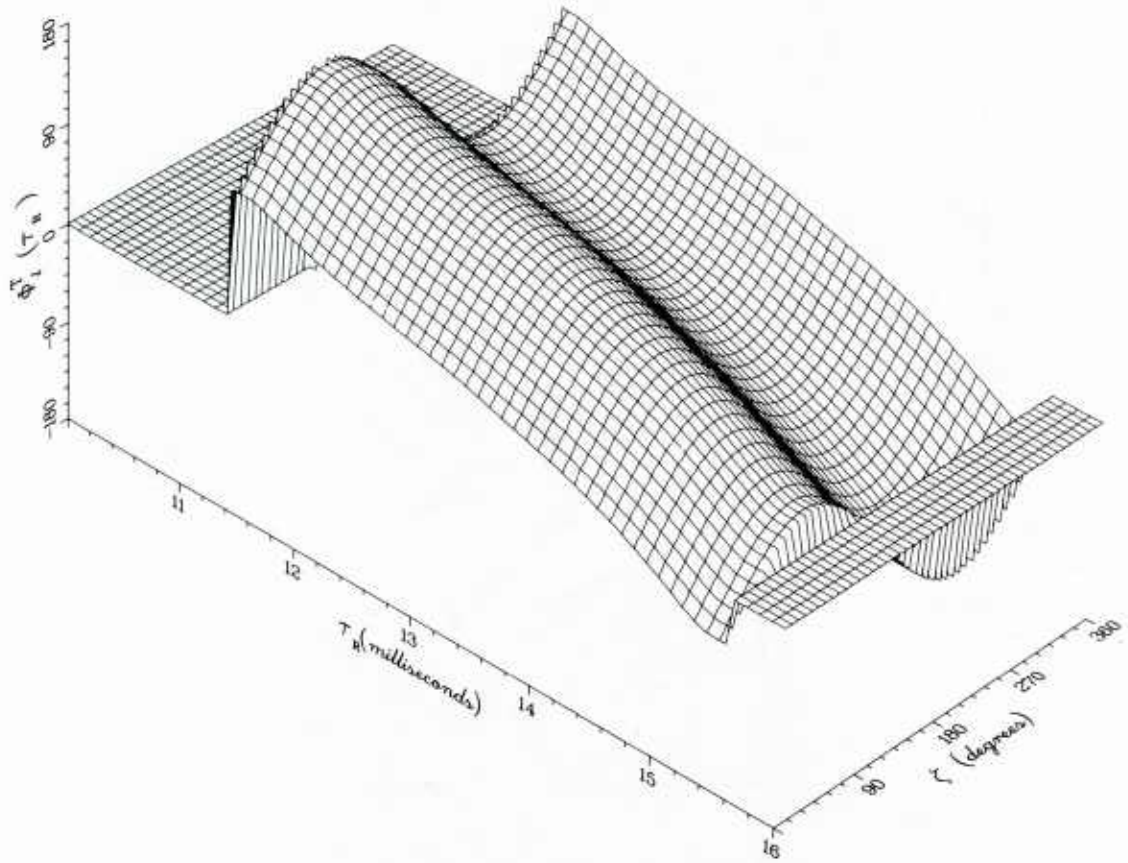


Fig. 96b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 45°$, calculated with the parameters of Table 6

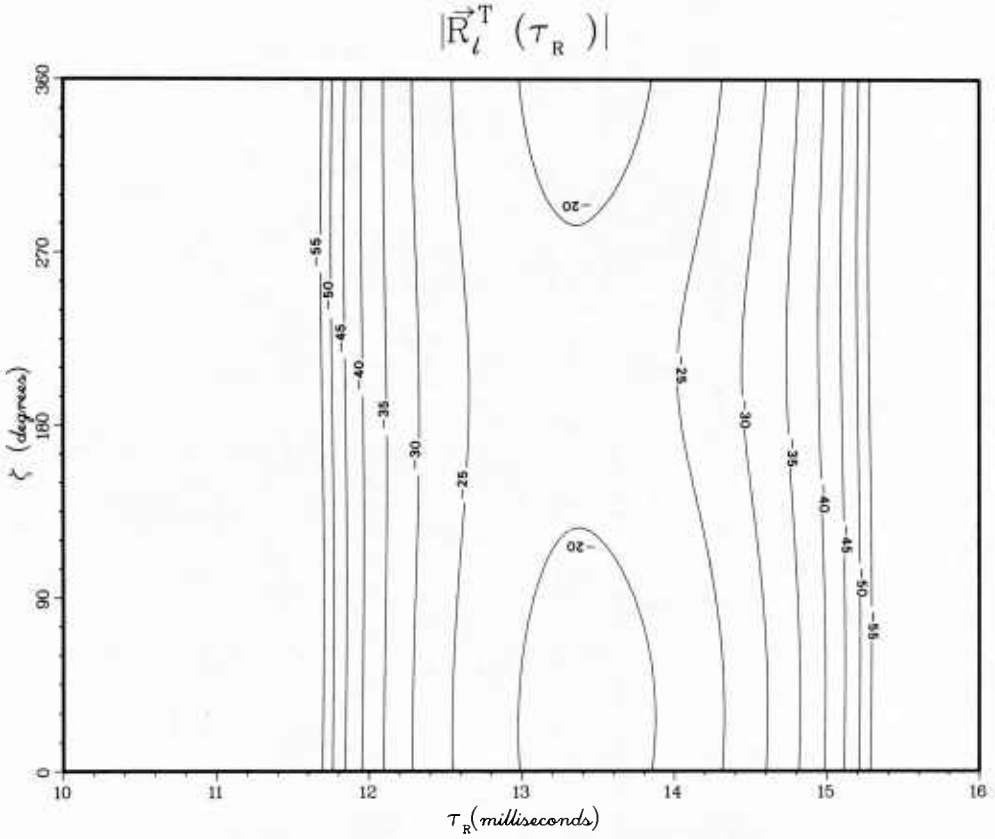


Fig. 96c: Amplitude contours corresponding to Fig. 96a

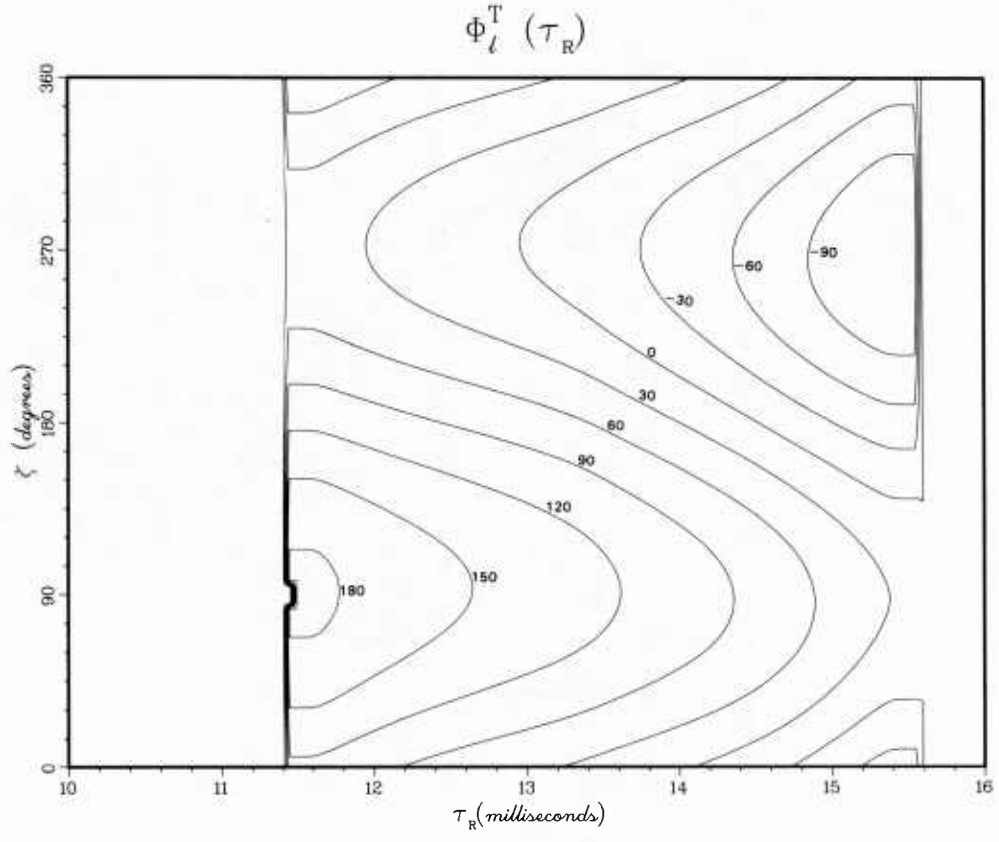


Fig. 96d: Phase behavior corresponding to Fig. 96a

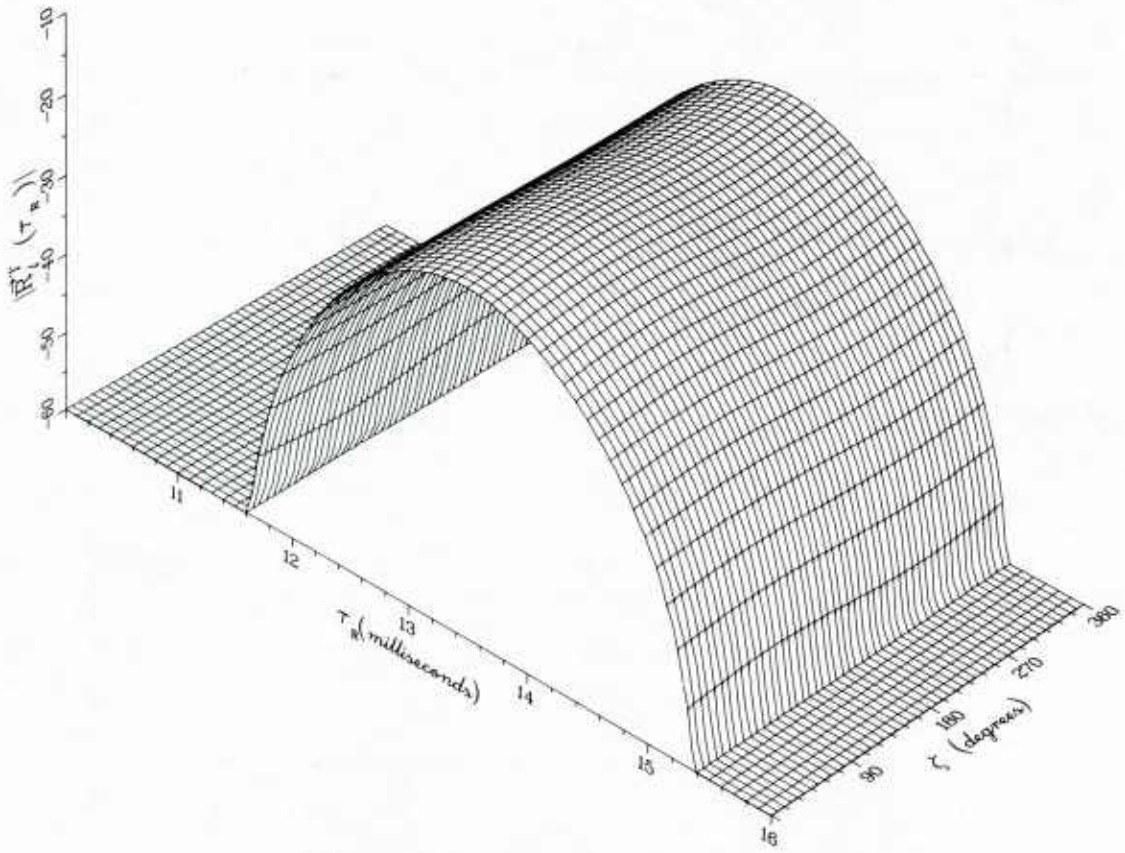


Fig. 97a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 0°$, calculated with the parameters of Table 6

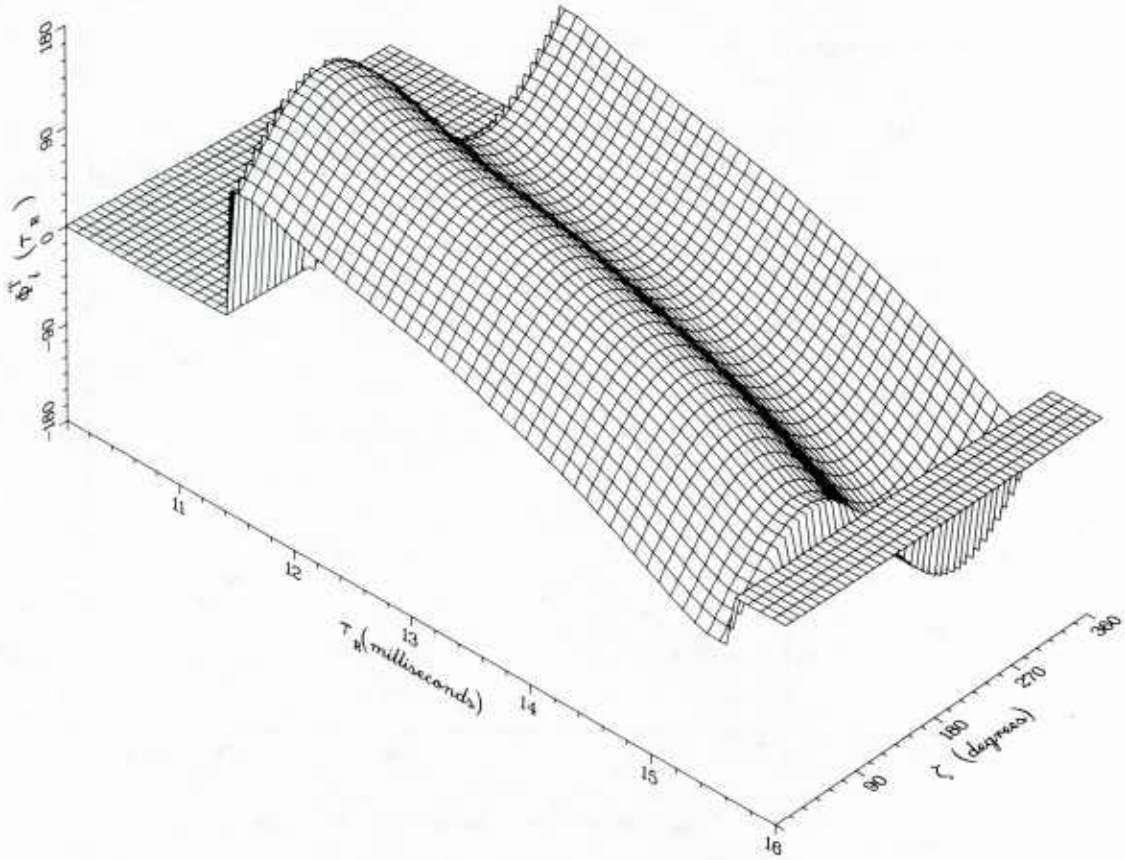


Fig. 97b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 0°$, calculated with the parameters of Table 6

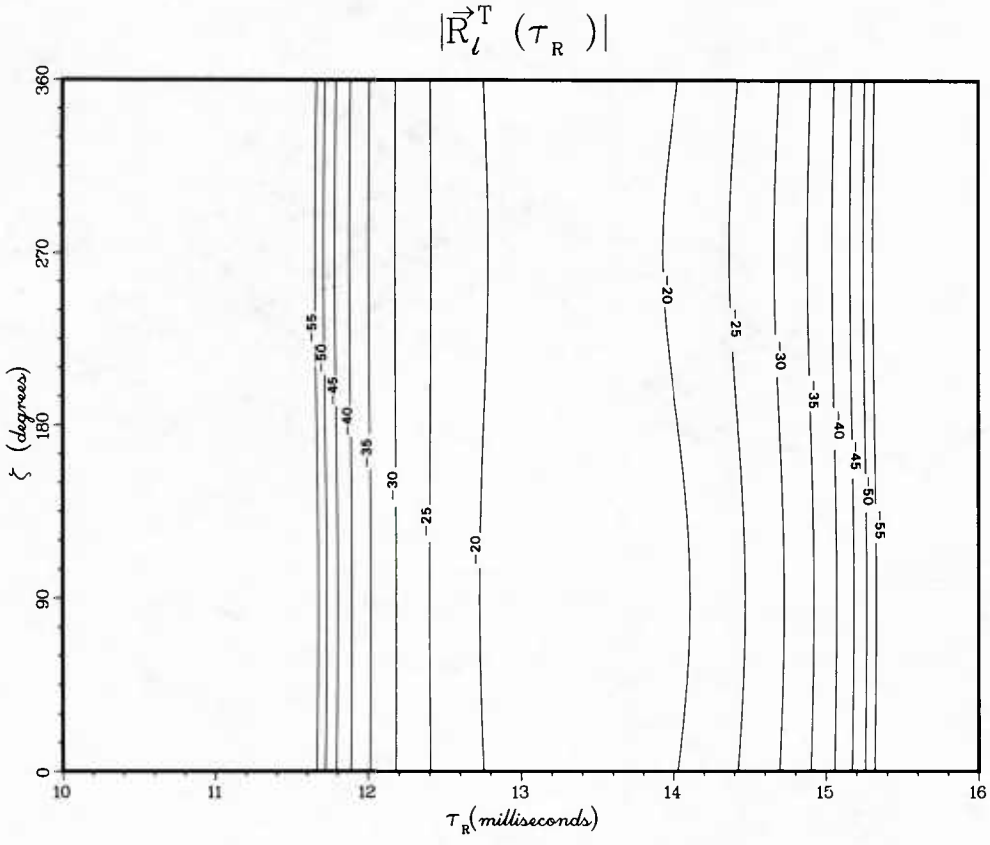


Fig. 97c: Amplitude contours corresponding to Fig. 97a

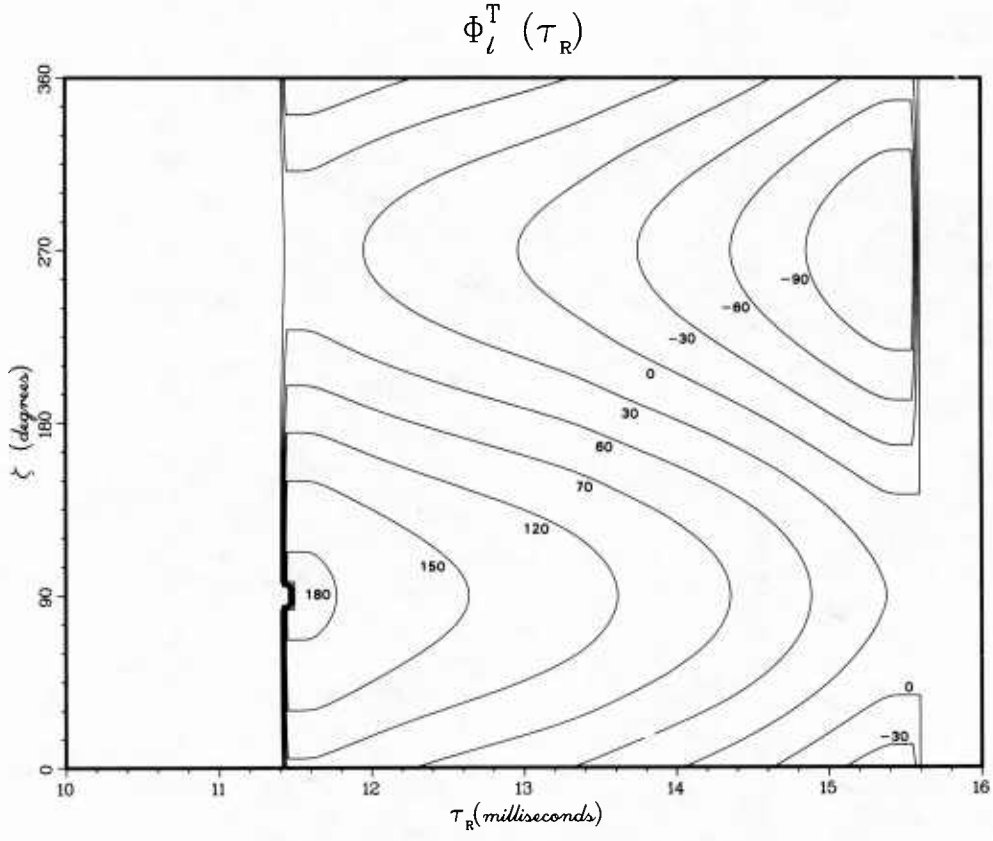


Fig. 97d: Phase behavior corresponding to Fig. 97b

Table 6 — Set of Quantities Used in Calculation of Correlation Vectors

Mode	Quantity	Value
1	$\lvert A_1^{ve} \rvert$	$1.26500262 \times 10^{-1}$
	$\tau_1^{p,ve}$	$1.33658211 \times 10^{-2}$
	$\tau_1^{g,ve}$	1.34400×10^{-2}
	$\lvert A_1^{he} \rvert$	$2.52495471 \times 10^{-3}$
	$\tau_1^{p,he}$	$1.33533157 \times 10^{-2}$
	$\tau_1^{g,he}$	1.34400×10^{-2}
2	$\lvert A_2^{ve} \rvert$	$1.11773228 \times 10^{-1}$
	$\tau_2^{p,ve}$	$1.32437617 \times 10^{-2}$
	$\tau_2^{g,ve}$	1.356666×10^{-2}
	$\lvert A_2^{he} \rvert$	$5.62805237 \times 10^{-3}$
	$\tau_2^{p,he}$	$1.32562578 \times 10^{-2}$
	$\tau_2^{g,he}$	1.356666×10^{-2}

Table 7 — Second Set of Quantities Used in Calculation of Correlation Vectors

Mode	Quantity	Value
1	$\lvert A_1^{ve} \rvert$	$1.26500262 \times 10^{-1}$
	$\tau_1^{p,ve}$	$1.33658211 \times 10^{-2}$
	$\tau_1^{g,ve}$	1.34400×10^{-2}
	$\lvert A_1^{he} \rvert$	$2.52495471 \times 10^{-3}$
	$\tau_1^{p,he}$	1.3315820×10^{-2}
	$\tau_1^{g,he}$	1.34400×10^{-2}
2	$\lvert A_2^{ve} \rvert$	$1.26500262 \times 10^{-1}$
	$\tau_2^{p,ve}$	$1.33908211 \times 10^{-2}$
	$\tau_2^{g,ve}$	1.356666×10^{-2}
	$\lvert A_2^{he} \rvert$	$5.62805237 \times 10^{-3}$
	$\tau_2^{p,he}$	1.3365820×10^{-2}
	$\tau_2^{g,he}$	1.356666×10^{-2}

Figure 94 shows the correlation performance for an antenna with $\gamma = 80°$. In the vicinity of $\tau_R = 14.0$ ms and $\zeta = 180°$ there is an abrupt drop of signal and jump in phase. Otherwise the $\lvert \bar{R}_l(\tau_R) \rvert$ function seems fairly well behaved.

Figure 95 shows the behavior of $\bar{R}_l^T(\tau_R)$ for an antenna with $\gamma = 60°$. The amplitude and phase variation of the correlation are much milder than for the cases with smaller values of γ, since the vertical electric-field components exceed the horizontal components by it least one order of magnitude at this droop angle. This trend is continued as γ is decreased to 45° and then to 0° in Figs. 96 and 97.

Figure 98 gives the value of $\bar{R}_l^{(1)ve}(\tau_R)$. It illustrates the case of pure first-order-mode propagation from a vertical electric-dipole source.

The values of the propagation quantities in Table 7 are similar to those of Table 6 except for an adjustment to produce maximum mode cancellation and interference effects. We have made the first-order and second-order vertically generated modes have opposite phase and equal magnitude. We have modified the phase of the horizontally generated modes so they are in the same phase plane with the vertically generated modes to produce maximum field-strength variation. Hence both severe effects appear in Figs. 99 through 104.

Figure 99 gives the behavior of $\bar{R}_l^T(\tau_R)$ for the propagation parameters of Table 3. The results are almost identical to those of Fig. 92, since $\gamma = 90°$ and the only changes made in the parameters of the horizontally generated modes are to the phase delay times $\tau_1^{p,he}$ and $\tau_2^{p,he}$.

Figure 100 shows the behavior of $\bar{R}_l^T(\tau_R)$ for the case of $\gamma = 85°$. The influence of the splitting of the vertically generated function is more evident on the phase plot than on the amplitude plot.

Figure 101 shows the behavior of $\bar{R}_l^T(\tau_R)$ for the case of $\gamma = 80°$. Here, the influence of the correlation-peak splitting of the vertically generated function is quite evident as the notch in the correlation peak makes its way from side to side through the values of τ_R and ζ. This is illustrated further in Figs. 102 and 103, which give the graphs for $\gamma = 60°$ and 45°.

Finally, Fig. 104 shows the split correlation peak for a purely vertical antenna being carried around an orbit. Here we find a nearly constant variation of $\lvert \bar{R}_l^T(\tau_r) \rvert$ versus ζ, and the phase of the correlation function $\phi_l^T(\tau_R)$ varies only moderately with ζ. As a function of τ_R, these variables show the expected splitting.

7.0 DISTURBANCES

As mentioned previously, long wavelength waves require an ionospheric reflection medium to provide for their world ranging capabilities. This reflecting ionospheric medium is caused by the natural interaction of the sun and the stars with the earth's atmosphere, mediated by the interplanetary medium and by the geomagnetic field of the earth. In total this constitutes an extremely complex system which is only poorly understood and monitored. An additional complication arises in the consideration of the ground—the lower boundary of the waveguide. The effective conductivity of the ground surface can vary over great ranges. The conductivity is controlled by the composition and layering of the earth, by the water content of the soil and rocks, by their roughness, and by their temperatures. One may consider the variation of ground conductivity with location as a disturbance from the placid intellectual picture of a homogeneous propagation path. In addition, the seasonal wetting and warming of vast areas of the earth should be of consequence to radio propagation forecasts. Further, the lower atmosphere contains elements which slightly retard the propagation of radio waves, form ducting layers and account for the famous "four thirds earth radius" law which becomes applicable even at low frequencies. So, the weather is an additional complicating factor.

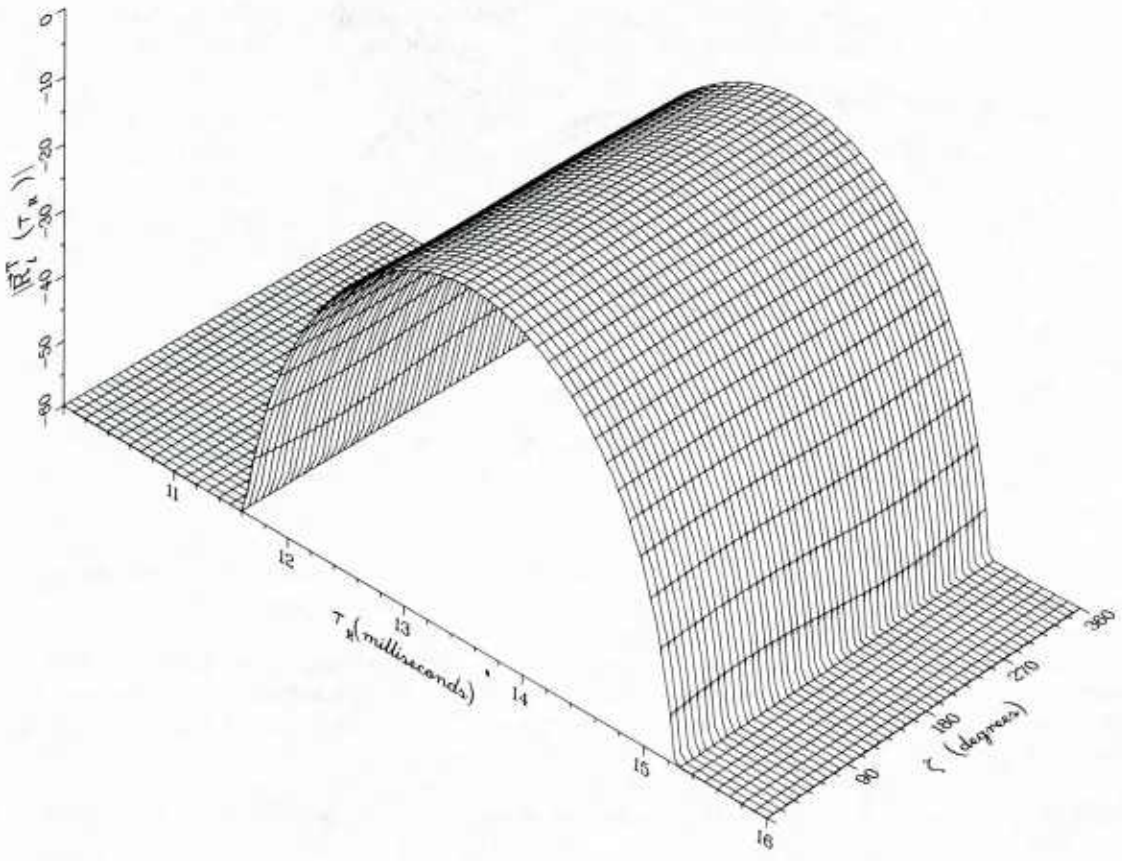


Fig. 98a: Amplitude behavior of $\vec{R}_l^{(1)}(\tau_R)$ for $\gamma = 0°$, calculated with the parameters of Table 6

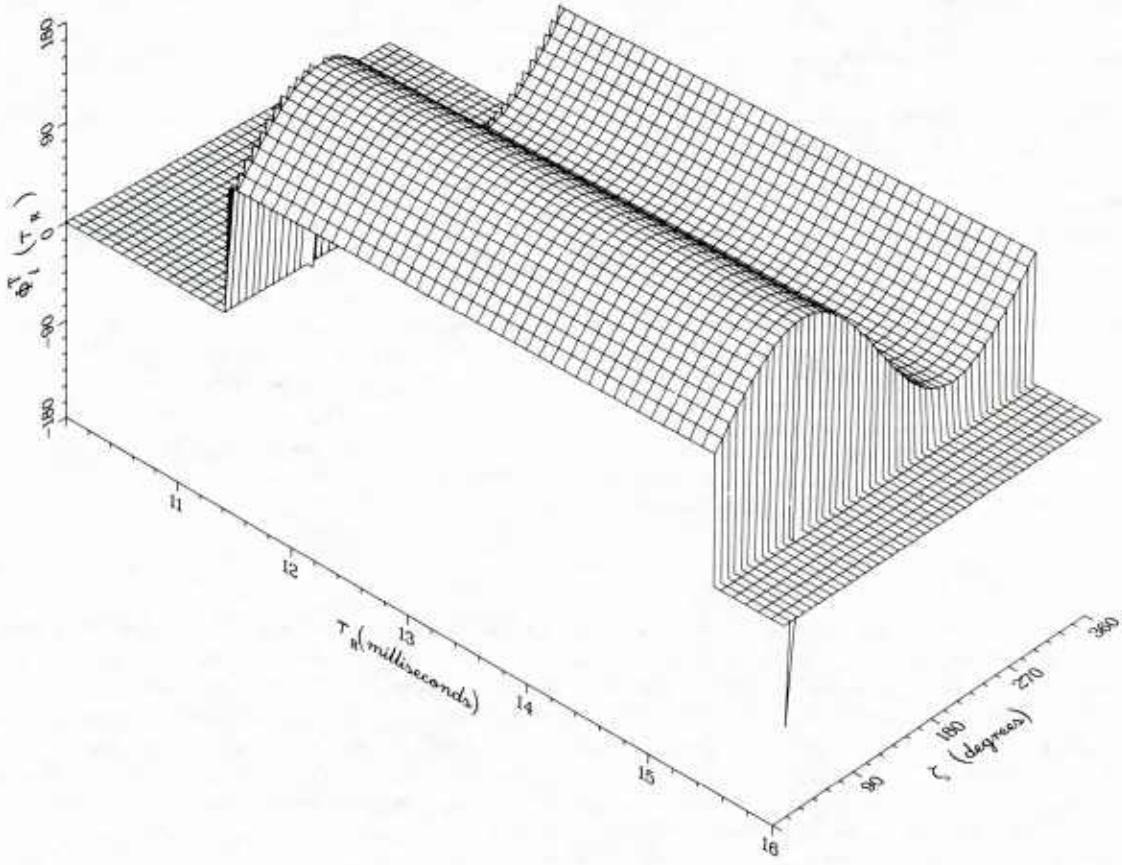


Fig. 98b: Phase behavior of $\vec{R}_l^{(1)}(\tau_R)$ for $\gamma = 0°$, calculated with the parameters of Table 6

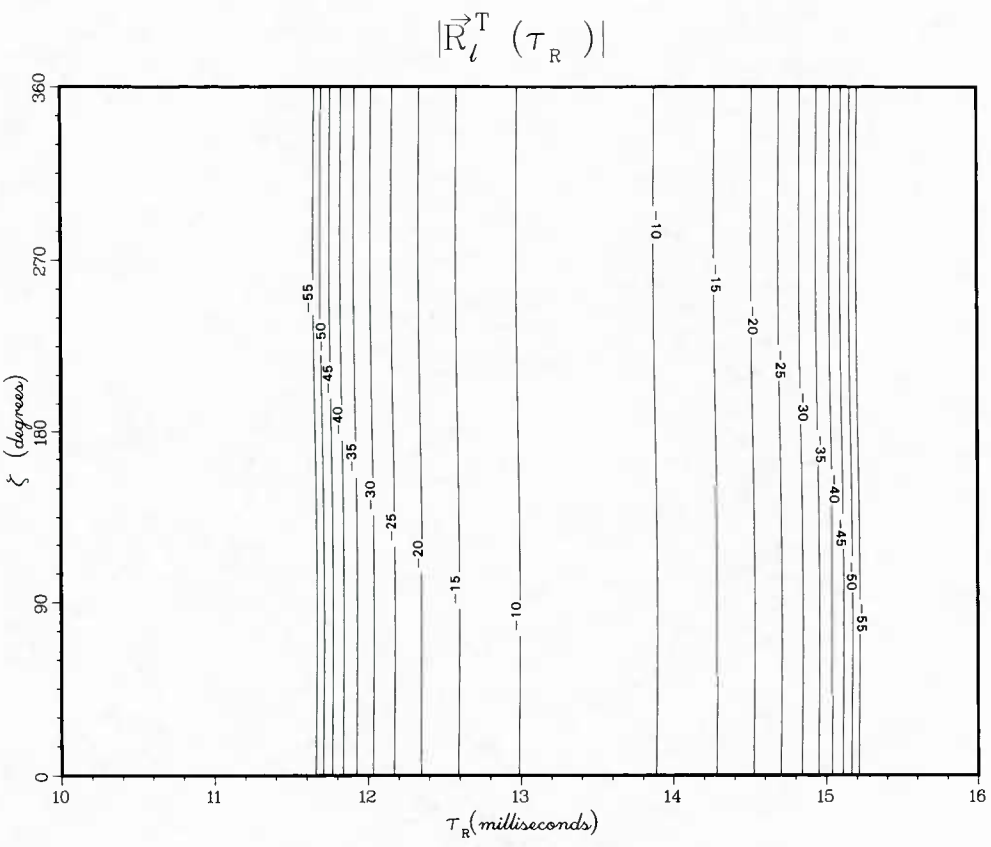


Fig. 98c: Amplitude contours corresponding to Fig. 98a

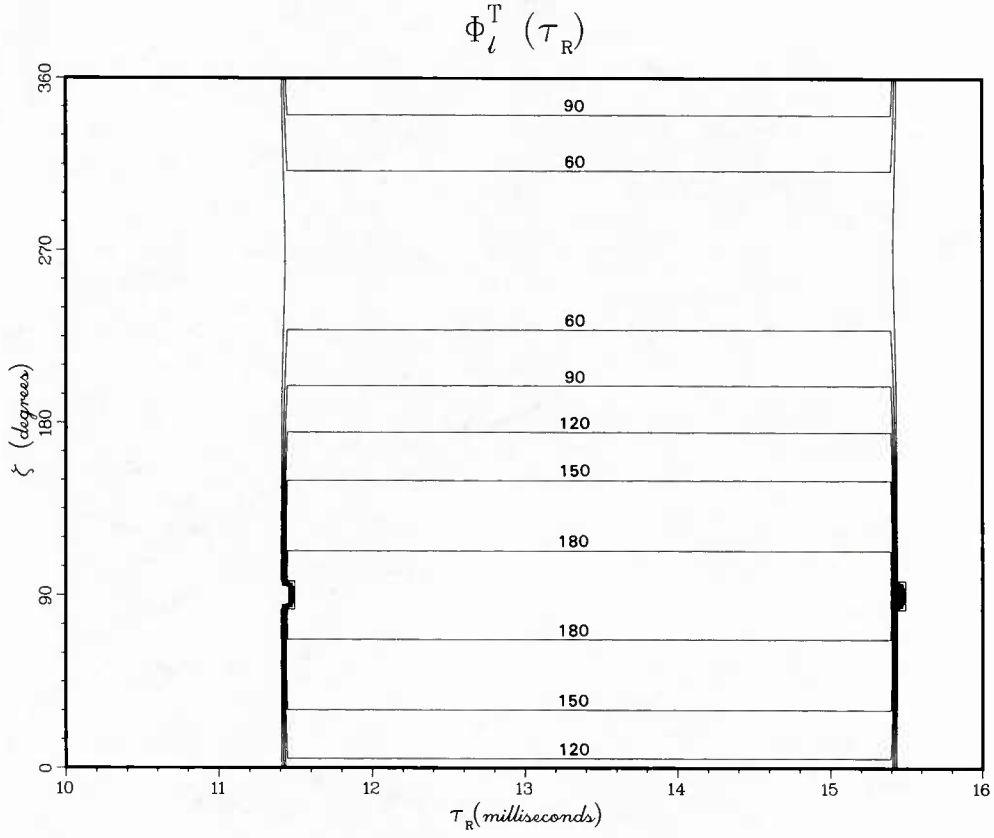


Fig. 98d: Phase behavior corresponding to Fig. 98b

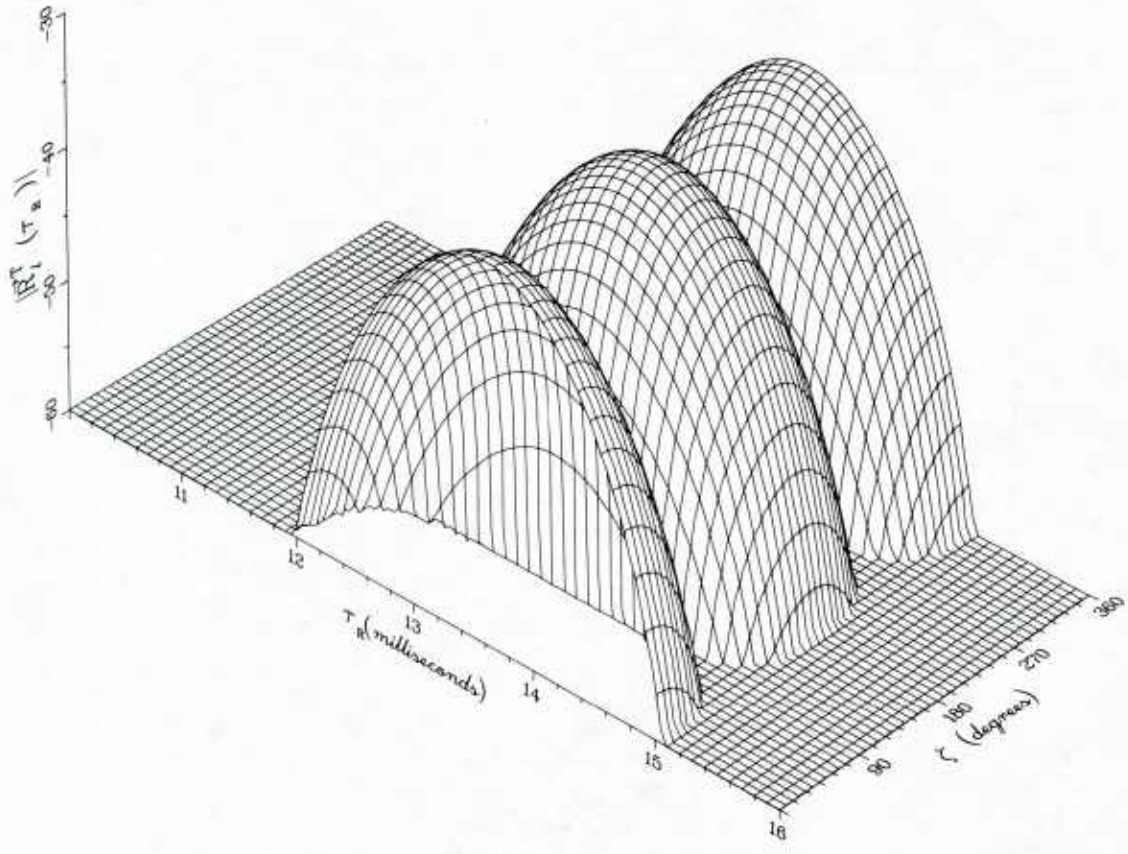


Fig. 99a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 90°$, calculated with the parameters of Table 7

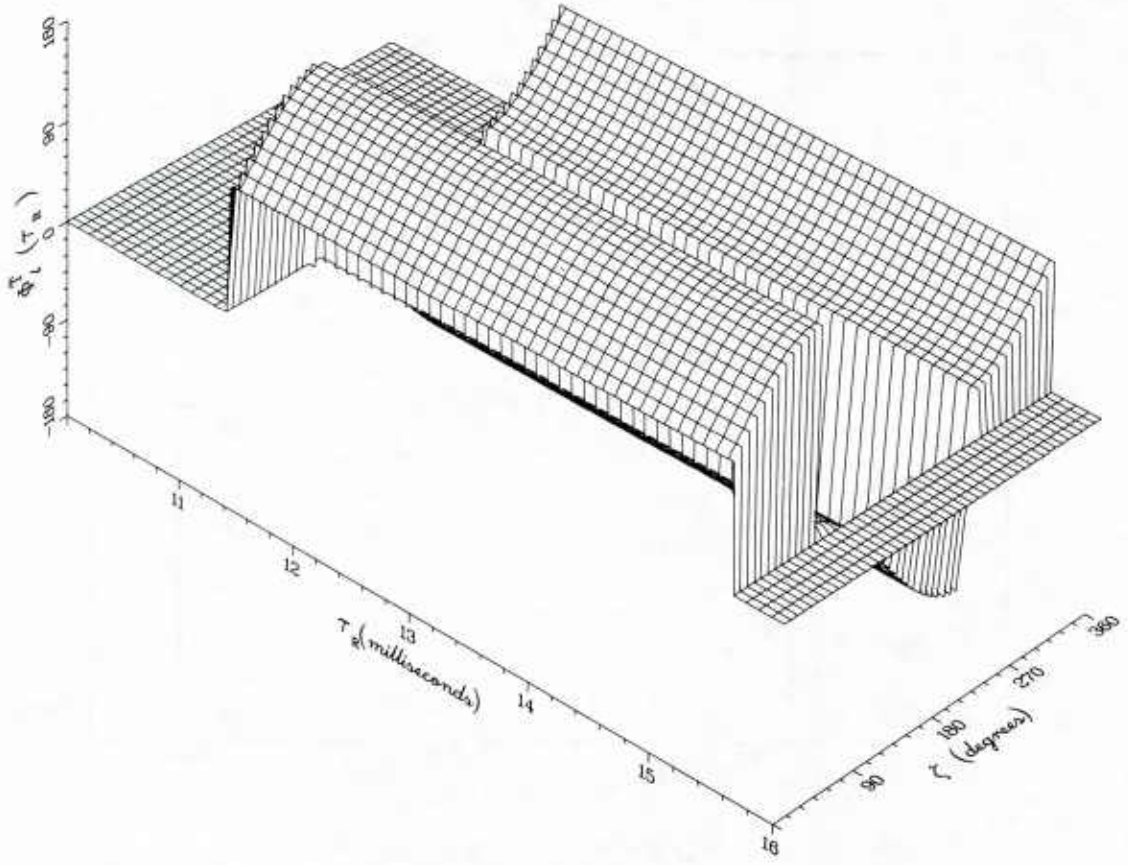


Fig. 99b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 90°$, calculated with the parameters of Table 7

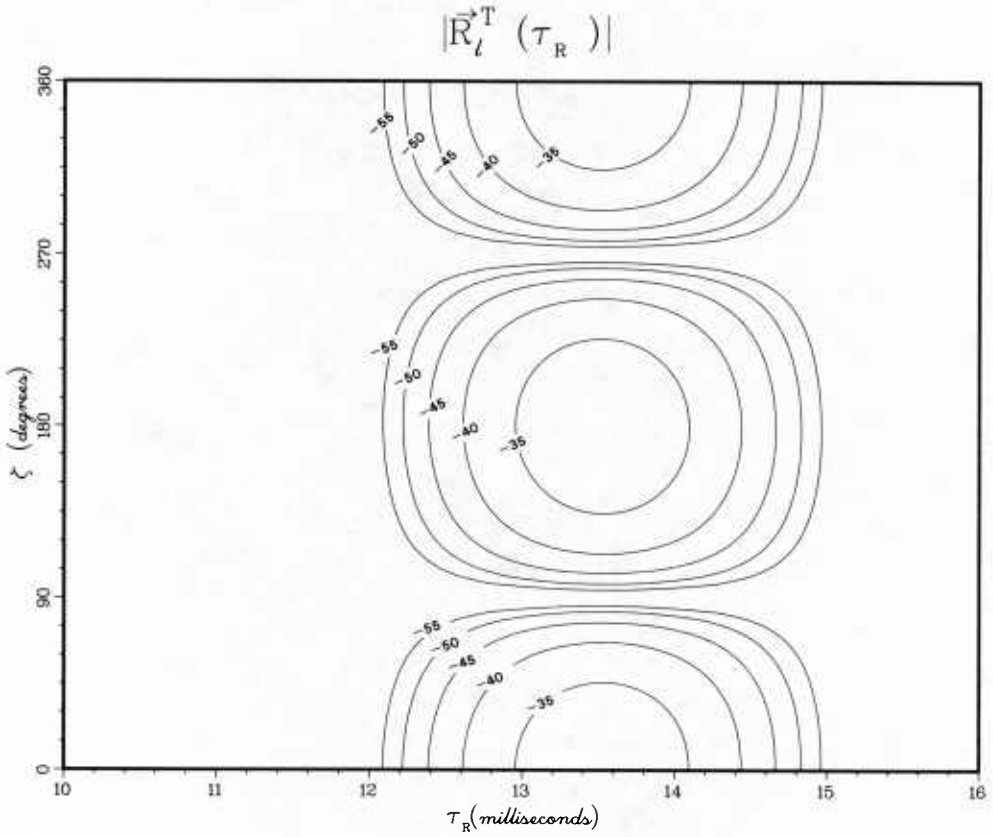


Fig. 99c: Amplitude contours corresponding to Fig. 99a

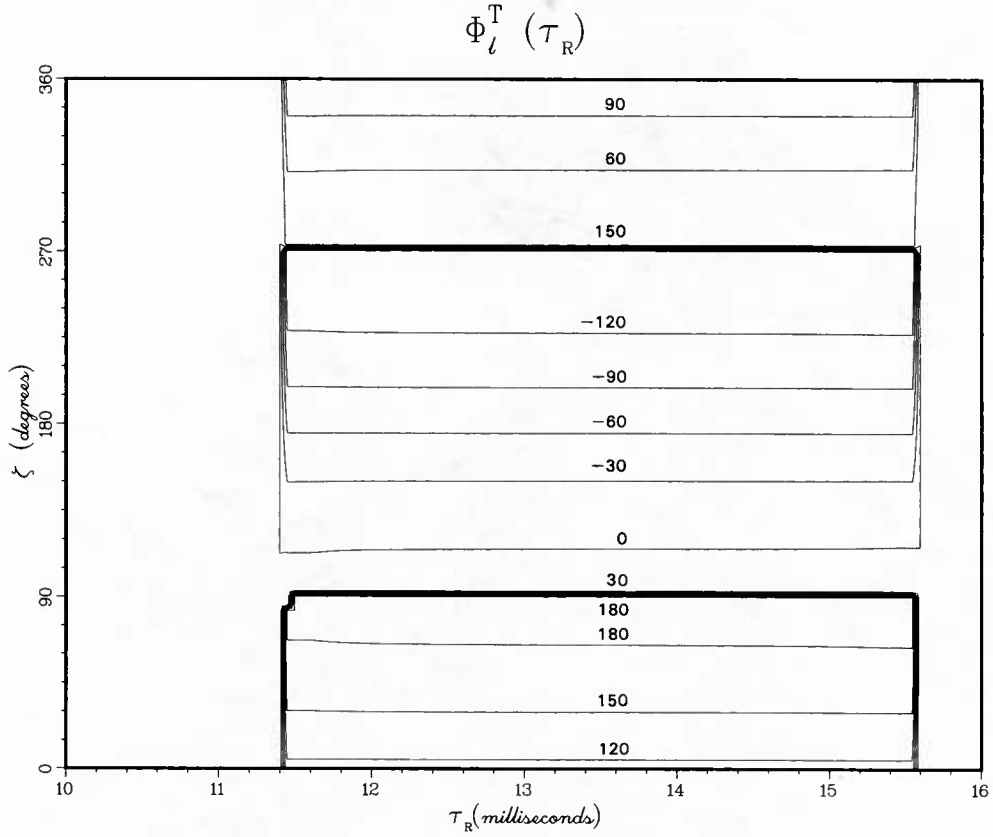


Fig. 99d: Phase behavior corresponding to Fig. 99b

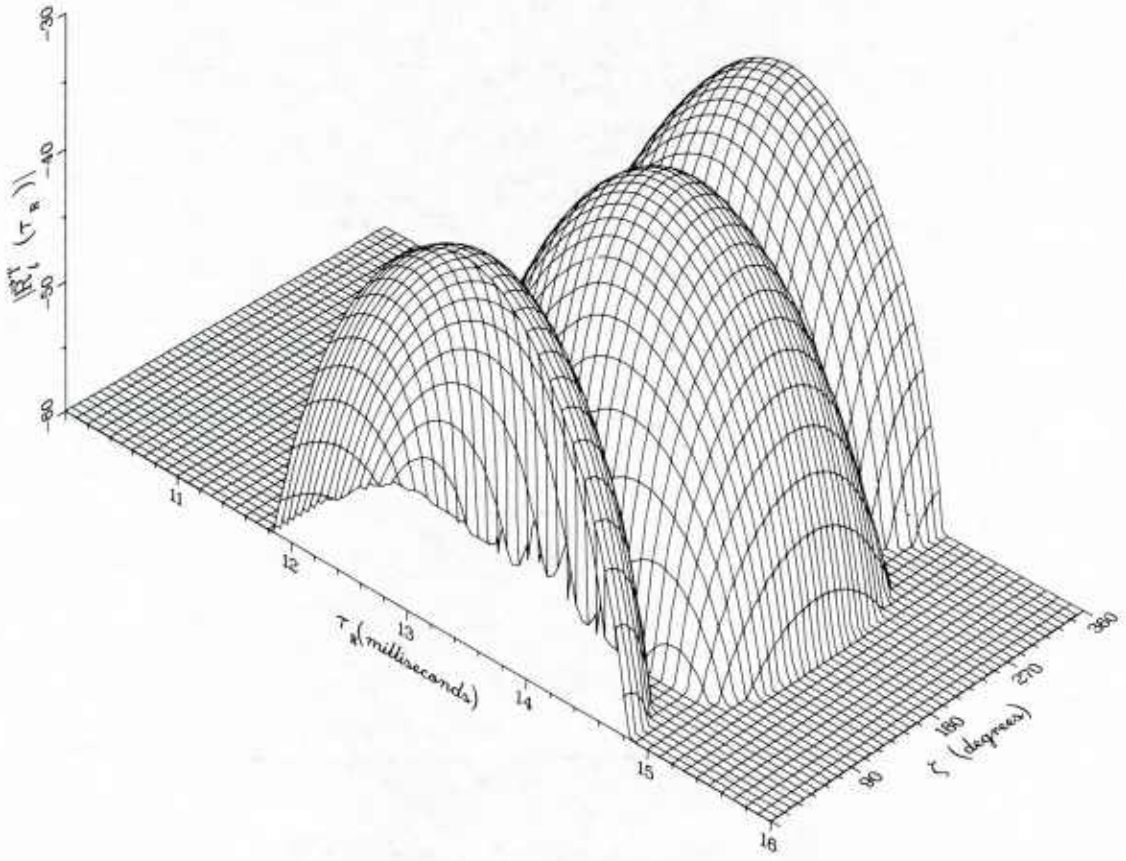


Fig. 100a: Amplitude behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 85°$, calculated with the parameters of Table 7

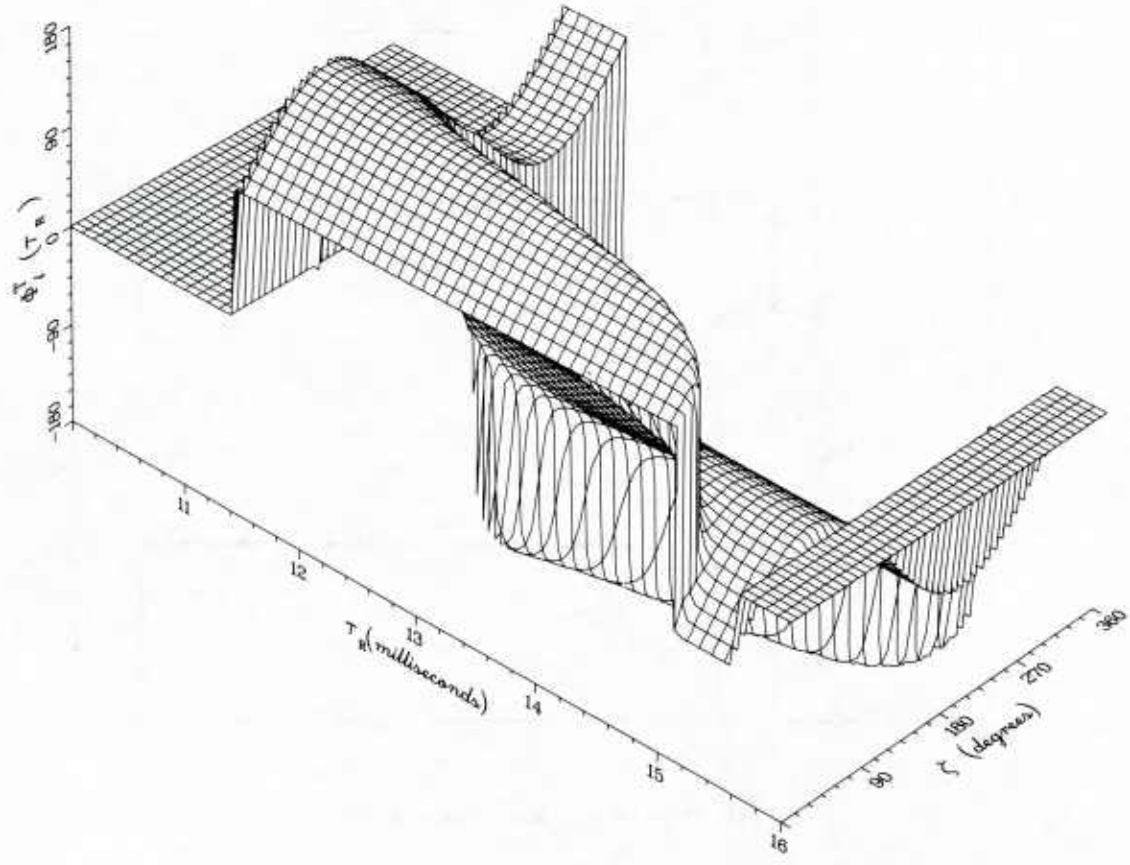


Fig. 100b: Phase behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 85°$, calculated with the parameters of Table 7

$|\vec{R}_l^T(\tau_R)|$

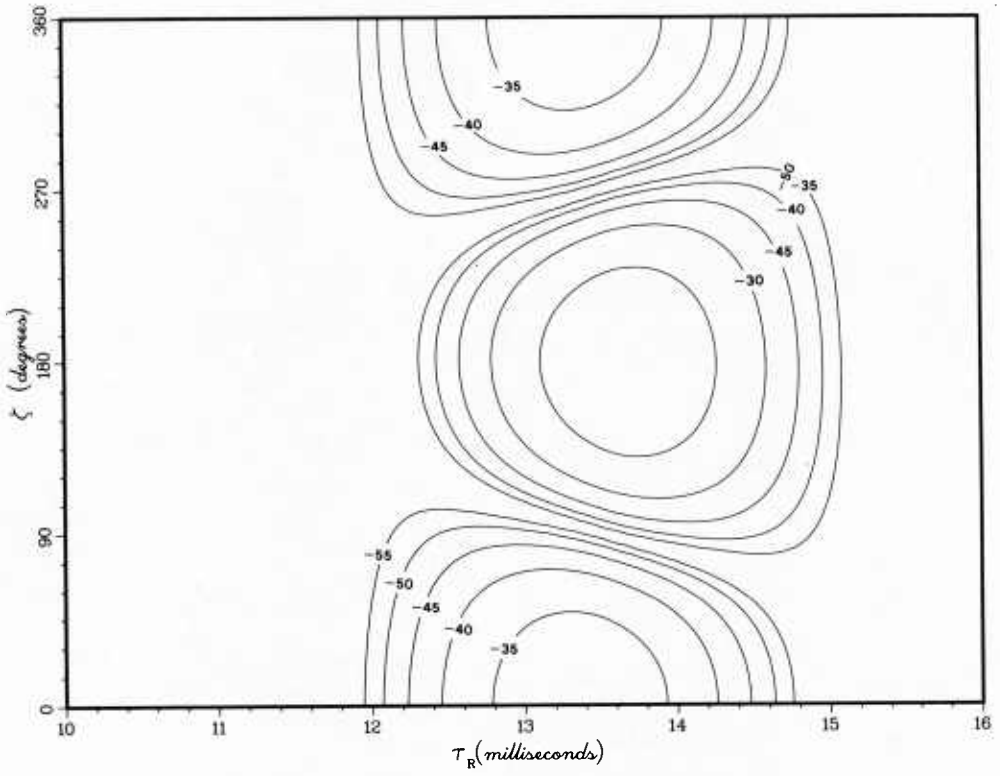


Fig. 100c: Amplitude contours corresponding to Fig. 32a

$\Phi_l^T(\tau_R)$

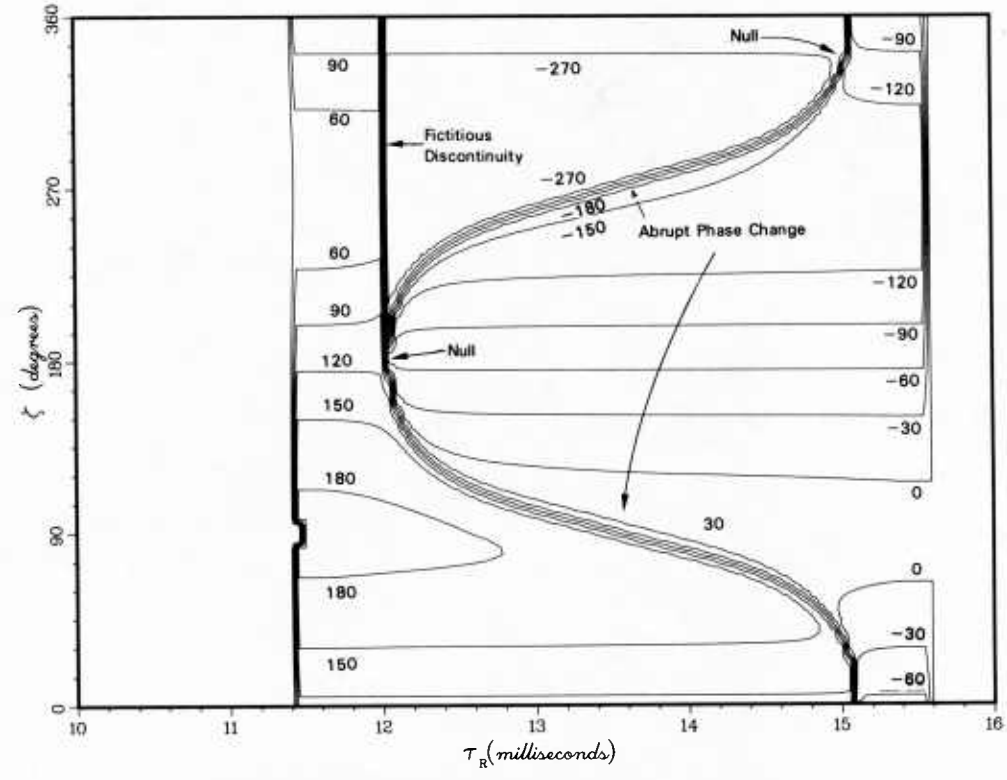


Fig. 100d: Phase behavior corresponding to Fig. 32b

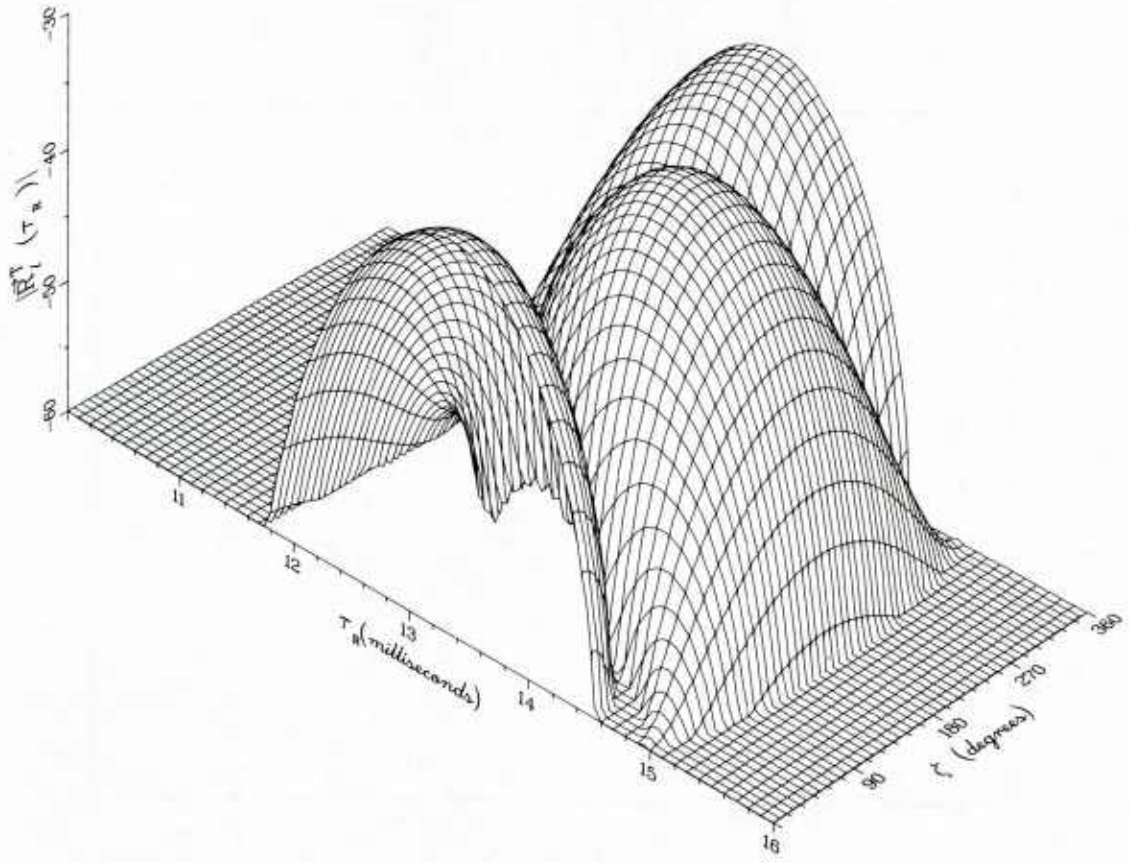


Fig. 101a: Amplitude behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 80°$, calculated with the parameters of Table 7

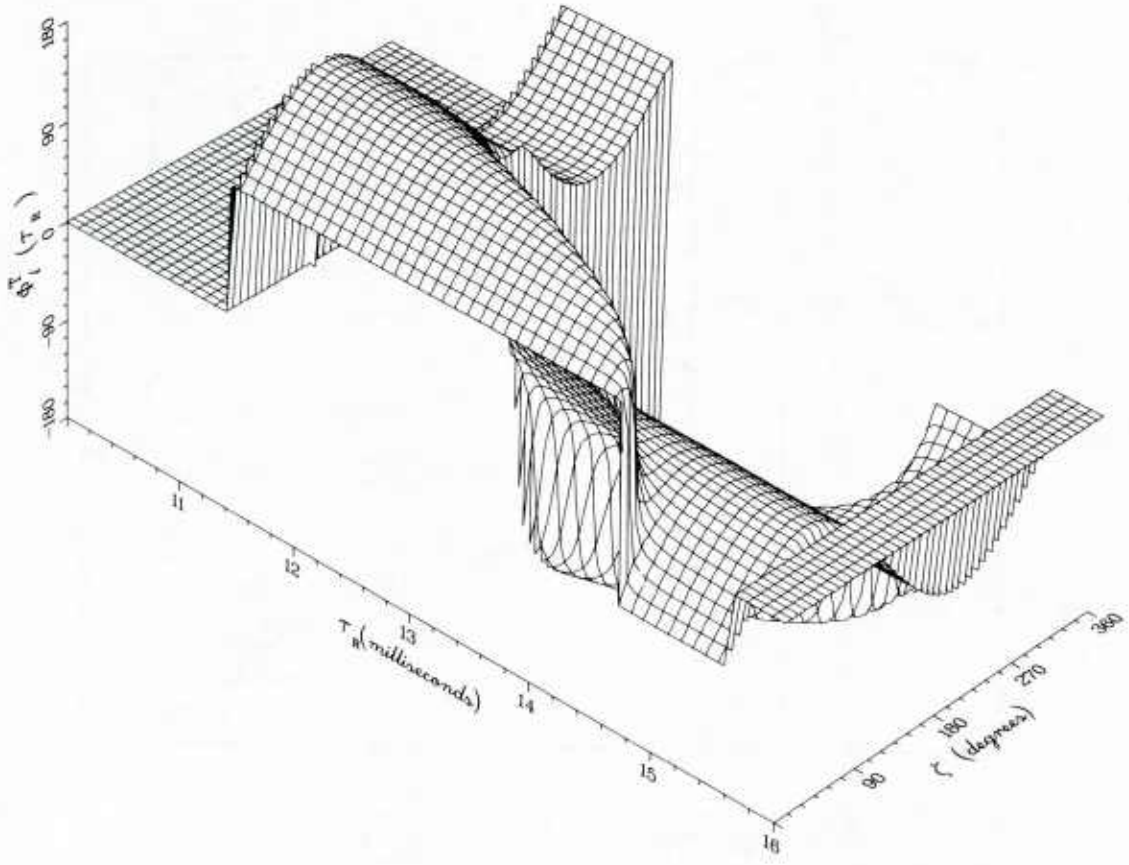


Fig. 101b: Phase behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 80°$, calculated with the parameters of Table 7

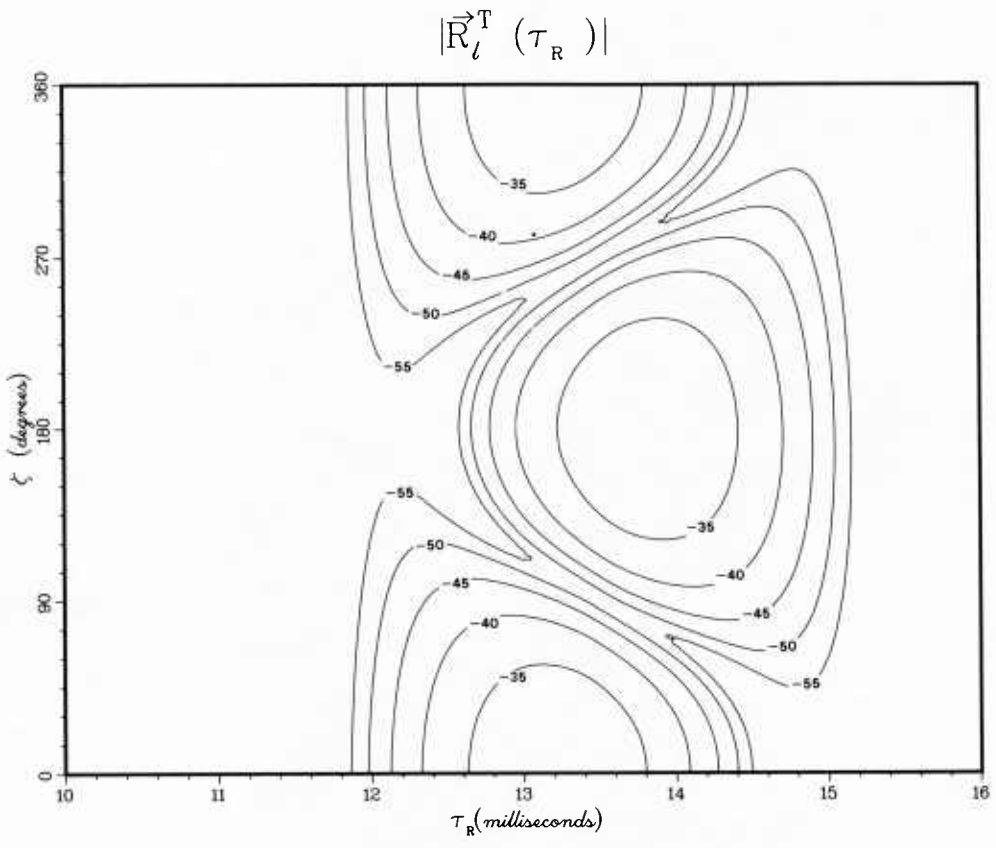


Fig. 101c: Amplitude contours corresponding to Fig. 101a

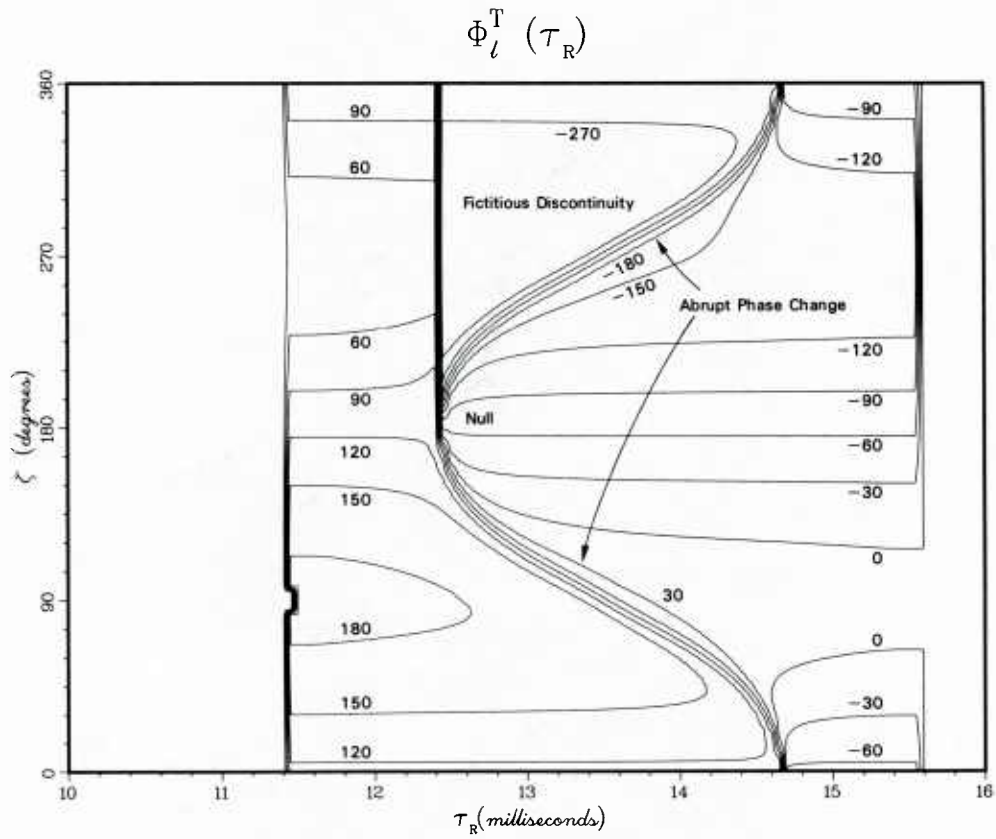


Fig. 101d: Phase behavior corresponding to Fig. 101b

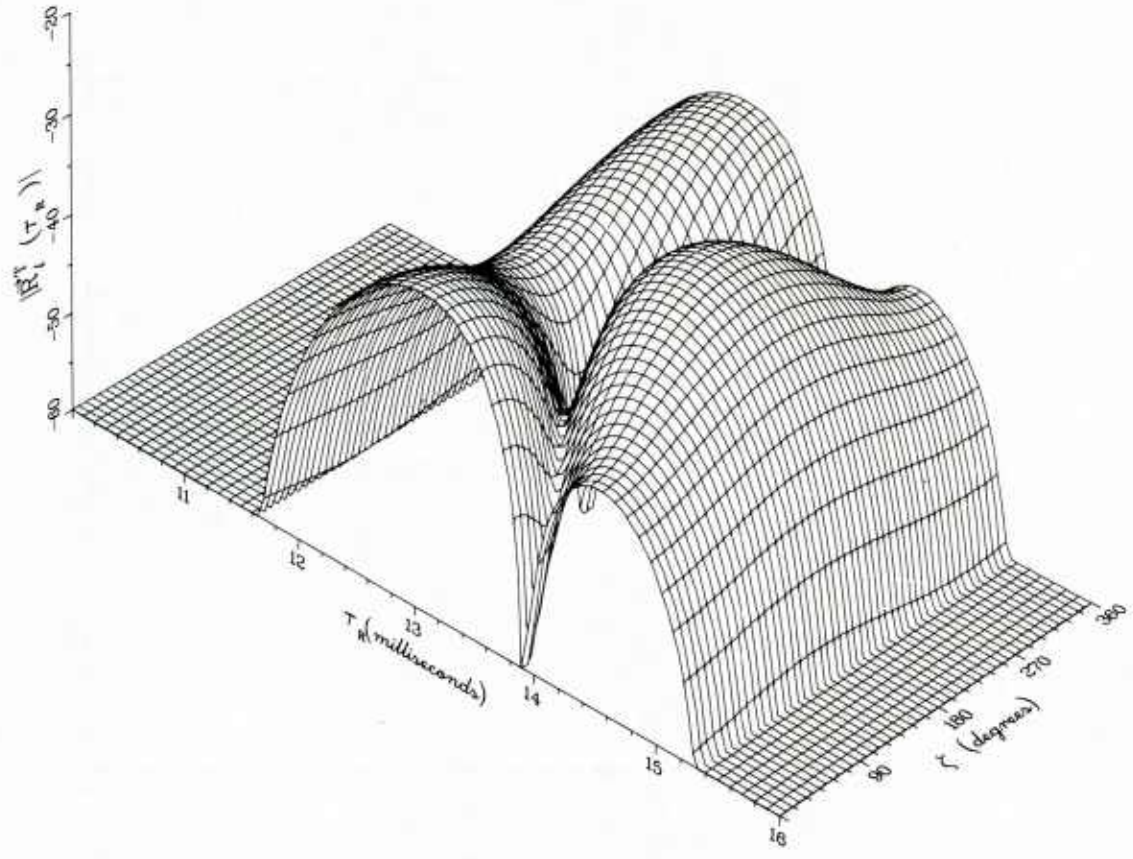


Fig. 102a: Amplitude behavior of $\bar{R}_l^T(\tau_R)$ for $\gamma = 60°$, calculated with the parameters of Table 7

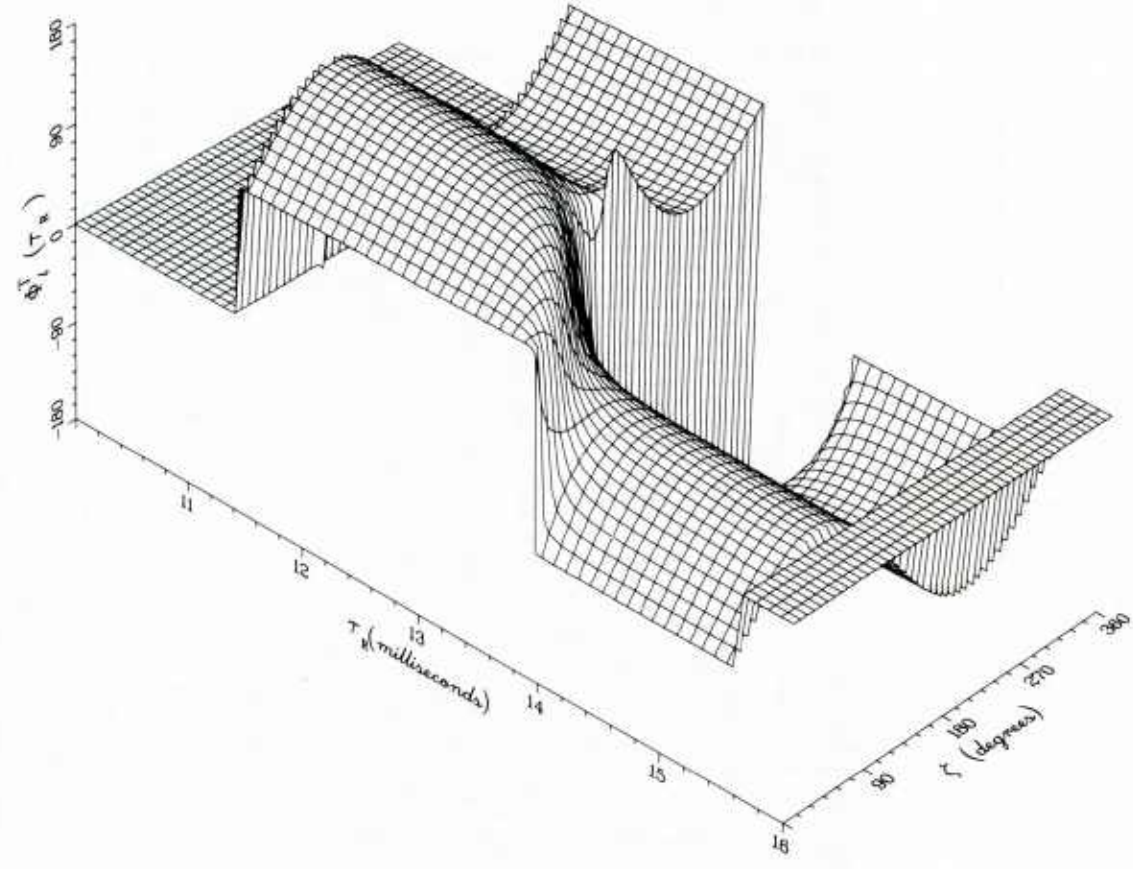


Fig. 102b — Phase behavior of $\bar{R}_l^T(\tau_R)$ R $\gamma = 60°$, calculated with the parameters of Table 7

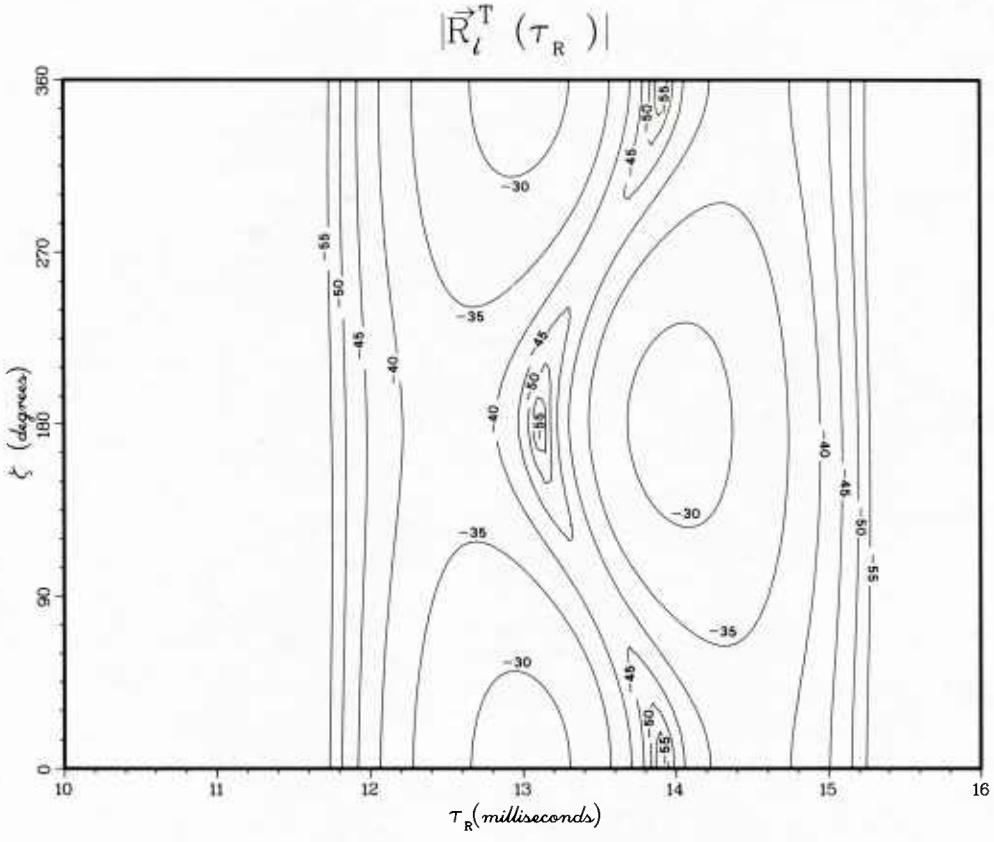


Fig. 102c: Amplitude contours corresponding to Fig. 102a

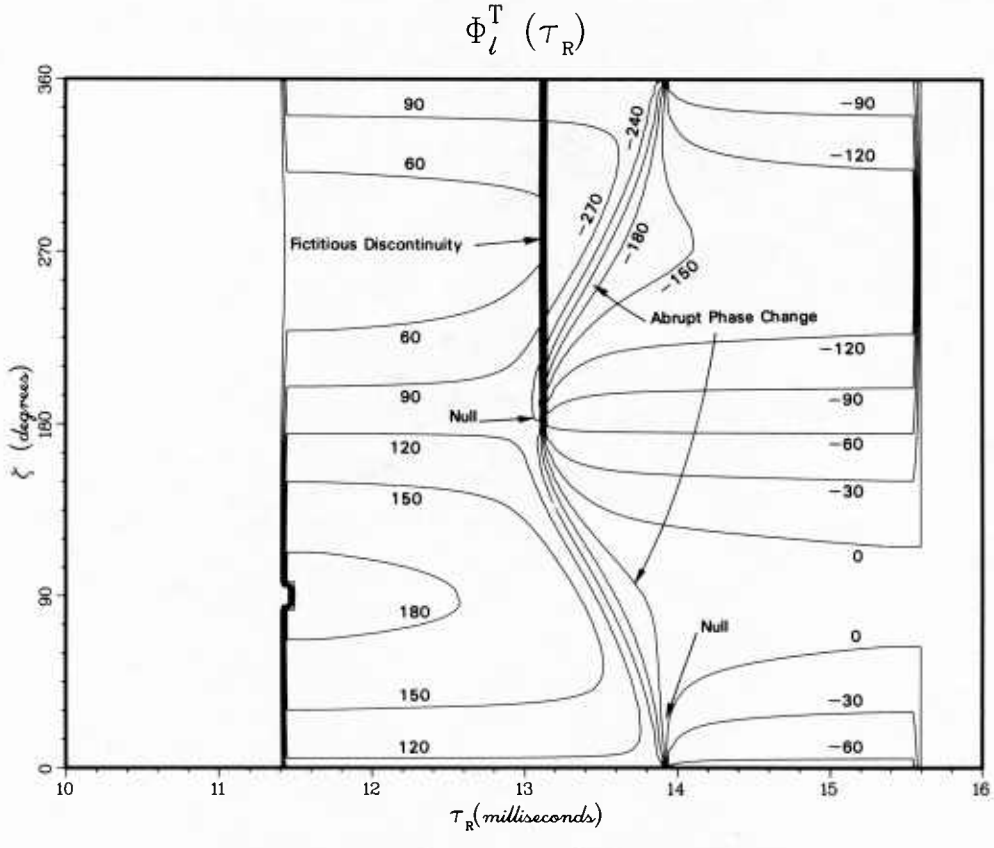


Fig. 102d: Phase behavior corresponding to Fig. 102b

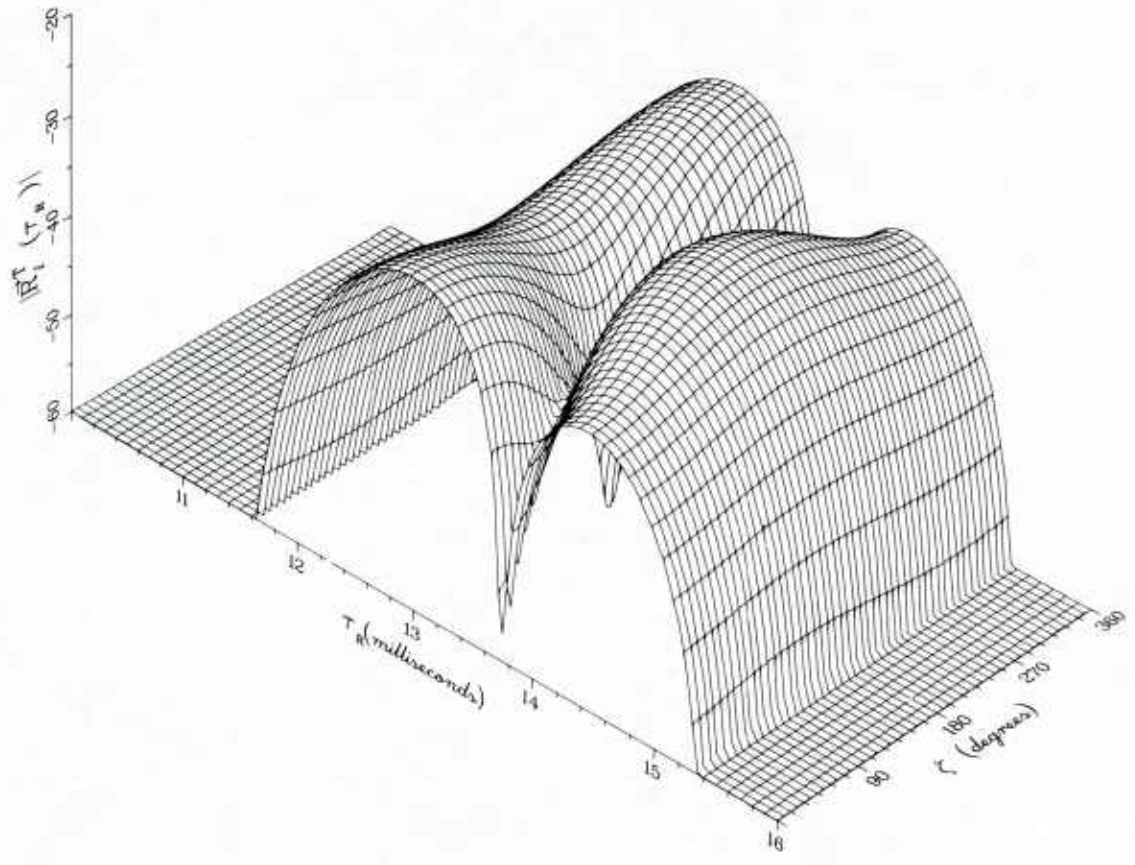


Fig. 103a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 45°$, calculated with the parameters of Table 7

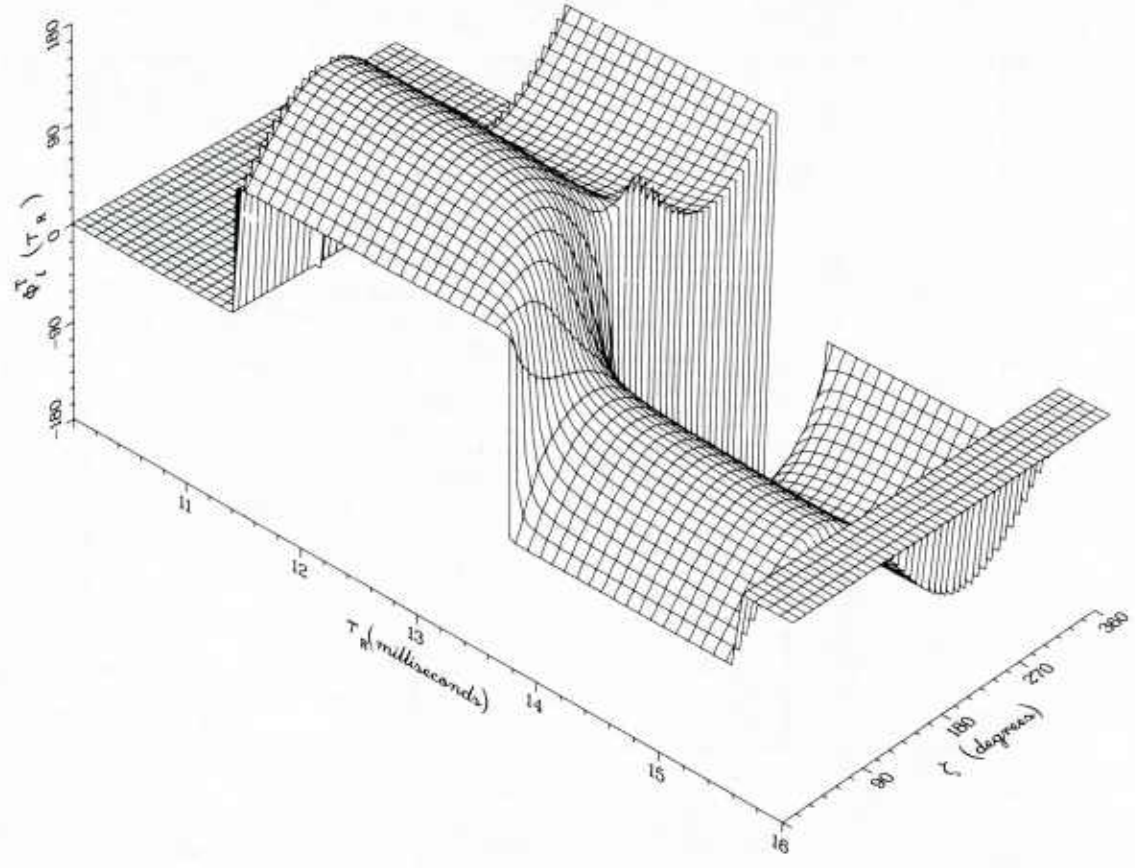


Fig. 103b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 45°$, calculated with the parameters of Table 7

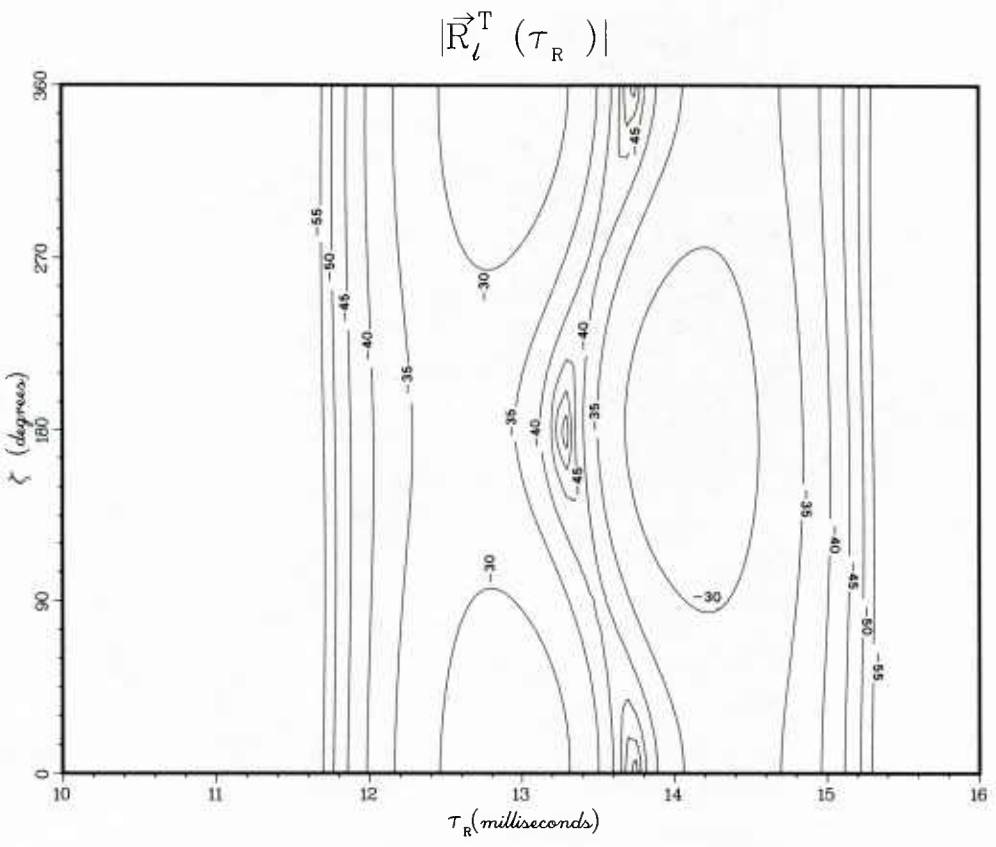


Fig. 103c: Amplitude contours corresponding to Fig. 103a

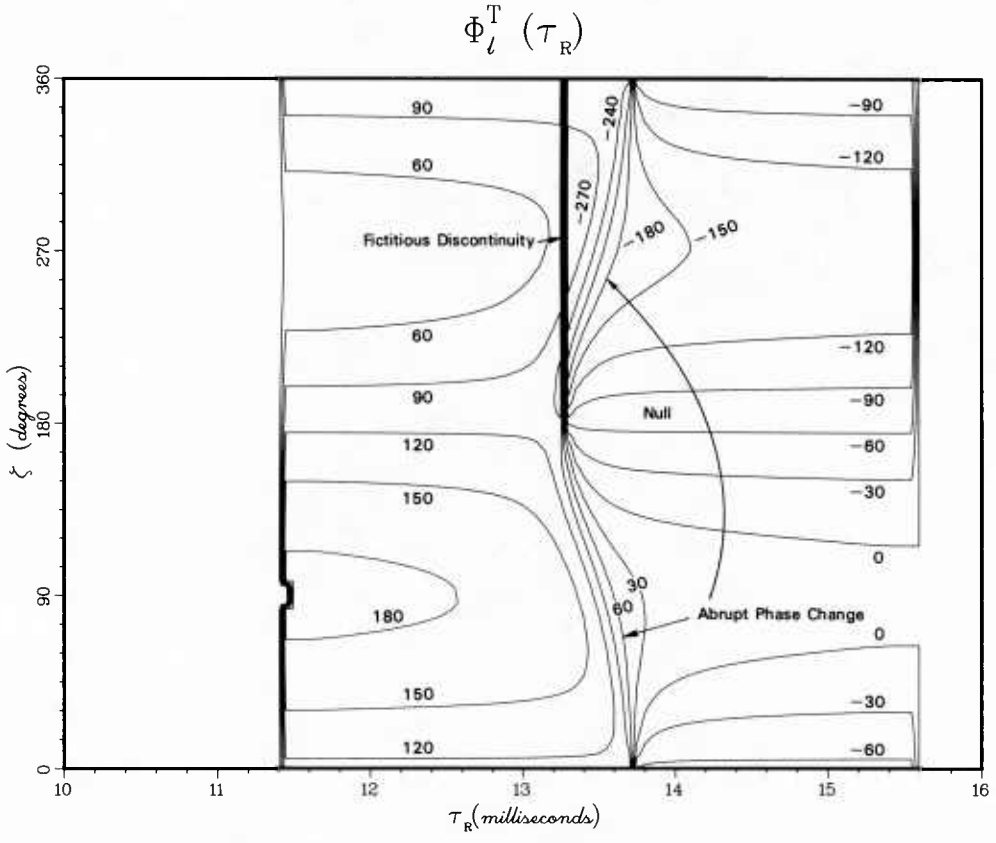


Fig. 103d: Phase behavior corresponding to Fig. 103b

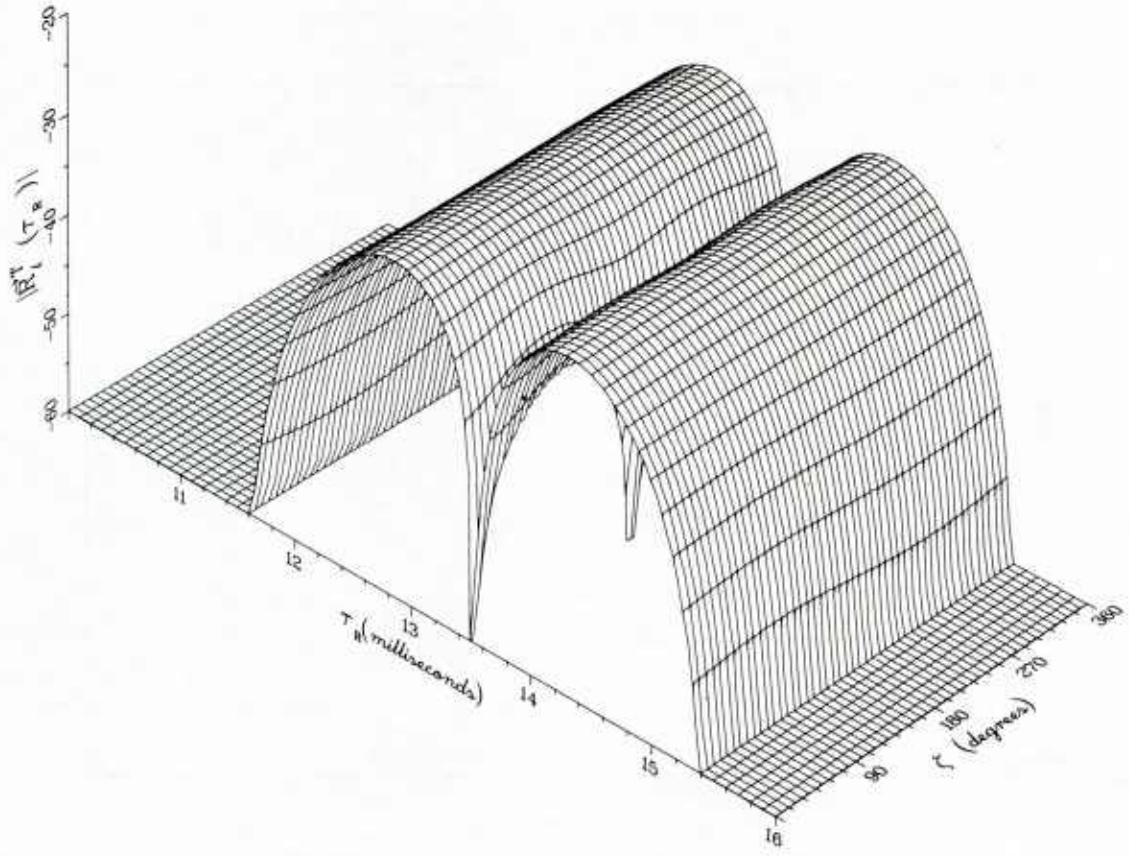


Fig. 104a: Amplitude behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 0°$, calculated with the parameters of Table 7

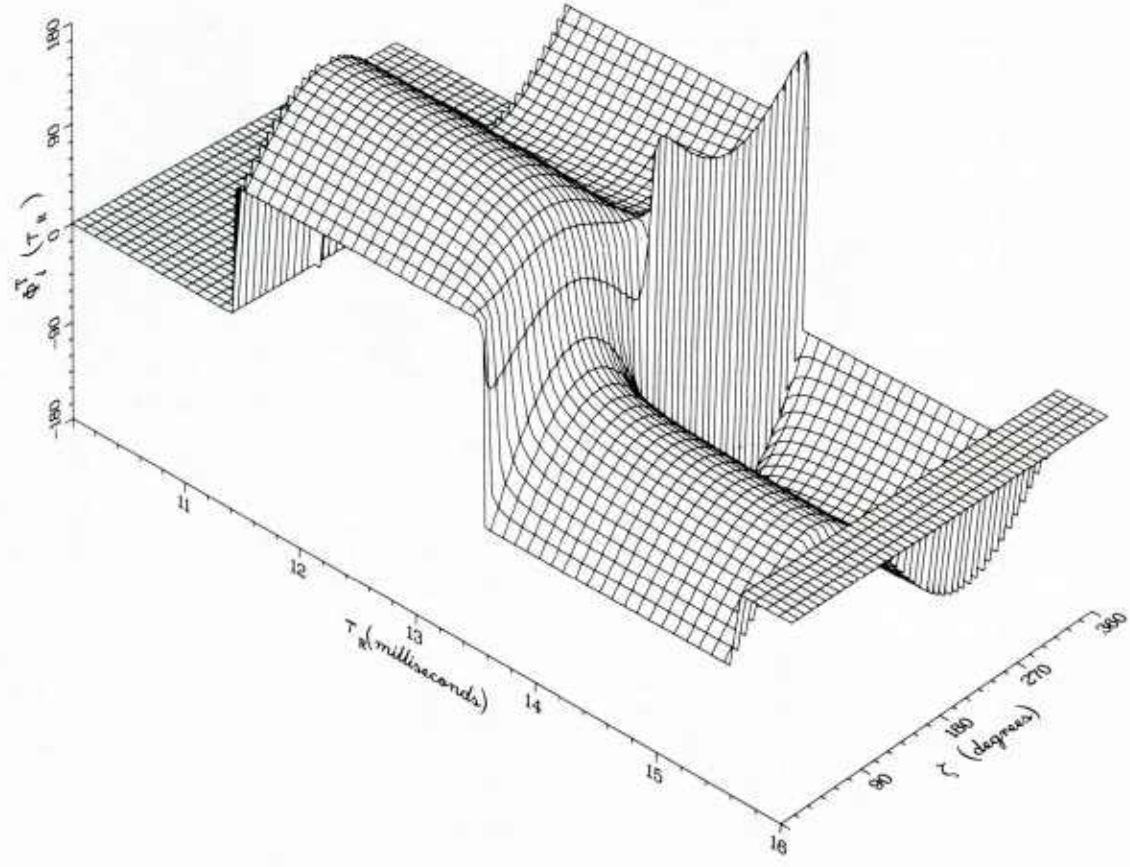


Fig. 104b: Phase behavior of $\vec{R}_l^T(\tau_R)$ for $\gamma = 0°$, calculated with the parameters of Table 7

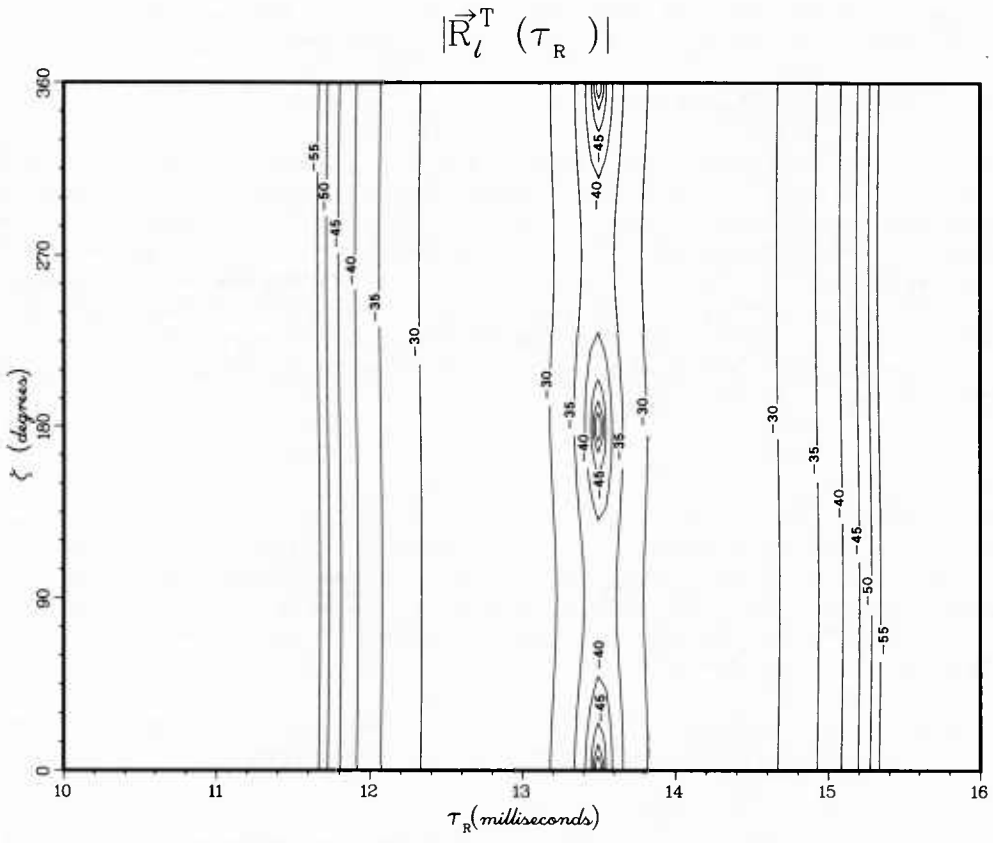


Fig. 104c: Amplitude contours corresponding to Fig. 104a

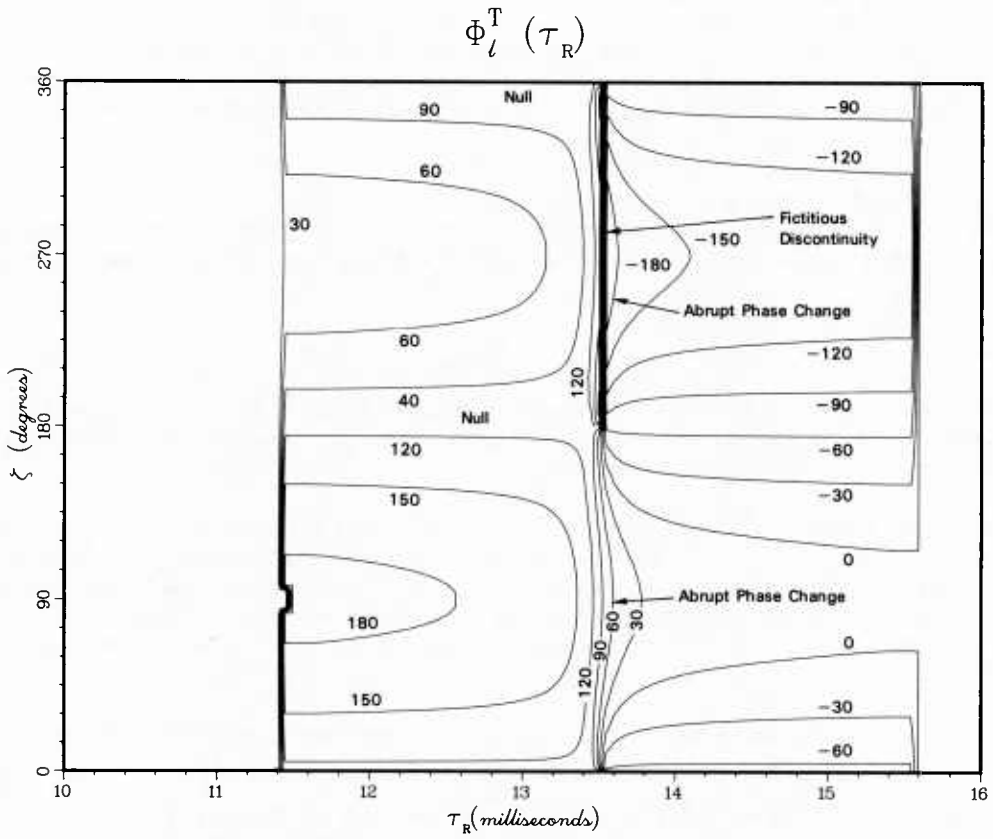


Fig. 104d: Phase behavior corresponding to Fig. 104b

Many things can be expected to disturb radio waves—but to what degree?—and how deleterious are the various disturbances? Some answers are known, but a lot is still to be revealed.

7.1 Solar Flares—Sudden Ionospheric Disturbances

Long wavelength waves have always been thought of as more reliable than high frequency waves because the solar flares and magnetic storms which tend to completely obliterate long-haul high frequency communications circuits were not able to shut down communications on long waves. Figuratively, the very low frequency waves effectively "ducked under" the enhanced ionization layer of the lower ionospheres which cause the severe high frequency signal degradation. To be sure, the solar flares are noticeable on long wavelength circuits; but they are normally seen as a change of signal level not as a complete loss of signal. Sometimes the signal will increase, sometimes it will decrease. The precise behavior is currently unpredictable without making exacting predictions of the locations of interference nulls under normal and disturbed conditions. On the whole I believe that the VLF signal is more likely to be enhanced than depressed during a Solar Flare for an all daytime path at temperate latitudes. [Ref. 59]

7.2 Polar Cap Absorption Events

Polar cap absorption (PCA) events are natural disturbances which affect VLF propagation for paths crossing the polar regions. PCA's are caused by a flux of charged particles (protons) emitted from the sun during a solar flare. The particles follow the lines of the magnetic field and effectively lower the height of the ionosphere in the polar regions because of the enhanced ionization. The onset of a PCA effect on VLF is more gradual than that of a sudden ionospheric disturbance, but the effects are more long-lasting. PCA events lasting five days are not uncommon and they occur fairly frequently. There were eleven PCA's between 24 March 1966 and 6 June 1967.

Papers by Westerlund, et al. (60) and Dalland, et al. (61) contain detailed discussions of PCA events and their effects on VLF signals. The following is a brief summary of the material of these references which seems relevant to a communications channel model.

During a PCA, the phase of a VLF signal normally advances to some maximum amount and later returns to its pre-PCA value. Phase advances of as much as 50 μs have been observed. The behavior of the amplitude of a VLF signal during a PCA is not completely predictable. For paths over water in the polar regions the amplitude variations are generally less severe and unpredictable in sign. However, for paths which involve a good deal of extremely low conductivity soil (Greenland or Antarctica) the amplitude consistently shows strong signal losses (10-20 dB) when a PCA occurs. Figure 105 shows the effects of three PCA events on the VLF signal over three Arctic paths. Note that the paths NPG-Stockholm and WWVL-Stockholm show severe attenuation during the PCA's. Both of these paths pass over the low conductivity ground of Greenland, The path from NPM-Stockholm does not pass over Greenland and relatively small changes in amplitude are observed.

The behavior of the VLF channel during PCA event is very significant for communications over some high latitude paths. A very low data rate modulation scheme could compensate for the very severe long duration attenuation which occurs. At the same time the phase variation is fairly slow over a message time and should not pose a severe obstacle to communication.

7.3 Transition Fading

Transition fading refers to signal fades which occur when the sunrise or sunset line (the terminator) crosses the propagation path between the transmitter and receiver. This phenomenon has been studied by Crombie (62, 63, 64) and others (65, 66). We will summarize here the material from these references.

In Crombie's model the difference in ionospheric heights on either side of the terminator causes modal conversion in the waveguide. The resultant interference on the side of the terminator away from the transmitter gives rise to the transition fading. The principal effect of the fading is to introduce multimode complications similar to those already observed at nighttime. During these times the amount of the multimode interference is determined by the relative amplitudes of the modes and the group and phase delays involved. The effect should be comparable to that observed during normal nighttime conditions when multimode effects are also important.

The transition fading is a dynamic phenomenon because the terminator moves along the path length. Therefore, the rates of change of phase and amplitude are important. Crombie (62) gives an observed maximum $\frac{d\phi}{dt}$ of $\sim 70 \pm 2°$/minute. However, such phase rates could become very large for very short times and a step function jump of phase is possible if a very sharp and severe modal interference null occurs (67). The rate of change of amplitude with time is normally in the neighborhood of 0.3 dB/min during a fade. Faster rates of change are, of course, possible for very deep fades. Typical fade depth is in the neighborhood of 10 dB; however, greater fades have been observed, and, of course, even perfect cancellation of modes is theoretically possible.

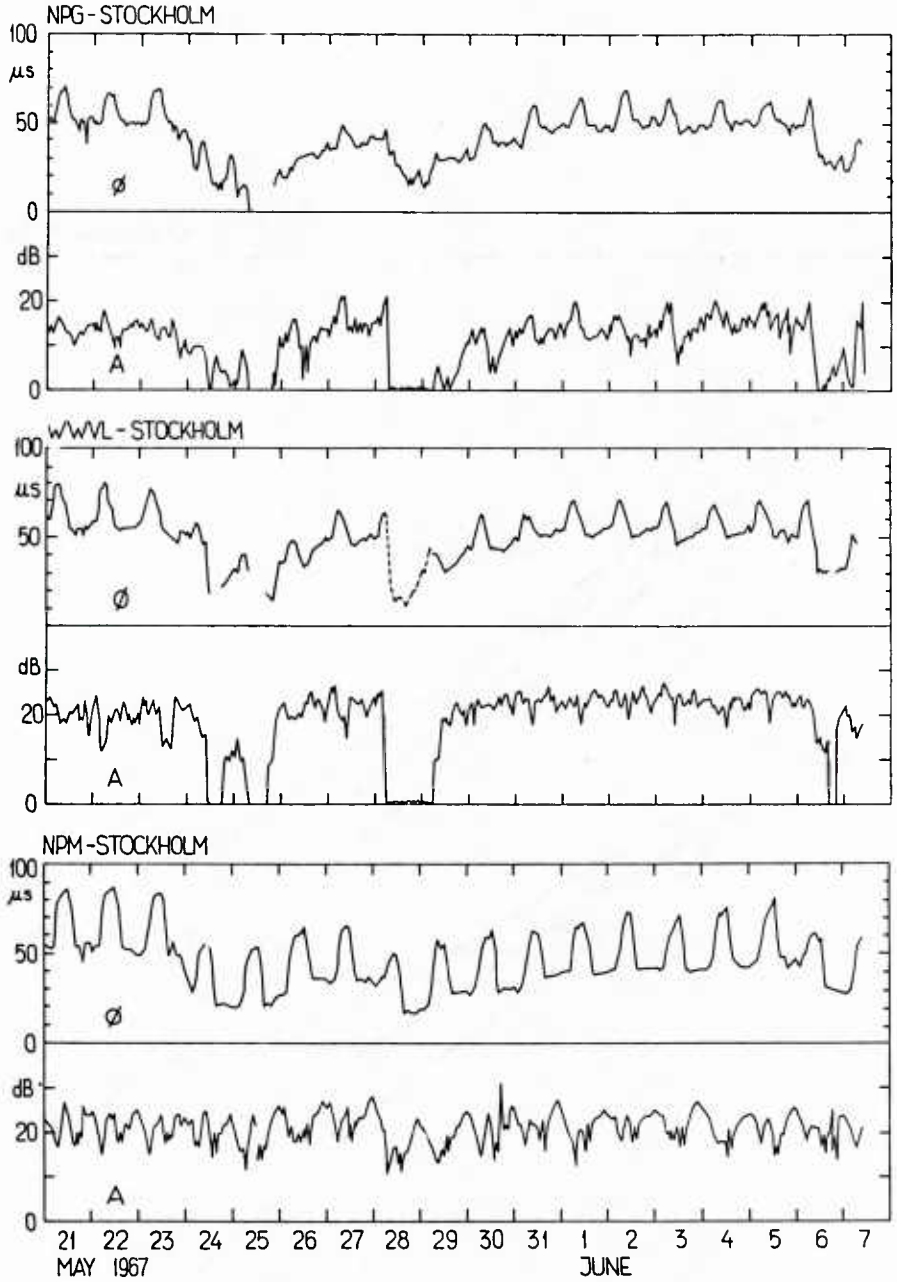


Fig. 105: VLF phase and amplitude data for the transmissions monitored during PCA's of 23 May 1967, 28 May 1967 and 6 June 1967

A statistical analysis of transition fading depths and durations has been performed by Schneider and Carlston (66). They have applied the statistics of extreme values (68) to data of Rhoads and Garner (69) for the path from Haiku to Washington, D.C., in May and June 1965. They find a good fit to a probability distribution for depth of fade and outage times given by type I and type III distributions of the statistical theory of extreme values.

The cumulative probability distributions for the depths of fade and the times of signal strength, 10 dB below the normal daytime level, are shown on Figs. 106 and 107. Experimental examples of the diurnal variation of received signal from Haiku to Washington, D.C. (taken from Ref. (69)), are shown in Figs. 108 and 109.

7.4 Electron Precipitation Induced Perturbations

Anyone who has recorded longwave field strengths will note that the nighttime traces are more irregular and "shakey" than the daytime cases—even though the fields are usually stronger and less affected by ambient noise (70).

The picture that is emerging to explain this feature (Ref. 71) of propagation behavior is curious. It appears that the earth's magnetosphere is filled with energetic particles in the famous Van Allen belts. These belts are continually filling up with

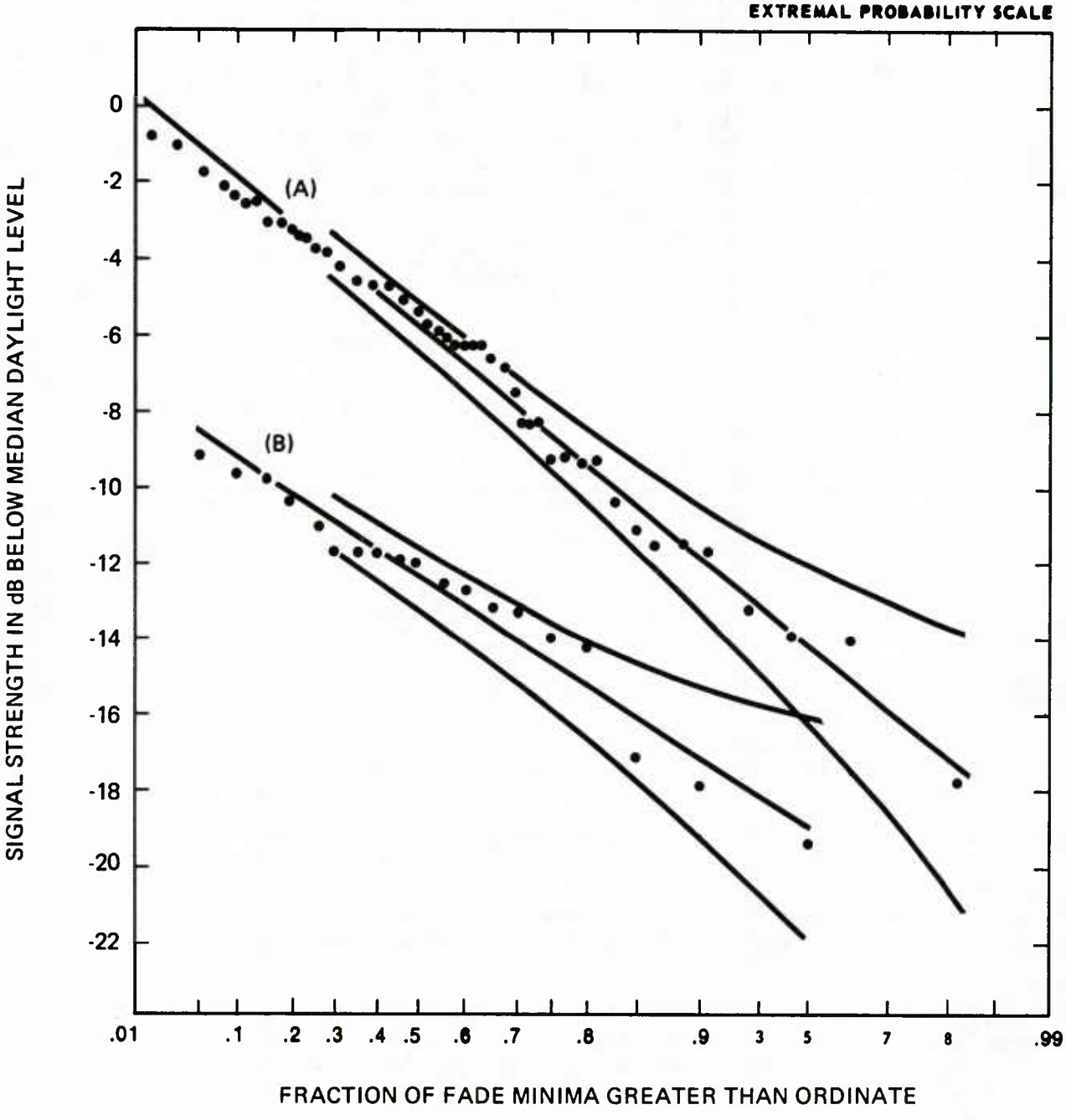


Fig. 106: (A) Distribution of transition fades for 10 consecutive days; (B) Distribution of the third sunrise fade for 19 consecutive days.

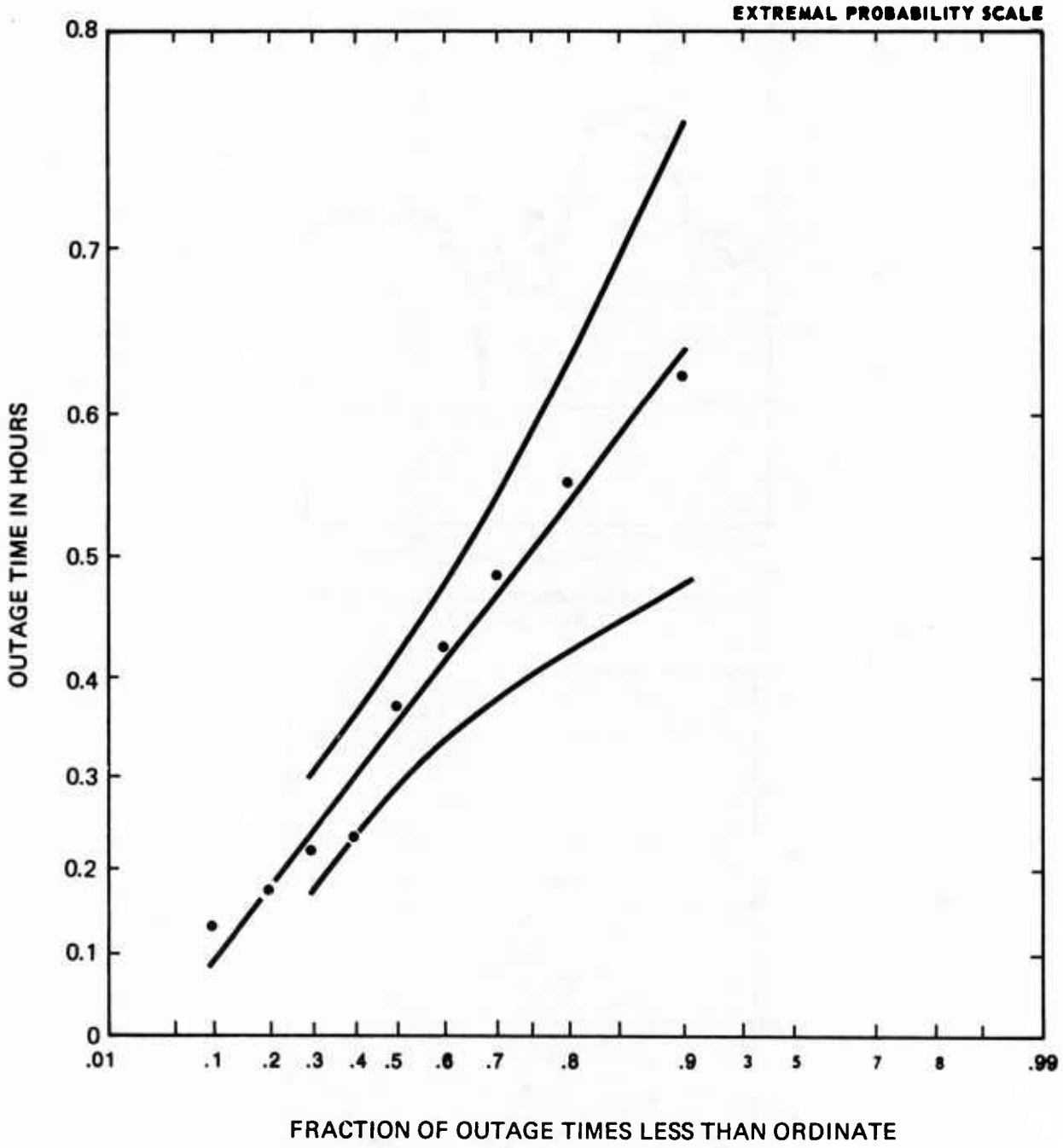


Fig. 107: Distribution of outage time at 10 dB below median daylight threshold level for 10 consecutive days. See Fig. 106(A).

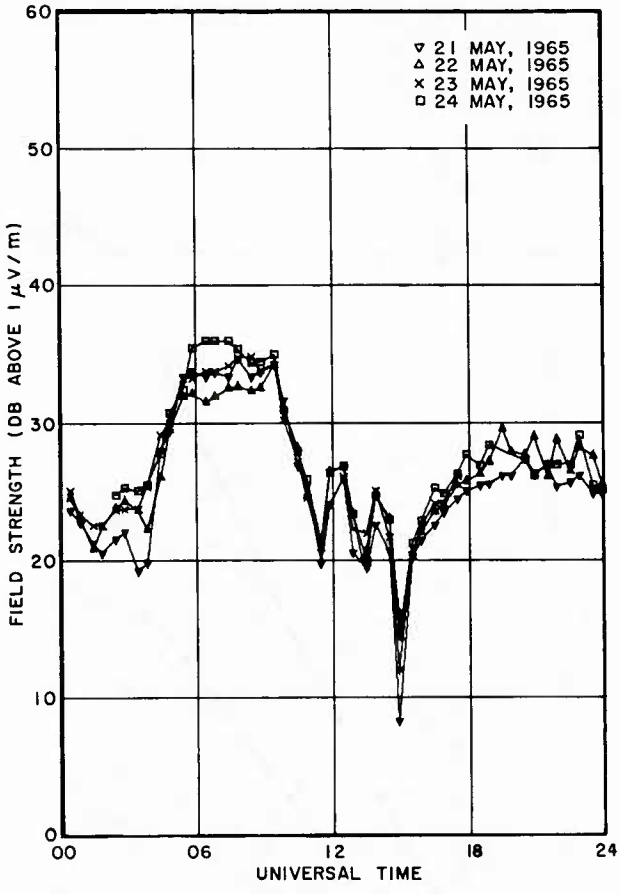


Fig. 108: Haiku (19.8 kc/s) data recorded May 21, 22, 23, 24, near Washington, DC

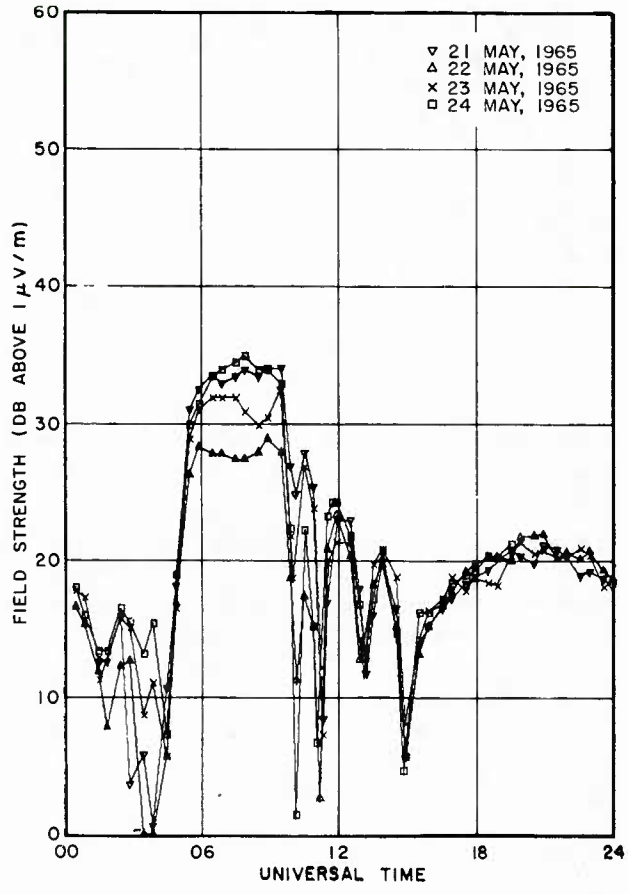


Fig. 109: NPM (26.1 kc/s) data recorded May 21, 22, 23, 24, near Washington, DC

particles. An equilibrium is reached which depends on the plasma density in the magnetosphere and the intensity of lightning generated whistler waves which enter the magnetosphere from the earth ionosphere waveguide (Ref. 72, 73). The propagation of whistler waves through the magnetosphere can shake or scatter the trapped electrons into their ''loss cone'', and they precipitate from the Van Allen Belts. When they fall from the magnetosphere, they strike the lower atmosphere and produce electrons in the nighttime D-layer. This sudden electron production causes a quick change in the ionospheric reflection height and thereby perturbs the waves propagating in the earth-ionosphere waveguide causing the abrupt events noted in nighttime recordings.

The reader is invited to learn more about this ''strange but (probably) true'' mechanism by studying Refs. 74 through 79.

The auroral zone is a region where the electron precipitation is greatly intensified over midlatitudes. It should be expected that these electrons shall produce disturbance also (Refs. 80 through 83 and Ref. 70).

7.5 ELF Nighttime and PCA Disturbances

The ELF longwave communications systems has an interesting disturbance mode that has not been exhibited by the VLF waves. Because of the huge wavelength of ELF (5Hz-3000Hz) waves their reflection coefficient matrix (Eq. (1)) at nighttime depends strongly on the electron density distribution up to the F-layer! This means that the ELF nighttime wave will be subject to the vagaries, that HF exhibits [Refs. 84,85] — except the ELF will be averaged over a larger Fresnel zone. This possibility was early pointed out by J. Galejs [Refs. 86, 87, 88]. It is being treated now in greater detail (and with greater respect) in recent years (Ref. 89, 90, and 91).

Ref. 92 treats the interesting possibility that the change in ELF propagation velocity under a strong PCA could bend the ELF waves so as to cause weakening and reenforcements of the waves in unpredicted ways.

7.6 C-Layer Disturbance

Under some circumstances during daytime, cosmic rays can produce a significant ionization layer beneath the D-layer. This is known as a C-layer [Refs. 93 and 94]. The formation and disappearance of this layer is somewhat of a mystery (but so is the D-layer). Obviously with the right intensity this layer will cause its own assembly of effects. The most notable effect is a predicted extra attenuation for waves that go twice through this layer before reflecting at the D-layer. Cosmic rays are known to vary in intensity with the eleven year solar cycle and with solar eruptions (Ref. 95).

7.7 Ground Conductivity and Weather Effects in Low Frequency Radio Propagation

In the low frequency propagation research community awareness is growing of the need for a better propagation model which takes ground conductivity, terrain, and atmospheric conditions into account (Refs. 96-102). For example, Fig. 110 taken from reference (99) shows the strong relationship between the time of arrival of a Loran-C groundwave signal transmitted from Carolina Beach (N.C.) to Fort Wayne (Ind.) and the atmospheric temperature recorded at Roanoke, Virginia very close to this propagation path. Reference (96) puts forth theoretical ideas to explain the weather related propagation delay in terms of changes in the gradient of atmospheric index of refraction. The disturbance of radio wavefields caused by irregular terrain and ground conductivity variations is often observed. For example, Fig. 111 from Ref. 103 shows data illustrating the ''recovery effect'' observed in passage over ground segments having dissimilar conductivities. Figure 111a shows the flight path of a receiving aircraft over Great Britain. Figure 111b shows the dramatic ''recovery effect'' as the groundwave passes over a river and over a land/sea boundary. The experimental path was chosen because of the flatness and homogeniety of the terrain, otherwise even greater irregularities would be evident in the portion of the flight over land.

Figure 112 (taken from Ref. 104) illustrates the variation of atmospheric index of refraction versus height which can produce ducting at radar frequencies. Such atmospheric changes should produce measurable effects at the edge of the effective coverage range of a groundwave communications or navigation system. Indeed the (100 kHz) Loran-C community has noticed this effect (Ref. 105).

8.0 NOISE

The prediction of atmospheric noise is a very important portion of the overall prediction of long wavelength system performance. The basic source of most long wavelength system noise is lightning in thunderstorms. During the 1960's the National Bureau of Standards conducted a long series of atmospheric noise measurements at stations around the world that led to the CCIR 322 publication (Ref. 106, 107).

Maxwell and others of the Westinghouse Georesearch Corporation utilized NBS data along with propagation models and the statistical records of thunderstorm occurrence to make a more realistic model of noise for the frequency range from 10 to 30 kHz (Ref. 108). This model was further improved and extended to predict atmospheric noise at any altitude and polarization within the earth-ionosphere waveguide and also up to nearly 60 kHz (Refs. 109-112).

The NBS noise data (Ref. 106) was taken through the LF and medium frequencies up to the HF range.

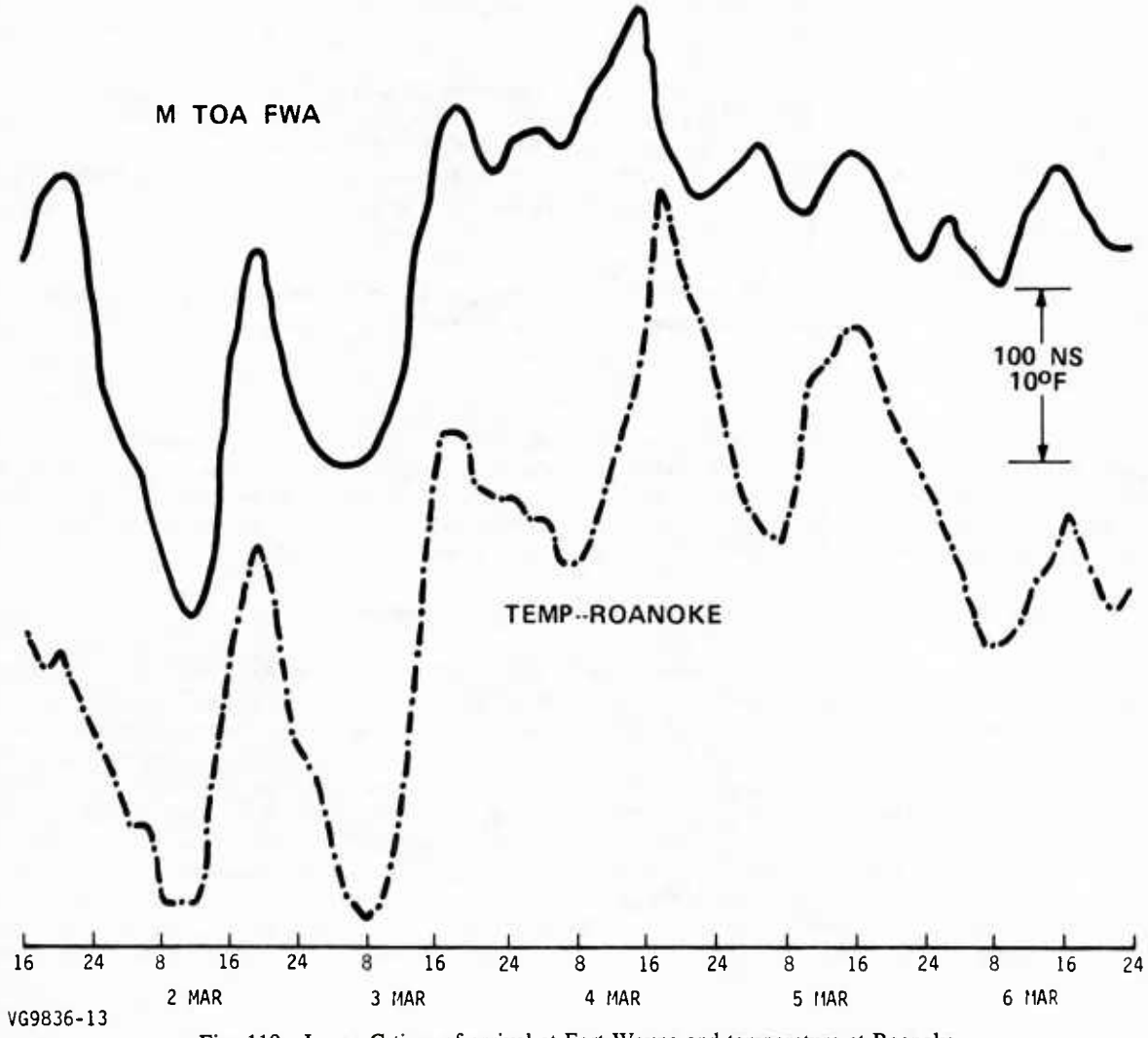
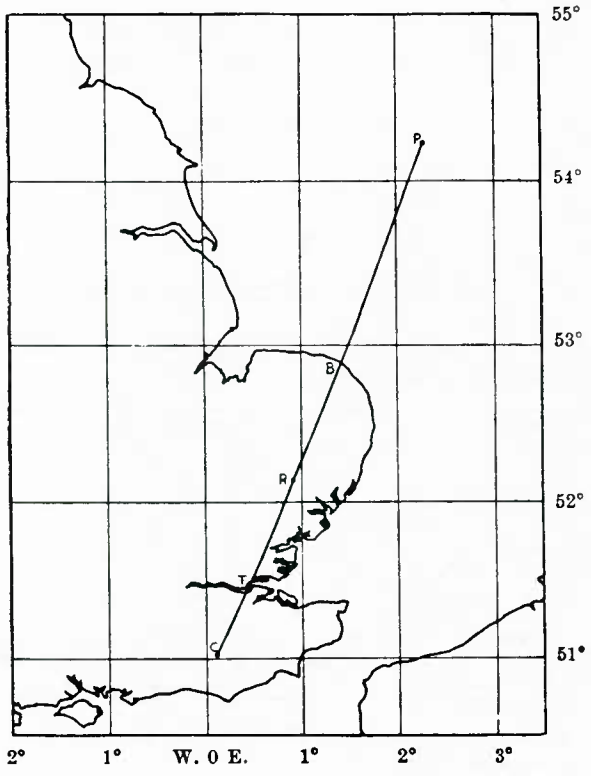
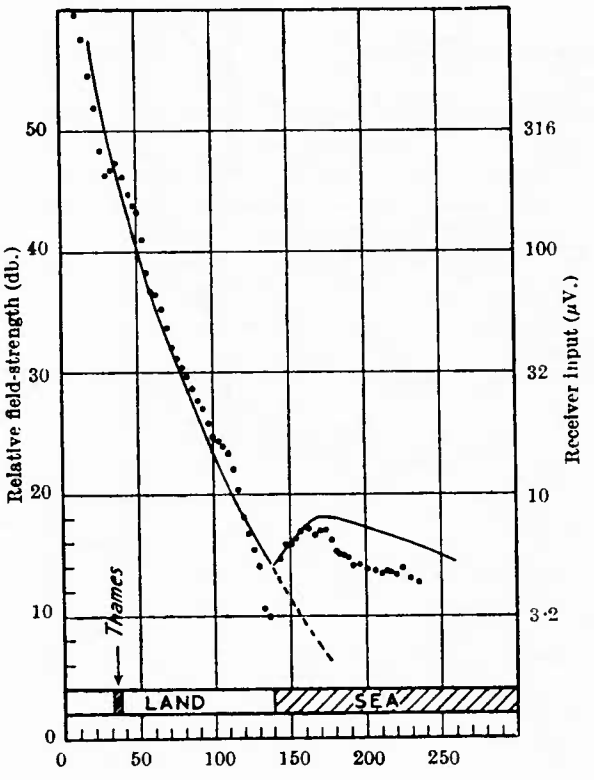


Fig. 110: Loran C time of arrival at Fort Wayne and temperature at Roanoke



(a)



(b)

Fig. 111: Data showing "recovery" effect on groundwave propagation on 1.1 MHz waves (Figure 1 from page 115 entitled "GROUND-WAVE PROPAGATION ACROSS A LAND/SEA BOUNDARY," by G. Millington from NATURE, No. 4157, Vol. 164, July 2, 1949, published by MacMillan and Co. LTD, used by permission.)

RADIO REFRACTIVE INDEX OF AIR

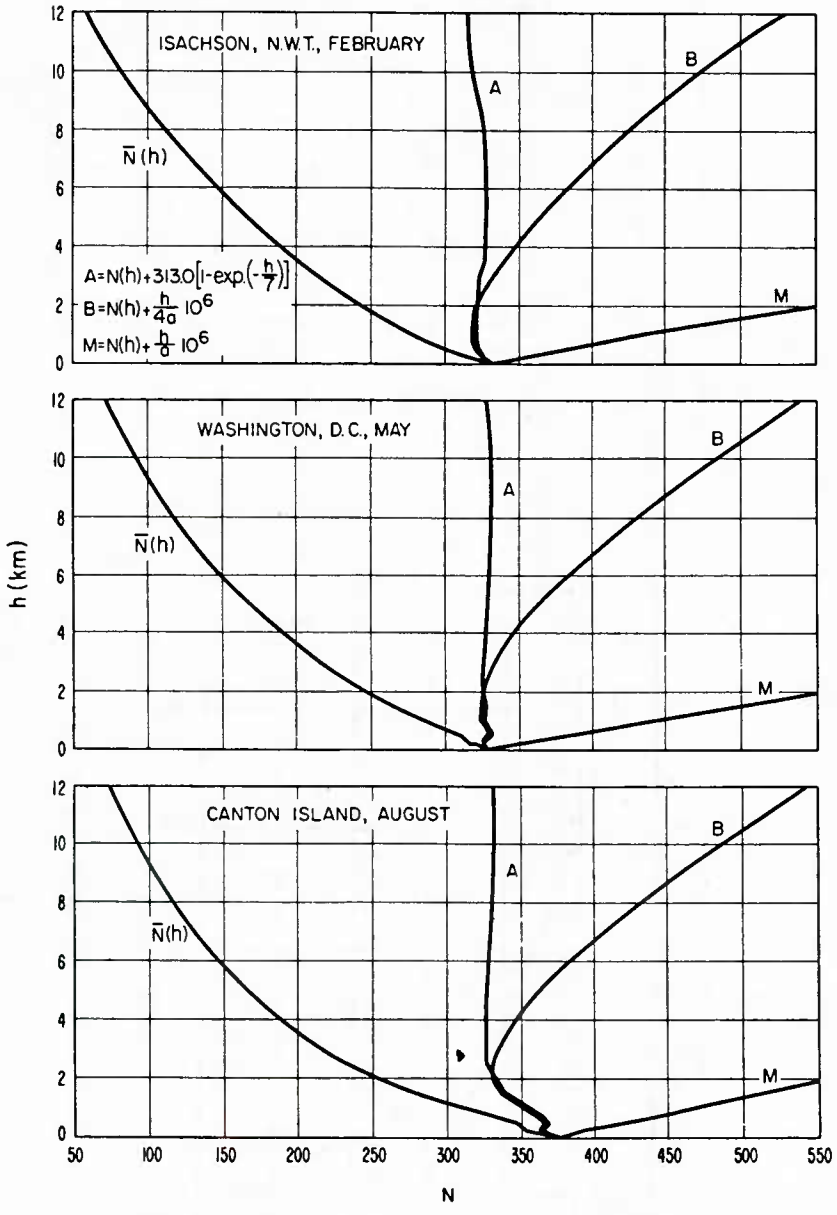


Fig. 112: Index of refraction data. $N = (n-1)10^6$

At ELF frequencies similar studies have been made (Refs. 114-119) on a sporadic basis. No official group has released worldwide maps of ELF noise and/or powerline interference. It would seem appropriate to extend the model of Ref. 112 down to ELF and up through MF to HF frequencies.

Some proportion of the noise is generated by currents flowing in the ionosphere and magnetospheres (Ref. 120). These noise sources are believed to be more important at polar receiver sites throughout the longwave band, but a clear understanding of precisely when, where and how much noise comes from extraterrestial sources is not known (Ref. 113).

8.1 Noise Mitigation

Because the highest noise pulses generated by lightning are strongly non-thermal in amplitude probability distribution various methods have been used to reject the energy contained in there pulses so as to reduce the "effective" root-mean-square (RMS) noise in a receiver much below that which would be seen by a "true" RMS noise meter. It is believed that Shannon's Theorem on Communication Channel Capacity is valid for gaussian noise statistics but that for non-Gaussian noise receiver can do better than look at the RMS noise. Noise theorists have defined a parameter called V_d which is the ratio of the RMS noise envelope to the average noise envelope. It is believed that an effective noise clipper circuit can reduce the noise level down to about the average levels. (This belief is always being sorely tested by the untractable performance of real systems in the real noise world.) Caution is advised for anyone who thinks that a few simple measurements will suffice to characterize a receiver's performance in noise for all time. Noise is noisy! It is variable in time and place and a good amount of experience with any non-linear receiver must be obtained before its performance can be relied on. It is even questionable how to characterize a receiver's performance in noise. Noise has an amplitude probability distribution and a time probability distribution; but these distributions may not be enough to characterize noise to permit a receiver's performance to be specified.

9.0 SYSTEM PERFORMANCE CALCULATION

At ELF Ref. 121 gives a good example of a longrange communication system performance calculation. At VLF and LF the system performance calculation is facilitated by computer programs (Refs 38, 46 and 47) which can be used to produce coverage maps such as those illustrated in Figs. 56 and 57 and in Figs. 113 through 116. One merely decides what S/N ratio is needed for reliable communication for a given keying rate receiver and antenna. One then selects the appropriate contours that describes this coverage.

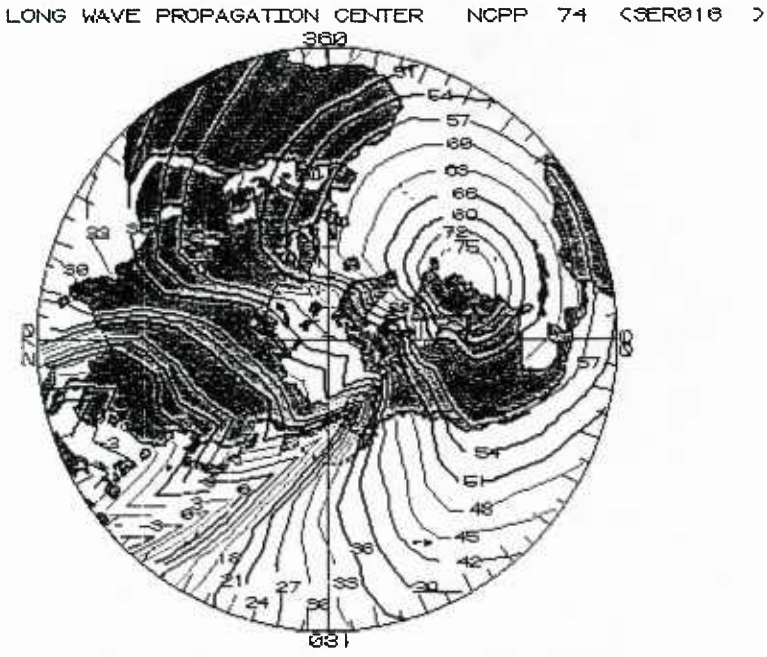


Fig. 113: Signal level contours in dB above one microvolt per meter for the NAA transmitter in Cutler, Maine (Lat 44.60°, Long −67.20°) for 1MW radiated power during July at 24.0 kHz and 90% time availability

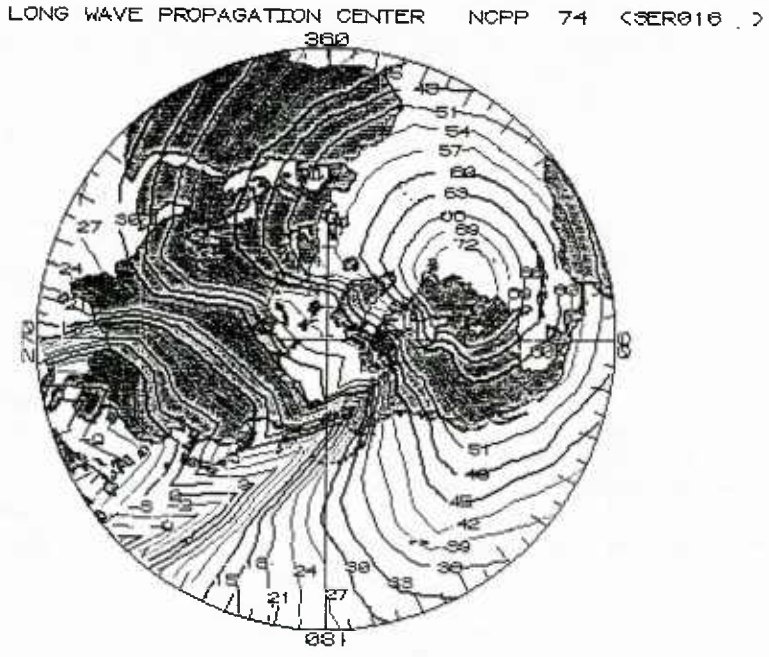


Fig. 114: Signal level contours in dB above one micrvolt per meter for the NAA transmitter in Cutler, Maine (Lat 44.60°, Long −67.20°) for 1MW radiated power during July at 24.0 kHz and 99% time availability

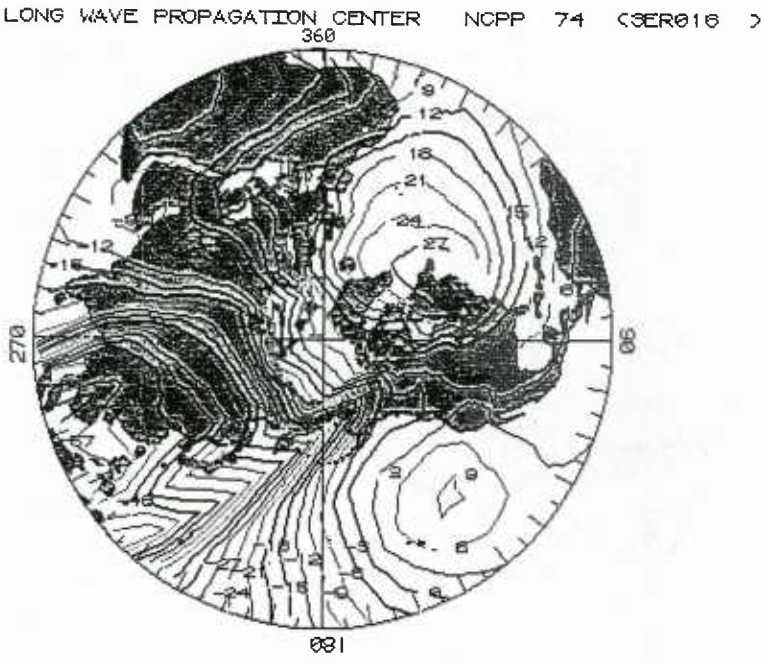


Fig. 115: Signal-to-noise level contours in dB for a 1 kHz bandwidth for the NAA transmitter in Cutler, Maine (Lat 44.60°, Long −67.20°) for 1MW radiated power during July at 24.0 kHz and 90% time availability

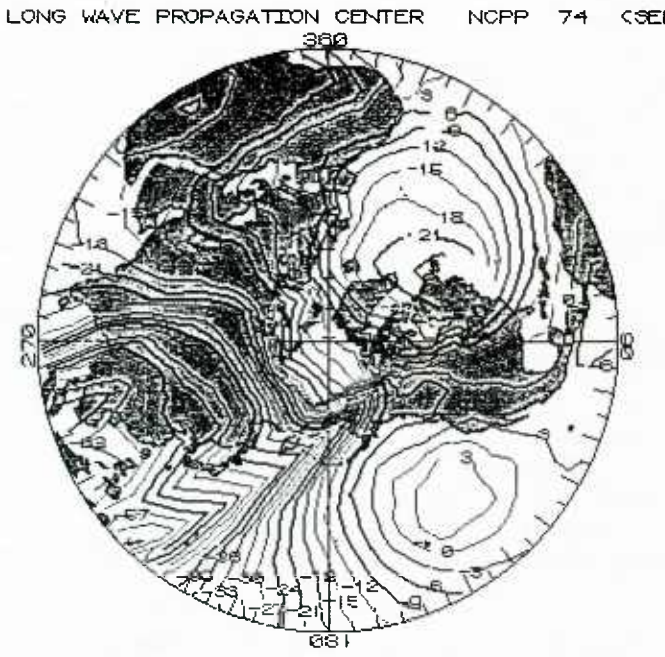


Fig. 116: Signal-to-noise level contours in dB for a 1 kHz bandwidth for the NAA transmitter in Cutler, Maine (Lat 44.60°, Long −67.20°) for 1MW radiated power during July at 24.0 kHz and 99% time availability

ACKNOWLEDGMENTS

I want to express my gratitude to my many colleagues at the Naval Research Laboratory, who have guided and helped me especially W.E. Garner, F.J. Rhoads, D. Baker, J.R. Davis, H. Beck, S. Knowles, J. Goodman, J.A. Murray, J. Brogden, J.P. Hauser, E.S. Byrd and A.J. Martin. I want to thank the engineers at the Defense Communication Agency (R. Heidig, A. Blankfield, G. Jones, Y.S. Fu, P. Bird and P. Sumptor) and at the Department of the Navy (T. Giadrone, R. Savarese, L. Woznak, J. Don Carlos, L. Hice, M. Deeble, and R. Boebel) who have sponsored propagation research. Also I want to express my appreciation to Boyd Burgess of RAE, Farnborough for recommending me as your course lecturer.

REFERENCES

1. LATHAM, Donald C., Assistant Secretary of Defense for Command, Control, Communication and Intelligence, "The C^3 Budget in Perspective," Signal, Vol. 40, No. 9, pp. 107-120, (May 1986).

2. WATT, A.D., VLF Radio Engineering, International Series of Monographs on Electromagnetic Waves, Vol. 14, Pergamon, 1967.

3. WHEELER, H.A., "Fundamental relations in the design of a VLF transmitting antenna," IRE Trans. Antennas Propagat., Vol. AP-6, pp. 120-122, Jan. 1958.

4. Department of the Navy, Naval Electronic Systems Command, Wash., D.C., Naval Shore Electronics Criteria: VLF, LF and MF Communication Systems, NAVELEX Report 101, 113. (Sup. of Doc., U.S. Gov't Printing Office).

5. BLAIR, Bruce G., "Strategic Command and Control and National Security," Signal, Vol. 39, No. 7, pp. 23-32 (Mar., 1985).

6. BLACK, K.M., et al., "TACAMO," Signal, Vol. 33, No. 1, September pp. 6-13.

7. BYSTROM, A., "The Evolution of TACAMO," Signal, Vol. 40, No. 3, pp. 33-40.

8. The entire number of IEEE Transactions on Communications, Vol. com-22, No. 4 (April, 1974) was devoted to the U.S. Navy ELF system.

9. BURROWS, Michael, "ELF Communications Antennas," IEE Electromagnetic Waves, Series 5, Peter Peregrinus Ltd, (Herts, England, 1978).

10. GROSSI, Mario D., "New Technology for ELF Radiators: A Review of Airborne, Rocket-Borne and Space Borne Antennas," in AGARD Conference Proceedings No. 305, "Medium, Long and Very Long Wave Propagation (at frequencies less than 3000 kHz)" J.S. Belrose, Ed. (1982).

11. DAVIES, Kenneth, "Ionosphere Radio Propagation," NBS Monograph 80 (1965) U.S. Govt. Printing Office, Wash., D.C.

12. DeWITT, R.N., F.J. Kelly and G.A. Chayt, "Lower Ionosphere Effects on the Propagation of Waves from an ELF/VLF source in the magnetosphere," Radio Science, Vol. 11, No. 3, pp. 189-197 (Mar 1976).

13. KELLY, F.J., and D.N.M. Rao, "Ionosphere Heating Effects on LF Propagation" in "Effect of the Ionosphere on Radiowave Systems" ed. by J.M. Goodman, F.D. Clarke, J. Aarons (Sup. of Doc., U.S. Govt., Printing Office, Wash., D.C.) based on Ionosphere Effects Symposium held in Old Town, Alexandria, Va. April 14-16, 1981.

14. BELROSE, J.S., "The Oblique Reflection of Low Frequency Radio Waves from the Ionosphere," pp. 149-168 in "Propagation of Radio Waves at Frequencies Below 300 Kc/s" edited by W.T. Blackband, The Macmillan Company (New York) 1964.

15. PITTEWAY, M.L.V., "The Numerical Calculation of Wave-Fields, Reflection Coefficients and Polarizations for Long Radio Waves in the Lower Ionosphere," Phil. Trans. R. Soc. A257, 219 (1965) and G.H. Smith and M.L.V. Pitteway, "Fortran Program for Obtaining Wavefields of Penetrating Nonpenetrating, and Whistler Modes of Radio Waves in the Ionosphere," in ELF-VLF Radio Wave Propagation edited by J. Holtet, D. Reidel Publishing Company (Dordrecht, Holland) 1974.

16. GALEJS, J., Terrestrial Propagation of Long Electromagnetic Waves, Pergamon Press, New York, 1972.

17. PAPPERT, R.A., W.F. Moler, and L.R. Schockey, "A Fortran Program for Waveguide Propagation which Allows for Both Vertical and Horizontal Dipole Excitation," NELC Interim Report 702, 1970.

18. SHEDDY, C.H., R. Pappert, Y. Gough, and W. Moler, "A Fortran Program for Mode constants in an Earth-Ionosphere Waveguide," NELC Interim Report 683, 1968.

19. PAPPERT, R.A., and L.R. Shockey, "A Program for Computing Earth Ionosphere ELF/VLF Excitation Factors for Satellite-borne Antennas," NELC Interim Report 741, 1974.

20. PAPPERT, R.A., "Effects of Elevation and Ground conductivity on Horizontal Dipole Excitation of the Earth-Ionosphere Waveguide," Radio Sci., 5 (No. 3), 579-590 (1970).

21. KELLY, F.J., G.A. Chayt, and D.J. Baker, "Waveguide-Mode Power Budget for an ELF/VLF Transmitting Satellite," NRL Report 8032, Sept. 17, 1976.

22. KELLY, F.J., J.P. Hauser, and F.J. Rhoads, "Computer-Program Model for Predicting Horizontally and Vertically Polarized VLF Atmospheric Radio Noise at Elevated Receivers," NRL Report 8479, Dec. 28, 1981.

23. WAIT, J.R., and K.P. Spies, "Characteristics of the Earth-Ionosphere Waveguide for VLF Radio Waves," NBS Technical Note 300, 1964.

24. BROOKES, C.M., J. McCabe, and F.J. Rhoads, "Theoretical VLF Multimode Propagation Predictions," NRL Report 6663 (1967).

25. KELLY, F.J., F.J. Rhoads, and I.P. Hansen, "On Statistical VLF Multimode Propagation Prediction," NRL Report 7239, (1971).

26. KELLY, F.J., F.J. Rhoads, and I.P. Hansen, "Statistical Multimode VLF Field Strength Predictions," in Agenda and Abstracts: Symposium on Long Wave Propagation 19, 20 and 21 January 1971, Menlo Park, California," NRL Memorandum Report 2202 (1971).

27. GARDINER, R.S., "Comparison of Predicted VLF/LF Signal Levels with Propagation Data," Defense Communications Agency, 960-TP-74-5 (21 Jan. 1974).

28. REED, W.G., "Determination of Effective Ionosphere Electron Density Profiles for VLF/LF Propagation," Defense Communications Agency, C650-TP-76-4, 1 Jan. 1976.

29. MORGAN, R.R., "World-Wide VLF Effective-Conductivity Map," Westinghouse Electric Corporation Report 8013 F-1, 1968.

30. HAUSER, J.P., W.E. Garner, and F.J. Rhoads, "A VLF Effective Ground Conductivity Map of Canada and Greenland with Revisions Derived from Propagation Data," NRL Report 6893, Mar. 4, 1969.

31. SWARM, H.M., D.K. Reynolds, and A.W. Biggs, "Analytical Study of ELF and VLF Radiation Fields from a Horizontal Dipole in Antartica," University of Washington College of Engineering, Dept. of Electrical Engineering, Technical Report No. 87 (Jan., 1964).

32. WAIT, J.R., (1962a), An analysis of VLF mode propagation for a variable ionosphere height, J. Res. NBS 66D (Radio Prop.), No. 4, 453-461.

33. WAIT, J.R., (1962c), Mode conversion in the earth-ionosphere waveguide, NBS Technical Note No. 151.

34. PAPPERT, R.A., and L.R. Shockey, Effects of Strong Local Sporadic E on ELF Propagation, NOSC Technical Report 282, 15 Aug. 1978.

35. PAPPERT, R.A., and R.R. Smith, Mode Conversion Program for an Inhomogenious Isotropic Ionosphere, NELC Interim Report No., 701, 1 April 1970.

36. PAPPERT, R.A., and L.P. Shockey, "WKB Fields Program for Lower ELF," NELC Interim Report No. 731, 2 Jan. 1973.

37. PAPPERT, R.A., and L.R. Shockey, "A Simplified Mode Conversion Program for VLF Propagation in the Earth Ionosphere Waveguide," NELC Interim Report 751, 1 October 1974.

38. HAUSER, J.P., F.J. Rhoads, and F.J. Kelly, "VLFACM Program Description and Operational Manual," NRL Report 8530, November 24, 1981.

39. HAUSER, J.P., and F.J. Rhoads, "Coverage Predictions for the Navy's Fixed VLF Transmitters," NRL Memorandum Report 2884, Sept. 1974.

40. WAIT, J.R., (1961), "A diffraction theory of LF sky-wave propagation," J. Geophys. Res., 66, No. 6 (1961).

41. WATSON, C.N., Proc. Roy. Soc. (London) A95, pp. 83 and 546.

42. HOLLINGWORTH, J., "The Propagation of Radio Waves," J. IEEE, 1926, 64, pp. 579-589.

43. BREMMER, H., Terrestrial Radio Waves (1949) New York, Elsevier Publishing Co., Inc.

44. WAIT, J.R., and A.M. Conda, "A Diffraction Theory for LF Sky-Wave Propagation-An Additional Note," Journal of Geophys., 66, (1961) 1725-1729 (see also pp. 1713-1724).

45. BERRY, L.A., G. Gonzalez, and J.L. Lloyd, "Wave Hop Series for an Anisotropic Ionosphere," Radio Science, 4, 1969, pp. 1025-1027.

46. BERRY, L.A., and J.E. Herman, "A Wave Hop Propagation Program for an Anisotropic Ionosphere," OT/ITS Research Report 11, April 1971.

47. BERRY, L.A., and J.E. Herman, "Fortran Program for Estimating LF-VLF Radio System Performance," OT TM 63, Office of Telecommunications, Institute for Telecommunications Sciences, Boulder, Colorado (Oct. 1971).

48. JOHLER, J.R., "Spherical Wave Theory for MF, LF and VLF Propagation," Radio Science, 5, (1970).

49. BERRY, L.A., G. Gonzalez, and J.L. Lloyd, "Wave Hop Series for an Anistropic Ionosphere," Radio Science, 4, No. 11 pp. 1025-1027.

50. JONES, T.B., and K. Mowforth, "A Review of the Analytical Techniques for Determining the Phase and Amplitude of a VLF Radio Wave Propagation in the Earth-Ionosphere Waveguide," AGARD Conference Proceedings No. 305, "Medium, Long and Very Long Wave Propagation (At Frequencies Less Than 3000 kHz)," edited by J.S. Belrose, February 1982.

51. MORFITT, D.F., and R.F. Halley (1970), "Comparison of waveguide and wave hop techniques for VLF propagation modeling," NWC Tech. Publication 4952, Naval Weapons Center, China Lake, Calif.

52. HUFFMAN, J.C., Lauaber, F.J. LeBlanc, and C.I. Meng, "Ultraviolet Remote Serving of the Aurora and Ionosphere," paper 4-4, in Effect of the Ionosphere on C^3I Systems ed. by J.M. Goodman, F.D. Clarke, J.A. Klobucher and H. Soicher Based on Ionospheric Effects Symposium held in Old Towne, Alexandria, Va. (1-3 May 1984).

53. WAIT, J.R., (1958), "Propagation of very-low-frequency pulses to great distances," J. Res. NBS, 61(3), 187-203.

54. WAIT, J.R., (1965), Propagation of pulses in dispersive media, J. Res. NBS, (Radio Sci.) 69D(1), 1387-1401.

55. KELLY, F.J., "Multimode and Dispersive Distortion in the Very-Low-Frequency Channel," Radio Science, Vol. 5, No. 3, pp. 569-573, March 1970.

56. JACKSON, J.D., (1962), Classical Electrodynamics, 1st ed., John Wiley and Sons, Inc., New York.

57. KELLY, Francis J., "VLF field strength variations from an airborne trailing-wire antenna," Radio Science, Vol. 5, No. 5, pp. 785-791, May 1970.

58. KELLY, F.J., J.P. Hauser, H.M. Beck, and F.J. Rhoads, "Multipath VLF Propagation Effects on Correlation Receivers," AGARD Conference Proceeding No. 305, Medium, Long, and Very Long Wave Propagation (at Frequencies less than 3000 kHz) ed. by J.S. Belrose.

59. WINGATE, B., "The Very Low Frequency Experimenter's Newsletter," 121 Chesapeake Avenue, Lake Hiawatha, N.J. 07034.

60. WESTERLUND, S., F.H. Reder, and Abom, C., "Effects of Polar Cap Absorption Events on VLF Transmissions" Planet and Space Sci., 17, pp. 1329-1374.

61. DALLAND, J., K.W. Eriksen, B. Landmark, and H.J. Weedon "The Influence of Polar Blackouts on VLF Circuits," NDRE Report 36, April 1961.

62. CROMBIE, D.D., (1964), "Periodic Fading of VLF Signals Reveived Over Long Paths at Sunrise and Sunset," Radio Science J. Res. NBS 68D, No. 1, 27-34.

63. CROMBIE, D.D., (1966), June Radio Science.

64. CROMBIE, D.D., "The Waveguide Mode Propagation of VLF Radio Waves to Great Distances," Conference on MF, LF and VHF Radio Propagation, 8-10 November 1967, pp. 138-150.

65. WAIT, J.R., (1968), "Mode Conversion and Refraction Effects in the Earth Ionosphere Waveguide for VLF Radio Waves," J. Geophys. Res., 73, 3537 (1968).

66. SCHNEIDER, A., and P.K. Carlston, (1969), Radio Science, 4, 377-380.

67. BICKEL, J., (Private Communication).

68. HOFFMAN, W.C., (1960), "Some Statistical Methods of Potential Value in Radio Wave Propagation Investigations," in Statistical Methods in Radio Wave Propagation, edited by W.C. Hoffman, 117-135, Pergamon Press, New York, 1960.

69. RHOADS, F.J., and W.E. Garner, (1965), "An Investigation of the Modal Interference of Very-Low-Frequency Radio Waves," NRL Report 6359, 27 October 1965.

70. BANNISTER, P.R., et al., "Scientific and Engineering Studies: Extremely Low Frequency (ELF) Propagation," Naval Underwater Systems Center, Newport Laboratory, Newport, Rhode Island.

71. HELLIWELL, R.A., (Private Communication).

72. KENNEL, C.F., and H.E. Petschek, "Limit on Stably Trapped Particle Fluxes," J. Geophys. Res., 71, 1, (1966).

73. BRICE, N., "Harnessing the Energy of the Radiation Belts," J. Geophys. Res., 176 No. 19 (1971).

74. TOLSTOY, A., and T. Rosenberg, (Private Communication).

75. POTEMRA, T.A., and A.J. Zmuda, "Precipitating Energetic Electrons as an Ionization Source in the Midlatitude Nighttime *D* Region," J. Geophys. Res.,Vol. 75, No. 34, (1970).

76. LARSEN, T.R., T.A. Potemus, W.L. Imhof, and J.B. Reagan, "Energetic Electron Precipitation and VLF Phase Disturbances at Middle Latitudes Following the Magnetic Storm of December 16, 1971," J. Geophys. Res., Vol.10, (1977).

77. LARSEN, T.R., J.B. Reagan, and W.L. Imhof, L.E. Montriand, and J. Belrose, S., "A Coordinated Study of Energetic Electron Precipitation and Region Electron Concentrations Over Ottawa during Disturbed Conditions," J. Geophys. Res., Vol. 81, No. 13, (May 7, 1976).

78. MONTBRIAND, L.E., and J.S. Belrose, "Changes in Electron Precipitation Inferred From Spectra Deduced From *D* Region Electron Densities During a Post-Magnetic Storm Effect," J. Geophys. Res., 81, No. 13, (May 1976).

79. IMHOF, W.L., R.R. Anderson, J.B. Reagan, and E.E. Gaines, "The Significance of VLF Transmitters in the Precipitation of Inner Belt Electrons," Jour. of Geophy. Res., pp. 11225-11234 (1981).

80. HARANG, L., "The Aurorae," John Wiley and Sons, Inc. (New York, 1951).

81. WHITTEN, R.C., and I.G. Poppoff, "Physics of the Lower Ionosphere," Prentice-Hall, Inc. (Englewood Cliffs, New Jersey, 1965).

82. MENG, C.I., "Electron Precipitation in the Midday Auroral Oval," Jour. Geophys. Res., 86, No. A4, pp. 2149-2174 (Apr. 1981).

83. MENG, C.I., "The Auroral Electron Precipitation During Extremely Quiet Geomagnetic Conditions," Jour. Geophys. Res., Vol. 86, No. A6, pp. 4607-4627, (June 1981).

84. GOODMAN, J.M., and D.R. Uffelman, "The Role of the Propagation Environment in HF Electronic Warfare," NRL Memo. Rpt. 4953, Nov. 1982.

85. GOODMAN, J.M., "A Survey of Ionospheric Models A Preliminary Report on the Development of an Ionospheric Model Thesaurus and Users Guide," NRL Memo. Rpt. 4830, July, 1982.

86. GALEJS, J., "F layer reflections and ion effects in the propagation of terrestrial ELF waves," J. Geophys. Res., Vol. 75, pp. 2529-2539, 1970.

87. GALEJS, J., "Terrestrial Propagation of Long Electro-magnetic Waves," New York, Pergamon (1972).

88. DAVIS, J.R., E.L. Althouse, and D.R. Uffleman, "Some possible propagation-associated constraints on ELF communications," NRL Report 7269, July 1971.

89. PAPPERT, R.A., "Effects of a large patch of sporadic-E on night-time propagation at lower ELF," Journal of Atmospheric and Terrestial Physics, Vol. 42, pp. 417-425. Pergamon Press Ltd. 1980. Printed in Ireland.

90. DAVIS, J.R., "ELF Propagation on Northern and Mid-Latitude Paths," in "ELF-VLF Radio Wave Propagation" ed. by J.A. Holtel (D. Reidel Publishing Co., Dordrecht, Holland 1974), pp. 263-277.

91. PAPPERT, R.A., "Broadside excitation of ELF by a Horizontal Dipole Beneath a Sporadic E. Environment: Theory" Radio Science, 21, pp. 106-116 (Feb. 1986).

92. FIELD, E.C., C.R. Warber, and R.G. Joiner, "Focusing and Shadowing of ELF Signals," Radio Science, 21, pp. 511-517 (May-June 1986).

93. BAIN, W.C., "The Use of VLF Propagation Results in Ionospheric Modelling," in ELF-VLF Radio Wave Propagation ed. by J.A. Holtel (D. Reidel Publishing Co., Dordrecht-Holland, 1974), pp. 151-162.

94. RASMUSSEN, J.E., P.A. Kossey, and E.A. Lewis, "Evidence of an Ionospheric Reflecting Layer Below the Classical Region" Jour. of Geophs Res., Vol. 85, No. A6, pp. 3037-3044, June 1980.

95. OKHLOPKOV, V.P., L.S. Okhlopkova, and T.N. Charakhch'yan, Two-Year Cosmic-Ray Variations, Geomagnetism and Aeronomy, Vol. 19, No. 3, 1979.

96. SAMADDAR, S.N., "Weather Effects on Loran-C Propagation," Navigation: Journal of the Institute of Navigation, Vol. 27, (Spring 1980) pp. 39-53.

97. OTT, R.H., L.W. Vogler, and G.A. Hufford, "Ground Wave Propagation over Irregular, Inhomogeneous Terrain: Comparisons of Calculations and Measurements," NTIA Report 79-20.

98. DOHERTY, R.H., and J.R. Johler, "Meteorological Influences on Loran-C Groundwave Propagation," J. Atmospheric and Terrestrial Physics, Vol. 37, pp. 1117-1124, 1975.

99. DEAN, W.N., "Diurnal Variations in Loran-C Groundwave Propagation," Proc. Ninth Annual Precise Time and Time Interval Applications and Planning Meeting," NASA - Goddard Space Flight Center, Greenbelt, Md., 1977.

100. DEAN, W.N., "Study of Temporal Behavior of Loran-C to A Data," Feb. 15, 1979.

101. HOROWITZ, S., and J.R. Johler, "Loran C/D Coordinate Prediction Dependence on Ground Electrical Properties," AGARD Conf No. 209 on Limitations of Navigation and Positioning Systems, Istanbul, Turkey, 20-22 Oct. 1976.

102. JOHLER, J.R., "Prediction of Ground Wave Propagation Time Anamolies in the Loran-C Signal Transmissions over Land," AGARD Conference No. 209 on Limitations of Navigation and Positioning Systems, Istanbul, Turkey, 20-22 Oct. 1976.

103. MILLINGTON, G., and N. Elson, "Ground-wave Propagation Across a Land/Sea Boundary," Nature, 164, pp. 114-116 (1949).

104. BEAN, B.R., and E.J. Dutton, "Radio Meteorology," National Bureau of Standards Monograph 92, p. 18 (U.S. Govt. Printing Office, 1966).

105. SAMADDAR, S.,(Private Communication).

106. CRICHLOW, W.Q., C.A. Samson, R.F. Disney, and M.A. Jenkins, "Quarterly Radio Noise Data," NBS Technical Note 18 series, 1960.

107. "World Distribution and Characteristics of Atmospheric Radio Noise," Proc. of the CCIR Xth Plenary Assembly, Geneva, 1963, Report 322, International Telecommunications Union, Geneva, 1964.

108. MAXWELL, E.L., D.L. Stone, R.D. Croghan, L. Ball, and A.D. Watt, "Development of a VLF Atmospheric Noise Prediction Model," Westinghouse Georesearch Laboratory Report 70-1H1-VLF NO-RI, 1970.

109. HAUSER, J.P., and F.J. Rhoads, "Analysis of a VLF Atmospheric Noise Prediction Model," paper given at USNC/URSI 1973 Meeting Aug. 21-24, Commission VIII, Session 1.

110. HAUSER, J.P., "Further Analysis of a VLF Atmospheric Noise Model," paper given at USNC/URSI 1974 Meeting, Oct. 14-17, Commission VIII, Session 2.

111. KELLY, F.J., J.P. Hauser, and F.J. Rhoads, "Atmospheric VLF Radio Noise at Elevated Receivers: Horizontal and Vertical Polarization," AGARD Conference Proceedings, No. 305, Brussels, September 1981.

112. KELLY, F.J., J.P. Hauser, and F.J. Rhoads, "Computer-Program Model for Predicting Horizontally and Vertically Polarized VLF Atmospheric Radio Noise at Elevated Receivers," NRL Report 8479, December 1981.

113. HAUSER, J.P., F.J. Kelly, and F.J. Rhoads, "Atmospheric Radio Noise Modelling at NRL," in Effect of the Ionosphere on C^3I Systems ed. by J.M. Goodman, F.D. Clarke, J.A. Klobucher and H. Soicher Based on Ionospheric Effects Symposium held in Old Towne, Alexandria, Va., (1-3 May 1984) pp. 375-386.

114. GINSBERG, L.H., "Extremely Low Frequency (ELF) Atmospheric Noise Level Statistics for Project Sanguine," IEEE Transactions on Communications, Vol. Com-22, No. 4, (Apr. 1974) pp. 555-561.

115. EVANS, J.E., and A.S. Griffiths, "Design of a Sanguine Noise Processor Based upon World Wide Extremely Low Frequency (ELF) Recordings," IEEE Transactions on Communications, Vol. COM-22, No. 4, (Apr. 1974) pp. 528-539.

116. DAVIS, J.R., and W.D. Meyers, "Observations of ELF Signal and Noise Variability on Northern Latitude Paths," NRL Report 7923, Nov. 11, 1975.

117. MEYERS, W.D., and J.R. Davis, "ELF Nonlinear Noise Processing Experimental Measurements: Part 1- Preliminary Results in July and October 1975," NRL Memorandum Report 3226, Mar. 1976.

118. DAVIS, J.R., and W.D. Meyers, "ELF Nonlinear Noise Processing Experimental Measurements, Part 2- Synoptic Sample and Seasonal Noise Variation in Norway," NRL Report 8039, Oct. 29, 1976.

119. GOLDSTEIN, J.A., W.D. Meyers, and J.R. Davis, "ELF Nonlinear Noise Processing Experimental Measurements: Part 3 - Synoptic Sample of Diurnal and Seasonal Noise Variation in Italy," NRL Memo Report 3543, July 1977.

120. ALEXANDER, J.K., and M.L. Kaiser, "Terrestrial Kilometric Radiation ....," Journal of Geophysical Research, 81, 5948-5956 (1976), 82, 98-104 (1977).

121. BERNSTEIN, S.L., M.L. Burrows, J.B. Evans, A.S. Griffiths, D.A. McNeill, C.W. Niessen, I. Richer, D.P. White, and D.K. Williams (1974), "Long Range Communications at Extremely Low Frequencies," Proc. IEEE, 62, 292-312.

APPENDIX I
THE IMPORTANCE OF THE ELF/VLF/LF BAND BAND

The Ground Wave Emergency Network (GWEN) will provide a nuclear hardened terrestrial communications system to carry warning data and emergency action messages and force direction orders to U.S. forces. When the first phase is operational in FY 87, GWEN will be a unique, low-cost, hardened communications system serving both national and nuclear CINC requirements. The second phase of the program will add more transmitters and receivers, and will provide a hardened link to the ICBM force and other strategic elements. We expect to procure about 50 more nodes for the network in FY 87 and to complete the system by the end of the decade.

To improve communications connectivity with strategic bombers in flight, we plan to equip them with Miniature Receive Terminals (MRTs) for receiving low frequency (LF) and very low frequency (VLF) communications. Although slower than transmissions in higher frequency bands, LF/VLF communications can be transmitted over much greater distances than can line-of-sight communications, and they are much less susceptible to nuclear effects and jamming than are existing satellite and high frequency (HF) communications. Flight tests of the terminals are planned to be completed in FY 87; production should be completed in early FY 88.

To communicate with ballistic missile submarines, the Navy maintains special radio relay aircraft continuously airborne over the Atlantic and Pacific oceans. These planes, called Take Charge and Move Out (TACAMO), transmit relayed messages to the submarines by a variety of communications media. In FY 89, we will begin replacing the EC-130s (modified C-130 transports) now flying the TACAMO mission with the faster, more survivable, more endurable and longer range E-6A (a derivative of the Boeing 707 airborne warning and control system [AWACS] airframe). At the same time, the SSBN force will be outfitted with improved VLF receivers. The E-6 A's ability to operate over vast expanses of the ocean will permit SSBNs and nuclear cruise missile equipped submarines to expand their operating areas and still receive messages from the President. Building toward a fleet of 15 E-6As, we are requesting funds to procure the third, fourth and fifth production aircraft in FY 87. Although the planes will be outfitted initially with communications equipment transferred from EC-130s, they eventually will carry a vastly improved VLF communications system.

Submerged submarines now must deploy an antenna at or close to the ocean's surface to receive messages, thus potentially increasing their susceptibility to detection. The use of extremely low frequency (ELF) communications, which can penetrate to great depths, will alleviate this constraint. The ELF Communications System consists of two transmitter sites (in Wisconsin and northern Michigan), operating in electrical synchronization with receivers aboard submarines. The ELF system's high reliability and continuous transmissions significantly will upgrade peacetime communications to deployed submarines and will support the transition to wartime operations. Receivers are being installed in some submarines this year so that the full system can be tested in FY 87. We are requesting funds for FY 87 to procure receivers for other submarines. The system should be fully operational by the end of FY 88.

From "The C^3 Budget in Perspective" by Donald C. Latham, Assistant Secretary of Defense for Command, Control, Communications and Intelligence.

HF GROUNDWAVE AND SKYWAVE PROPAGATION

BY

KLAUS-JUERGEN HORTENBACH
DEUTSCHE WELLE, Cologne
Federal Republic of Germany

SUMMARY

The principles of HF ground- and skywave propagation are reviewed with regard to regular conditions as well as anomalous phenomena which may affect the signal characteristics and thus may have an influence on the performance of digital communication systems. Particular consideration is given to the statistics of signal fading as a basis for system design.

1 INTRODUCTION

In the high frequency range electromagnetic energy can be propagated by ground- or skywaves. Ground wave propagation is more effective the lower the frequency, providing very stable signals with, however, relatively short coverage ranges. Furthermore, there are difficulties in constructing efficient antennas. Skywaves provide long-distance propagation but the signal can be most unstable. The ionosphere, being the propagation medium of skywaves, is subject to the most extensive variations and the most severe limitations imposed by nature. Caused by solar control, there are regular variations with the time of the day, the season and the 11-year's sunspot cycle. Superposed to these regular variations short term variations occur which up to date are not yet predictable with sufficient reliability. Moreover, a large variety of special propagational phenomena may become effective.

The quality of a link is usually described in terms of the link availability and the errors produced during a transmission of information which in digital transmission is normally characterised by the bit-error-rate. We may distinguish three different propagational phenomena which may become the main sources of bit errors:

- received field strength
- noise and/or interference
- signal fading

The received field strength must be known in relation to the noise and interference level because the signal-to-noise ratio is one of the main factors governing system design. The value to be required depends strongly on the type of service and the modulation used. The received field strength depends on the attenuating properties of the propagation medium and on the transmitting and receiving equipment, e.g. transmitter power, transmitting and receiving antennas but not on the type of signal.

Noise and interference are one of the main degrading factors. We may distinguish three types of noise:

- natural noise which in the HF range mainly originates in the earth's atmosphere from lightning discharges and is propagated by the ionosphere,
- man made noise, produced by industrial equipment and ignition
- internal noise of the receiving equipment which at HF is usually negligible against external noise.

Interference is produced by dislocated users of the same or the adjacent frequency channels and may be intended as in the case of jamming or not. The effectiveness of interference depends on the conditions of the propagation path between the interfering source and the receiver location.

The third degrading factor, the signal fading, is produced by the propagation medium. Fading may be non-selective, causing short- or long-term outages by decreasing the signal-to-noise ratio of the entire signal spectrum, or frequency selective, causing signal distortions of analog signals by affecting different parts of the signal spectrum and intersymbol interference in digital transmissions.

In section 2 we shall discuss the splitting of the field radiated by an antenna on the ground into ground waves and sky waves. Section 3 gives a review of ground wave propagation and section 4 of sky wave propagation. Particular consideration is given to the effects of fading on the signal.

2 THE RADIATION FIELD OF AN ANTENNA ON THE GROUND

The radiation field of an antenne in the vicinity of flat ground is described by Sommerfeld's theory [1] which was extended to a spherical earth by van der Pol and Bremmer [2]. Norton [3] has shown that the exact solution of the problem can be split into two parts which can be interpreted as a spacewave term and a groundwave term. This splitting is possible only in a formal way as both wave types taken separately are no solutions of the wave equation. Nevertheless it gives a good insight into the mechanism of wave propagation.

According to Norton, the space wave is the resultant of a direct wave and a ground reflected wave. The amplitude od the direct wave decreases inversely with distance as in free space. Amplitude and phase of the ground reflected wave depend on the reflecting properties of the ground as given by Fresnel's reflection coefficients and on the difference in path length between direct and reflected wave. Fresnel's reflection coefficients are functions of polarisation, wave frequency, angle of incidence and the electrical properties of the ground, characterised by its relative dielectric constant and conductivity, the latter two being in turn dependent upon the moisture content and the temperature of the ground. The ground reflection does not take place at a single reflection point but rather within an area surrounding the geometrical reflection point. This area is usually considered to be the first Fresnel zone, i.e. an area within which contributions from all other points lag in phase by 180 degrees or less. Contributions from within the zone reinforce each other but contributions from points outside the zone differ widely in phase and tend to cancel. Thus, the contributions from within the zone are mainly responsible for the reflected wave. Undulations of the ground and obstacles within the zone give rise to scattering of the incident energy thus leading to a weakening of the energy reflected into the specular direction. This loss is accounted for by the scattering coefficient by which the Fresnel reflection coefficients are multiplied or by using appropriately modified values of the ground constants.

The third component of the wave field, the ground wave term in Norton's solution arises because the earth is no perfect reflector. Some energy is transmitted into the ground where it sets up ground currents associated with an electromagnetic field which is called groundwave.

Norton's groundwave is not equivalent to and should not confused with Zenneck's surface wave. Zenneck's surface wave can not be realised physically, nor is it contained in Sommerfeld's exact solution for the wave propagation over a plane earth with finite electrical ground constants. The Zenneck wave has the same rotating electrical field as has Norton's wave but, at large distances, Zenneck's wave is attenuated exponentially and has a phase velocity which depends in the ground constants and is greater than the velocity of light, whereas Norton's wave is attenuated with the inverse of the distance squared and has a phase velocity which is equal to the velocity of light and independent on the ground constants [4].

3 GROUNDWAVE PROPAGATION

When the heights of both the transmitting and the receiving antennas are small compared to the wavelength the spacewave disappears at an elevation angle of zero degrees, i.e. in the direction tangential to the earth's surface, as the direct and the ground reflected waves have equal amplitudes and opposite phases and so give zero resultant. In this case the ground wave has its maximum. When the antenna heights are increased there is a further phase change of the reflected ray due to the extra raypath length so that the space wave and the reflected wave no longer cancel. Hence the resultant signal must be computed as the vector sum of the direct, the ground reflected and the surface wave. The effect is modelled by the so-called height-gain factor applied to the groundwave field strength. The heigt-gain factor increases with frequency and can be significant in the HF range for typical antenna heights. Fig.1 gives the vertical radiation pattern of a short vertical antenna showing the contributions of the spacewave and the groundwave. The cosine diagramme is obtained for an infinitely conducting surface. For this case the spacewave term represents the solution of the wave equation while the groundwave term disapears.

3.1 Groundwave Propagation Over Smooth Homogeneous Earth

The space wave decreases with the inverse of the distance, the ground wave with the inverse of the distance squared. Hence at large distances and non-zero elevation angles the strength of the space wave exceeds that of the groundwave. The ground wave is diffracted to follow the curvature of the earth. The diffraction decreases with the ratio of the wavelength to the radius of the obstacle. Thus, ground waves are stronger the lower the frequency. Diffraction is also influenced by the imperfect conductivity of the ground. Energy is absorbed by currents induced in the earth so that energy flow takes place from the wave downwards. This produces wavefront tilting which assists in the bending of the wave. The loss of energy dissipated in the earth leads to attenuation dependent on conductivity and dielectric constant. With horizontal polarisation the wave attenuation is still stronger than with vertical polarisation due to the different behaviour of Fresnel's reflection coefficients for both polarisations.

For the fieldstrength calculation different approaches are necessary depending on the range of distances involved:

- Close to the transmitter the field near the surface can be described by the Sommerfeld flat earth theory with corrections applied for the effective radius of the earth. This approach is valid out to distances of the order $10\,\lambda^{1/3}$ km and heights of the order of $35\,\lambda^{2/3}$ m, where λ is the wavelength in metres.

- In the line-of-sight region the field is composed of the sum of the direct and reflected waves. These are found from an integral representation by the method of stationary phase which is equivalent to geometrical optics.

- The diffraction region occurs at ranges greater than about $10\,\lambda^{1/3}$ km near the surface or beyond the radio horizon for elevated terminals. The field is represented by the residue or waveguide mode series. Associated with each mode is a propagation constant which determines the variation of the mode with distance, and a height gain function which determines the variation with height. Solutions in closed form are possible only for a troposphere with linear height variation of the refractive index. For an exponential troposphere numerical integration must be applied.

To facilitate calculations for practical purposes the CCIR has published sets of groundwave propagation curves showing the vertical fieldstrength component of the radiation field produced by a short vertical monopole versus distance for several types of surfaces like wet and dry ground, sea and fresh water and ice [5]. Fig.2 shows one of these curves for relatively wet land with a ground conductivity of 0.03 S/m and a dielectric constant of 30. The curves refer to a smooth homogeneous spherical earth and antenna heights which are small compared to the wavelength. A tropospheric refractive index, decreasing exponentially with height, has been taken into account (see Sect.3.3).

3.2 Groundwave Propagation Over Mixed Paths

The curves shown in Fig.2 relate to a combination of ground constants valuable for the entire path, a condition rarely met in practice. For an all-land path with variable ground constants, sufficient accuracy can normally be obtained by using the propagation curves appropriate to weighted mean values of these constants. This approach is not sufficiently accurate, however, for mixed land- sea paths because of the marked differences in the constants. This case is characterised by recovery effects in the regions of transition from land to sea, i.e. an increase of the fieldstrength despite of increasing distance, and an intensified decrease at the transitions from sea to land. For this case an empirical method is recommended. In this, sections of curves corresponding to the length of path traversed with different sets of ground constants are combined to produce a continuous single curve of the received signal strength. The strength should be independent of direction of propagation, and so the process is repeated for propagation in the reverse direction and a mean of the two sets of results taken. This method may also be used for propagation over land where different conditions apply on separate sections of the path [5].

Other methods involve the solution of integral equations by means of a computer [6]. These methods also allow for a given path profile different from smooth spherical earth to be taken into account.

3.3 Tropospheric Refraction

At HF the troposphere has some effects on groundwave propagation. Assuming a spherical earth surrounded by concentric refracting layers this refraction can be taken into account by giving the earth an effective radius greater than the true radius. Based on a refractive index changing exponentially with height the effective earth radius is 4/3 of the real radius for frequencies above about 10 MHz. Below 10 MHz the effective radius decreases and reaches a value of 5/4 at 1 MHz [7]. As mentioned before, the effect of refraction has been included in the propagation curves in [5].

3.4 Electrical Characteristics Of The Surface Of The Earth

If groundwave fieldstrengths are to be calculated, it is essential to know the ground characteristics along the path. The electrical characteristics of any medium can be expressed by three parameters: the permeability, the permittivity and the conductivity. The permeability of the ground can normally be regarded as equal to the free space value so that in most propagation problems we are only concerned with permittivity and conductivity.

As electromagnetic waves penetrate the earth's surface, the sub-surface structure of the ground down to levels corresponding to the penetration depth of the wave is also essential for wave propagation. The penetration depth is defined as the depth at which the wave has been attenuated to 1/e (37%) of its value at the surface. It is a function of frequency and varies widely with the type of ground. At 10 MHz it ranges from about 10 cm for sea water up to some hundreds of meters for fresh water ice

(Fig.3,[8]).The sub-surface structure is rarely homogeneous, but rather consists of several layers of different thickness and different conductivities and permittivities. If the penetration depth is less than the thickness of the layer, the underlying strata have little influence. If the penetration depth is much greater than the top layer thickness, propagation is also determined by the electrical characteristics of the lower strata. If the conductivity of an underlying strata is less than that of the top layer, the propagation may be better than would be with homogeneous ground of good conductivity. On the other hand, an underlaying layer of higher conductivity than that of the top layer may deteriorate the propagation. The effect of multiple layers can be taken into account by introducing effective parameters representing an equivalent homogeneous structure when using the propagation curves in [5].

Among the factors which determine the effective electrical characteristics of the ground are [9]

- Nature of the soil: variations seem to be due not so much to the chemical composition of the soil as to its ability to absorb and retain moisture. At 10 Mhz the conductivity varies between about 0.02 mS/m for fresh water ice and 5 S/m for sea water. The relative permittivity varies from about 3 for fresh water ice or very dry ground up to 80 for fresh or sea water (Fig.4,[8]).

- Moisture content: the moisture content seems to be the major factor determining the electrical constants. However, at depths of one meter or more, the wetness of the soil at a particular site seems to be substantially constant over the seasons of the year. Although it may increase during rain, the drainage of the soil and surface evaporation soon reduces it to its normal value. Seasonal variations of the groundwave fieldstrength have been shown by measurements to be negligibly small.

- Temperature: At low frequencies the temperature coefficient of the conductivity of soil is in the order of 3% per degree Celsius, while that of the permittivity is negligible. As the range of temperature variation during the year decreases rapidly with depth temperature effects are likely to be important only when the ground is frozen to a considerable depth.

- Frequency: In the HF range the electrical characteristics vary little with frequency, with the exception of the conductivity of pure water and fresh water ice.

- General geological structure: As the effective electrical characteristics along a path are determined not only by the surface soils but also by that of the underlying strata, it is of great importance to have knowledge of the general geological structure of the area concerned. For this reason world maps of the ground conductivity have been prepared, e.g. in [10].

3.5 Irregular Terrain, Vegetation

Terrain irregularities and surface objects cause wave reflections, diffraction and absorption. If the variations in terrain height about some mean value is small in comparison to a wavelength and the slopes of the terrain are small, then the propagation curves [5] like Fig.2 for smooth earth can be used if the actual terrain surface impedance is replaced by a higher effective value. Large objects like isolated mountain ridges can be treated by the knife-edge diffraction theory [11]. As was mentioned in section 3.2 computerised methods have been developed which permit fieldstrength calculations not only for inhomogeneous but also for irregular terrain [6].

Propagation in wooded terrain can lead to additional absorption due to currents being induced in the trees which varies with season depending on the amount of foliage and its humidity and snow coverage. At HF the attenuation is of the order of 0.01 dB/m [12].

3.6 Delay Dispersion Of Ground Wave Signals

The ground wave is normally considered to be very stable in amplitude and phase and thus is regarded to be a very reliable carrier of information. However, even the groundwave exhibits a certain amount of time dispersion which limits the bandwidth of a signal and should therefore be considered especially in digital communications.

The largest bandwidth which can be supported without excessive distortion is defined as the coherence bandwidth. It is roughly equal to the inverse of the time dispersion (delay smear) of a narrow pulse. When a narrow pulse, i.e. a wideband signal, propagates over a smooth, nonperfectly conducting surface, there are several mechanisms which result in time dispersion of the transmitted signal:

- Amplitude and phase of the groundwave depend on the impedance of the surface which varies with frequency. The phase therefore varies nonlinearly with frequency which results in time dispersion of a transmitted pulse as each frequency component will propagate with a different group velocity.

- If the propagation distance is sufficiently long, diffraction and refraction

around the spherical earth will cause some time dispersion as these effects are also frequency dependent.

- When the surface is rough, the radio energy received at a point does not travel solely by the direct path from the transmitter, but some of the energy radiated off the great circle path is scattered towards the receiver. The scattered ground waves arrive from all azimuthal directions, with a spread in the time of arrival and are referred to as the diffuse component of the groundwave field.

In a recent publication [13] the time dispersion was derived to be expected for propagation over rough surfaces and, more specifically, over a rough ocean surface for which the roughness was characterised in terms of the gravity wave spectrum. The results are presented as a function of distance and sea state. It could be shown that at short distances surface roughness tends to increase the amount of delay dispersion. Beyond a certain distance, it depends on roughness: a small-scale roughness surprisingly causes a reduction of the delay dispersion, a large-scale roughness causes increased delay dispersion as expected.

The delay spread of a narrow pulse propagated over a rough ocean surface was found to be less than 200 nanoseconds for distances up to 300 km and wind speeds up to 30 kn. This implies that the coherence bandwidth of the groundwave signal is in excess of 5 MHz which would allow the transmission of 500-kHz-signals with little distortion.

4 HF SKYWAVE PROPAGATION

In chapter 2 we dealt with the division of a radiated electromagnetic field into a groundwave component and a spacewave, consisting of a direct and a ground reflected wave. A space wave at a frequency within the HF range, travels along a rectilinear propagation path (apart from some tropospheric bending) until it enters the ionosphere where it may be reflected back to the earth. A wave which has been reflected by the ionosphere is commonly called skywave. In this chapter we shall consider skywaves and their propagation medium, i.e. the ionosphere and its influence on a signal.

The ionosphere affects a wide range of frequencies, from ELF waves which are reflected at the lower bounds of the ionosphere up to ESF waves when penetrating the ionosphere. The most critical range is the HF range, i.e. the range between 3 and 30 MHz, as in this range the transition from reflection to penetration occurs.

4.1 Formation And Structure Of Ionospheric Layers.

The ionosphere is the ionised region in the earth's atmosphere extending from about 50 km to roughly 2000 km above the surface. The principal source of ionisation is UV- and X-ray radiation from the sun producing free electrons and ions from the low-density neutral atmospheric gases. The ionisation rate depends on the intensity of solar radiation and thus is a function of geographic location, time of the day, season and the phase of the 11-year's sunspot cycle. The production of free ionisation is counterbalanced by ionisation loss processes, principally the collisional recombination of electrons and positive ions and the attachment of electrons to neutral gas atoms and molecules. Due to the interactions of these processes with the density and the constituents of the atmospheric gases the ionisation density varies with height thereby forming different discernible peaks between which the ionisation density varies smoothly. We may distinguish three main peaks which we call layers and which are identified by the letters D, E and F in order of increasing altitude. Subdivisions of these layers may occur under certain coditions, e.g. F1 and F2 layers (Fig.5).

The D-region spans the approximate altitude range 60 to 90 km. The ionisation varies strictly with the sun's zenith angle being maximum at noon with typical electron densities of 10^8 to 10^{10} per cubic meter and nearly disappearing at night. There is a pronounced seasonal variation with a maximum in summer. In the winter, often anomalously high ionisation appears, probably resulting from changes in the neutral atmosphere composition. The D-region ionisation is too low for reflecting HF waves. Nevertheless, the penetrating wave excites oscillations of the free electrons which collide with ions and neutral particles the masses of which are much larger than that of the electrons. By this process absorption of wave energy takes place. Thus, during the daylight hours the D-region is the main source of attenuation in skywave propagation.

The altitude range of the E-layer is between 90 and 130 km with a maximum electron density of approximately 10^{11} per cubic meter at about 110 km. It also shows a strong solar zenith angle dependence with maximum density near noon, in summer and during solar cycle maximum. At night, only a small residual level of ionisation remains which is able to reflect MF- but not HF-waves. The E-layer playes an important role for short-distance communications below 2000 km, longer distances are bridged nearly exclusively by F-layer reflections.

The F-region extends upward from about 130 km. During the daylight hours it may be subdivided into the F1- and the F2-layer. The F1-layer is the region between 130 and 210 km altitude with a maximum electron density of about $5 \cdot 10^{11}$ per cubic meter. The ionisation follows the sun's zenith angle as does the E-layer, but with a different dependence. The F2-layer is the highest ionospheric layer with typical electron

densities exceeding 10^{12} per cubic meter at daytime and $5 \cdot 10^{10}$ per cubic meter at night. It departs markedly from solar angle dependence since it is strongly influenced by winds, diffusion and other dynamic effects. Among the most notable features of this layer are the maximum electron density occurring at times other than solar noon , the noon densities in winter exceeding the corresponding summer values, an 'equatorial anomaly' showing significant latitudinal gradients of electron density within 20 to 30 deg either side of the geomagnetic equator and a F-region 'trough' - a pronounced depression in electron density extending some 5 to 10 degrees equatorwards from the auroral oval on the night side of the earth.

The topside of the ionosphere ranges from the F2 layer peak up to about 2000 km. It contains about two thirds of the total ionisation of the ionosphere and is important for transionospheric communications between satellites or spacecraft and ground stations.

4.2 Regular Skywave Propagation

4.2.1 Principles Of Ionospheric Reflection -

The ability of the ionosphere to provide propagation support is related singly to the condition that its refractive index at radio frequencies is different from that in free space. Neglecting the geomagnetic field and particle collision the ionospheric refractive index as a function of frequency f is given by

$$n^2 = 1 - (f_h/f)^2 \qquad (1)$$

where f_h is the plasma frequency given by

$$f_h = \frac{N_h e^2}{4\pi^2 \epsilon_o m} = 80.5\, N_h \qquad (2)$$

with N_h = electron density
e,m = charge and mass of electron
ϵ_o = permittivity of free space

As the refractive index is a monotonically decreasing function of the height above the earth's surface, the phase front of an incident wave will propagate at a given height with larger phase velocity than at lower heights. Consequently, the ray is bent down towards the earth. The ray path can be calculated by geometrical optics using Snell's law

$$n_h \sin \varphi_h = \text{const.} \qquad (3)$$

where n_h and φ_h are the refractive index and the angle of incidence at any given layer height h. At the point of incidence with the angle φ_o we have n=1 and at the summit point $n=n_h$ and $\varphi_h = 90°$ (Fig.6). Applying Snell's law we find

$$n_h = \sin \varphi_o \qquad (4)$$

Inserting (4) into (1) and solving for f we obtain

$$f = f_h \sec \varphi_o \qquad (5)$$

This is the well-known secant law which is a relation between the frequencies at vertical and oblique incidence and the angle of incidence for all ray paths which are reflected at the height level h. The frequencies in (5) are called equivalent.

As the phase velocity of a wave in the ionosphere is less than in free space, a virtual path with a virtual maximum height h' can be defined (Fig.6). According to the theorem of Breit and Tuve the delay time of a radio wave on its curved path TBR through the ionosphere with the true height h is equal to the time delay of a wave travelling on the triangular path TAR with the velocity of light. Thus, the virtual height h' is the magnitude which is measurable by ionospheric sounding techniques.

At a certain height H the electron density in a layer reaches a maximum N_H. With (2) the highest frequency which can be reflected at vertical incidence by a layer with maximum electron density N_H is $f_{\perp,max} = f_H \sim \sqrt{N_H}$. Applying the secant law (5) we obtain the corresponding maximum frequency which can be propagated obliquely with an angle of incidence φ_o

$$f_{\wedge,max} = f_{\perp,max} \sec \varphi_o \qquad (6)$$

Higher frequencies penetrate the layer and do not return to the earth.

From ionograms we have the virtual height $h' = h'(f_\perp)$ given as a function of frequency at vertical incidence. It shows normally an increase of the virtual height with frequency until the layer is penetrated at the critical frequency. Using the above mentioned relationships we can now determine the corresponding curve for oblique incidence. This is done by use of the 'transmission curves'.

From Fig.6 we have

$$h' = \frac{D/2}{\tan \varphi_o} = \frac{D/2}{\sqrt{\frac{1}{\cos^2 \varphi_o} - 1}}$$

Using the secant law we obtain

$$h' = \frac{D/2}{\sqrt{(f/f_\perp)^2 - 1}} \qquad (7)$$

These transmission curves $h'(f_\perp)$ relate the virtual height h' with the equivalent frequency $f_\perp$ for vertical incidence where the path length D and the wave frequency f are parameters. Equating (7) to the ionogram, which is another function $h' = h'(f_\perp)$ we find the possible propagation path with the corresponding virtual heights, ground ranges and transmission frequencies. This can be done graphically as is shown in Fig.7 for a constant ground range. From the intersection of both curves we obtain a new curve $h' = h'(f_\wedge)$ with the following properties:

- it is shifted entirely to higher frequencies,
- the highest frequency at vertical incidence does not correspond to the highest frequency at oblique incidence,
- the curve has two branches joining at the highest frequency being reflected obliquely (junction frequency).

The two branches correspond to two different angles of elevation and are termed 'high-angle ray' or 'Pedersen-ray' and 'low-angle ray'. The latter is reflected at comparably low height, the former penetrates deeper into the ionosphere, comes closer to the ionisation maximum where the gradient of the refractive index is small and hence travels a long distance in the ionosphere. Increasing the frequency both paths approximate each other until they coincide at the junction frequency.

The relations discussed so far are strictly valid only for a flat earth and a flat ionosphere. For use with a spherical earth and ionosphere correction factors have to be introduced.

Fig.8 shows the relation between elevation angle and ground range for a fixed frequency which is greater than the critical frequency at vertical incidence. For low elevation angles the propagation path is long, the ground range is large and the penetration depth is small. With increasing elevation angle, the range decreases and the ray penetrates deeper into the layer until a minimum possible distance is reached (path no.6), which is called skip distance. Within this distance no reception by regular reflection is possible. The skip distance increases with the wave frequency and in the limit for a sufficiently high frequency can extend to the maximum ground range possible for rays launched at grazing incidence; in that case all rays escape into space. On the other hand, when the wave frequency is decreased down to the critical frequency for vertical incidence, all rays are reflected and the skip zone disappears.

It follows for a fixed point of reception that there is a maximum frequency at which the waves can be reflected to it. This is the frequency at which the ground range is equal to the skip distance. This frequency is the maximum usable frequency for this distance which is identical with the junction frequency in Fig.7.

When the elevation angle is further increased beyond the value which corresponds to the skip distance, the ground range increases again rapidly by virtue of the high angle rays which are reflected from a greater height and have a significant length of near-horizontal path close to apogee (nos.7,8,9). The ground range of such rays may well be in excess of 7500 km in temperate regions and 10000 km in equatorial regions [14]. The band of elevation angles providing high-angle rays is usually only a few degrees. As can be seen from Fig.8 this small bundle of rays is spread over a wide distance range which results in high spatial attenuation. So the strength of the high-angle rays tends to be less than that of the low-angle rays. This is partially compensated by less absorption as the D-layer is not penetrated. Thus, the high-angle ray may have appreciable amplitude which may give rise to interference with low-angle rays. The high-angle ray paths can exist until the reflection is from the layer maximum. Rays with larger angle of elevation penetrate the ionosphere and escape into space.

4.2.2 Multipath Propagation -

So far we have considered only single-hop single-layer propagation. In practice reception is by a multitude of rays propagating on different paths. Multipath propagation can arise from

- low- and high-angle paths,
- multi-hop propagation,
- multi-layer propagation,

- magnetoionic splitting

where the splitting into low- and high-angle paths has already been discussed in the preceding section. Generally, these effects are existent in combination.

4.2.2.1 Multi-hop Propagation -

Due to the curvature of the earth the range attainable after one ionospheric reflection is limited. The maximum range attainable arises from rays launched at grazing incidence and depends principally on the layer height of maximum electron density and, to a less extent, on the form of the model ionosphere and on the wave frequency. For typical E-, F1- and F2-layers, the maximum range is 2000, 3400 and 4000 km respectively. These values represent upper limits when account is taken of the poor performance of practical antennas at low elevation angles.

Distances beyond the values given can be bridged by consecutive reflections between the ionosphere and the earth's surface. On one hand, the possible number of hops is limited by increasing signal attenuation which is mainly due to D-layer absorption and ground reflection losses. On the other hand, the elevation angle increases with the hop number for a given ground range, resulting in a decrease of the path MUF. Generally, the geographic variations of the ionospheric characteristics at the different ionospheric reflection regions have to be taken into account.

4.2.2.2 Multi-layer Propagation -

During the daylight hour a regular E-layer is present and propagation is possible via the E- and/or F-layer. We have seen that the determining factors as to whether reflection or penetration occurs are the ionisation of the layer, the wave frequency and the angle of incidence. A given ground range yields different incidence angles for different layer heights. The greater the layer height, the steeper the angle of incidence to achieve propagation to a fixed range, and therefore the lower the MUF. This means that although the critical frequency of the E-layer is less than that of the F1-layer which in turn is less than that of the F2-layer, the E-MUF can be the greatest of these three separate layer-MUF's. This is most likely to be the case in the summer daytime at low solar activity over path ranges of 1000 to 2000 km. Again, the F1-MUF may exceed the F2-MUF beyond the maximum E-range at distances of 2000 to 3000 km. For propagation between a pair of fixed terminals the path MUF is the greatest of the individual MUF's for reflection from the different layers. For a particular propagation path to be possible, the wave frequency must be below the path-MUF and, in the case of F-modes, also the E-layer must not screen it. Screening of the 1F2-mode, but not of the 2F2-mode because of the lesser path obliquity, is a common summer daytime occurrence at certain frequencies.

The strongest mode on a long path is usually the lowest possible order F2-mode unless the antenna discrimitates against this. Higher-order F2-modes traverse the ionosphere a greater number of times to suffer more absorption and ground reflection losses, so that they tend to be weaker. A given range can be spanned by fewer F- than E-hops. Modes involving more than two reflections from the E-layer are rarely of importance. Reflections from the F1-layer arise only under restricted conditions and the 1F1 mode is less common than the 1E and 1F2 modes. The 1F1 mode can be important at ranges of 2000 to 3400 km, particularly at high latitudes. Multiple-hop F1 modes are very rare in practice because the necessary ionospheric conditions to support an F1-layer reflection do not tend to occur simultaneously at separated positions.

Because of the large distances involved with multiple-hop paths the ionospheric conditions will not be constant along the entire path. The geographical changes in ionisation cause so-called mixed modes with successive reflections from different layers. Mixed modes are a common feature of transequatorial paths and east-west paths across a daylight-darkness boundary. A catalogue of possible propagation paths is given in Fig.9, showing single-layer modes (a,b), which can be reflected by either the F- or the E-layer, and mixed modes (d-f). When the wave frequency exceeds the E-layer MUF only slightly, the wave does not penetrate the layer along a rectilinear path but will be bent downwards thus introducing a path asymmetry (Fig.9c, or Fig.8, paths nos.10-14). Fig.9 g and h show tilt supported propagation paths which shall be discussed in section 4.4.1.

4.2.2.3 Magnetoionic Splitting -

In all our preceding discussions we have neglected completely the influence of the magnetic field of the earth on radio wave propagation. It must be born in mind that the equivalence theorems discussed earlier are valid only for this case unless correction terms allowing for the magnetic field are introduced. Because of the presence of the geomagnetic field, the ionosphere is a doubly refracting medium. Magnetoionic theory shows that a ray entering the ionosphere is split into two separate waves due to the influence of the magnetic field which are called ordinary (o-) and extraordinary (x-) wave. The o-wave is refracted less than the x-wave, becomes reflected from a greater height and so has lower critical frequency and MUF. Both waves experience different amounts of refraction so that they travel independently along somewhat displaced raypaths with different phase velocities. The polarisations

of both waves are generally elliptical with the o-wave having an anticlockwise sense of vector rotation viewed in the direction of propagation for an angle of less than 90 degrees between the ray and field directions. For these conditions the vector rotation of the x-wave is clockwise. The rotation senses reverse for greater angles between the ray and field directions. The o-wave suffers less absorption and so is usually the stronger and hence the more important one. Its path is close to that which would arise if there were no field. For many practical purposes it is sufficient to take only the o-wave into account.

4.2.3 Propagation Losses -

In the preceding sections we were concerned with the ability of the ionosphere to support wave propagation and have found that the usable frequency range has an upper limit which is called MUF. Next we shall review the factors governing the magnitude of the received fieldstrength. On the way from the transmitter to the receiver the energy of a wave suffers different kinds of losses which are frequency dependent and lead to a lower limit of the usable frequency range. We shall review in the following the most important types of losses which are governing the received fieldstrength.

Spatial attenuation arises from the spreading of the power flux over an increasing area during signal propagation. In free space the power flux density is inversely proportional to the square of the path length. For skywaves, however, ionospheric refraction causes focussing and defocussing. Defocussing is associated with signals propagated by high-angle rays (Sect.4.2.1). Focussing occurs at frequencies close to the MUF due to the refractional conditions in the layer but also at low elevation angles and at near-antipodal distances. The latter effects arise due to the spherically shaped ionosphere acting as a concave mirror. The focussing can be derived from the ray path geometry [15]. Often, the effect is covered by other losses but it may become a considerable contributiuon to the signal strength on quasi-antipodal paths, where focussing gains of some 10 to 20 dB have been observed [16].

Ionospheric absorption is one of the most important factors influencing the received skywave signal strength. As is shown by the magnetoionic theory, a wave travelling through the ionosphere suffers absorption which arises from collisions between the free electrons, oscillating under the influence of the incident wave, and the neutral and ionised particles present in the ionosphere. As the absorption is proportional to the product of electron concentration and collision frequency, most absorption occurs in the D- and lower E-region. Additional absorption arises near the height of reflection where the refractive index is small. This is known as deviative absorption in contrast to the non-deviative absorption which occurs during the penetration of the layer. The absorption is proportional to the inverse of the wave frequency squared, thus providing a lower limit of the usable frequency range. Absorption is low at nighttime because of the reduced D- and E-region ionisation. The non-deviative absorption reaches a maximum around local noon in the summer, but the influence of the deviative absorption can modify the resultant seasonal variations. At temperate latitudes in winter, the non-deviative absorption becomes anomalously high and subject to greater day-to-day variability than would be expected from a normal solar control. This effect is referred to as winter anomaly. Especially high absorption can arise in auroral zones due to increased electron concentrations produced by energetic electrons incident during ionospheric disturbances (Sect.4.3.2). Absorption losses have a very wide span of magnitude, reaching in practice from near-zero values up to quasi infinity.

The losses of multi-hop modes associated with reflections on the ground are determined by the Fresnel reflection coefficients. Ground reflection losses depend on the conductivity and the relative dielectric constant of the ground, the wave polarisation,the frequency and the elevation angle. The order of magnitude is up to about 10 dB per hop for land and 1 dB for sea water. Additional losses are due to the spreading of wave energy by the convex shape of the reflecting surface of the earth. This divergence is, however, counteracted by ionospheric focussing. Further losses can arise from scattering when the reflecting surface is very rough and the elevation angle not too low.

Because of the magnetoionic splitting in the ionosphere the polarisation of the downcoming wave differs from that of the upgoing wave and is generally elliptically polarised (Sect.4.2.2.3). As the receiving antenna is usually linearly polarised, only a fraction of the wave power is coupled to the receiving antenna. Polarisation coupling losses are dependent on the geomagnetic field and on the angle between the field and the direction of propagation and can exceed 20 dB under certain conditions [17].

After the different propagation losses have been considered, the signal strength is still dependent on the gains and the radiation characteristics of the transmitting and receiving antennas. In system design the aims are to choose antennas with beam-widths and directions of maximum gain most appropriate to the principal propagation path encountered. This will be the path with the lowest losses, which is generally the path with the lowest possible number of hops. However, with a pair of antennas given, the stronger signal may be received on a higher-order mode despite of the higher losses associated with that path, when the antennas are highly directive and so oriented as to favour these modes by higher gain values.

As the propagation losses are almost all frequency dependent with the ionospheric absorption being the most restrictive one, the lower frequencies are generally more attenuated than the higher ones. It is therefore appropriate to define a lowest usable frequency (LUF) which will provide a specified level of fieldstrength or signal-to--noise ratio at a specified percentage of the time. Both LUF and MUF give the lower and upper limits of a useful frequency range which can be used for a radio link between given terminal locations. Contrary to the MUF, the LUF not only depends on ionospheric characteristics but also on operational equipment like antennas and transmitter power.

4.2.4 Variations Of The Usable Frequency Range -

The useful frequency range defined in the previous section represents average conditions only and is subject to change with geographic location, time of the day, season and solar activity. Superposed to these regular variations there are short- and long-term statistical variations of different origin (sect.4.5). To account for these variations the probability distribution function of MUF and LUF have to be determined. Alternatively, a magnitude called 'circuit reliability' may be specified which provides a unique measure of the statistical variations of the useful frequency range. The circuit reliability may be defined as the joint probability that a propagation path exists and that the signal-to-noise ratio achieved exceeds a specified value. For the prediction of the usable frequency range atlases of ionospheric characteristics [18] and computer programs have been established by the CCIR [19,20,21]. Fig.10 shows typical variations of the usable frequency range as given by the circuit reliability with the time of the day, the season and the sunspot activity.

4.3 Irregular Phenomena

In the preceding section we considered ionospheric propagation by regular layers under undisturbed conditions which is highly predictable by the determination of monthly median values of the usable frequency range and expected fieldstrength. Among the irregular propagation phenomena we may discern ionospheric irregularities with statistical occurrence and ionospheric disturbances which are attributable to solar events.

4.3.1 Ionospheric Irregularities -

Various physical processes generate irregular sructures in the ionosphere with strong temporal and geographical variations. Even under magnetically quiet conditions, the ionosphere exhibits day-to-day variations due to changes in the ionising ultra-violet flux from the sun and the solar wind and its interactions with the magnetosphere and ionosphere. The total effect is very complex and the complete causal chain has not yet been established.

Superimposed on these variations are sometimes variations with a time scale of the order of 10 minutes to a few hours due to travelling ionospheric disturbances (TID's) which are in fact the ionospheric signature of gravity waves in the neutral atmosphere. Large TID's have been found to originate at high latitudes in association with large geomagnetic disturbances while smaller ones would seem to be of local meteorological origin. TID's can change foF2 by 1 or 2 MHz and introduce multiple Doppler components which result in long period fading and consequent decrease in circuit quality. However, the effects of TID's can often be removed by the use of either space or polarisation diversity.

Sporadic E (Es) is an anomalous ionisation with statistical occurrence at E-region altitudes which is considerably above normal E-layer ionisation and may be strong enough to reflect even VHF waves. It may appear sometimes irregular and patchy, sometimes smooth and disk-like and has little direct relationship to solar ionizing radiation. At temperate latitudes sporadic-E occurs most frequently in summer daytime with a maximum in the mid-morning hours and near sunset. The seasonal minimum is near mid-winter or the spring equinox. In auroral latitudes it is most prevalent at night. The ionisation is distributed in patches ranging in spatial extent from several kilometers to 1000 km and with thickness typically 500 to 2000 m. A comprehensive world-wide survey of sporadic-E can be found in [22]. Propagation by sporadic-E may become beneficial at times when the regular E- and F2-layers do not exist. Thus it may extent the useful frequency range but, on the other hand, cause interference by unwanted transmitters which would not occur under normal conditions. Moreover, Es-layers can screen normal F-modes, thus interrupting long-distance circuits. A method for calculating sporadic-E fieldstrength has been developed and is recommended by the CCIR [23].

In ionospheric sounding it is often observed that an echo pulse reflected from the F-region has a much longer duration than the transmitted pulse. Because of its appearance on ionograms the phenomenon is called 'spread F'. It is caused by the scattering of the signal from irregularities embedded in the ionosphere. Individual irregularities may have scale-sizes from about 3 m to more than 10 km but also entire patches of irregularities with dimensions of some 100 km may exist. The irregularities tend to be field-aligned and travel with velocities of some tens to some hundreds of m/s. The strong gradients of the electron density within the irregularities give rise to scattering or total reflection of HF waves off the great circle propagation path between transmitter and receiver and to a large time delay dispersion of transmitted

pulses. Movements of the irregularities incidently may simulate moving targets on a direction finder or trans-horizon radar. Interference between propagation paths influenced by irregularities moving with different velocities generates severe flutter fading. Spread-F is essentially an evening and nighttime phenomenon. At low latitudes the maximum of occurence is between about 2100 and 0100 local time, at middle latitudes after midnight. In low latitudes the occurence is greater in summer than in winter, in higher magnetic latitudes it is more frequent in winter than in summer.

4.3.2 Ionospheric Disturbances -

Ionospheric disturbances are produced by solar flares emitting electromagnetic and corpuscular radiation into the space. In the earth's atmosphere this radiation causes several phenomena affecting radio propagation in various ways. Among those affecting HF propagation the most important are short wave fadeouts (SWF), also called 'sudden ionospheric disturbances' (SID), geomagnetic and ionospheric storms and polar cap absorption (PCA).

Short wave fadeouts are produced by X-ray and UV radiation penetrating into the D-region and increasing the electron density. This results in a sudden increase of absorption which may last from several minutes to hours. Short wave fadeouts occur only on the sunlit side of the earth and are most frequent at solar maximum. The effect depends on the frequency as well as on the length of the path through the absorbing region. Generally, higher frequencies are less affected.

Geomagnetic storms are disturbances of the geomagnetic field produced by charged solar particles arriving with a delay time of 20 to 40 hours after the onset of a solar flare. This modifies the existing system of ionospheric and magnetospheric currents and gives rise to geomagnetic and, indirectly, ionospheric storms. Ionospheric storms are characterised by a depression of the daytime critical frequency of the F2-layer, an increase of absorption up to a total radio blackout at high latitudes, enhanced spread-F and sporadic-E. During sunspot maximum years the storms tend to be shorter in duration (2 to 3 days) and more severe while during sunspot minimum they usually last longer with less intensity. All effects are most pronounced in the auroral zone and vanish almost completely near the equator.

As well as geomagnetic and ionospheric storms caused by solar flares, there exists a class of storms known as recurrent storms because of their marked tendency to recur after a solar rotation (27 days). This class of storms is generated by high-speed solar wind streams emitted from coronal holes and travel at several times the velocity of the background solar wind. Recurrent storms are a feature of the declining phase of the solar cycle, when both the coronal holes and their associated wind streams can last up to 10 solar rotations. The ionospheric effects of recurrent storms seem to be essentially the same as those of flare-induced storms except that they tend to be somewhat less severe and to last longer.

Polar cap absorption events are caused by solar protons which originate in major solar flares. The influx of high-energy protons into the relatively unshielded polar regions considerably increases D-region ionisation resulting in strong absorption of HF radio waves leading to a complete radio blackout in polar regions. PCA's start 15-200 minutes after the flare and usually last between 1 and 3 days. The frequency of occurrence is highly correlated with sunspot activity.

4.4 Other Propagational Phenomena

Besides ionospheric irregularities and disturbances there exist several propagational phenomena which may become important under certain conditions by increasing or decreasing the useful frequency range and the quality of an HF radio communication link.

4.4.1 Ionospheric Tilts -

Ionospheric tilts are horizontal gradients in electron density acting as a tilt of the reflecting ionospheric mirror. Such gradients may be caused by variations in layer height and/or critical frequency. Tilts have the effect of making the path asymmetric between transmitter and receiver, with the reflection point shifted towards a point of higher density, thereby increasing the effective MUF. Other effects are great-circle deviations being most severe when the path is tangential to the auroral oval. There is some evidence that rays being refracted by a tilted layer do not return to the ground but travel above the earth's surface until re-entering the ionosphere to undergo further refraction back to the ground ('super modes', Fig.9,g-h). It is believed that this effect could be one of the reasons for unexpectedly high fieldstrengths being observed on long-distance paths, e.g. with round-the-world echos.

4.4.2 Guided Propagation -

Guided propagation is the limiting case of super mode propagation when the distance between successive ionospheric reflections becomes infinitesimally small. It has been shown by theory and proved by satellite experiments that this mode of propagation is possible instead of just sky-wave multi-hop path propagation [14]. The concavity of the reflecting ionosphere can act as a waveguide with either a single boundary within the F-layer or being double-walled by the F- and the E-layer. Full wave mode theory has shown that the MUF for ground-to-ground communications is increased by a factor 1.5 if such a 'whispering-gallery' propagation is guided by the F-layer and by a factor 2.6 for a day-time E-region duct. For a day-time F-region with a maximum plasma frequency of 9 MHz the attenuation of round-the-world signals is only about 11 dB. The wave coupling into and out of the duct may be accomplished by irregularities like ionospheric tilts, sporadic-E, equatorial F-region troughs, meteor trails etc. Like the supermodes also the waveguide propagation mode is believed to provide surprisingly high fieldstrengths on long-distance paths.

Another form of ducting is along ionisation irregularities aligned with the earth's magnetic field. This was suggested to explain long-delay echos ($t<0.4$ s) in HF radar observations [14].

4.4.3 Scatter Propagation -

A radio wave propagated via the ionosphere may be partially scattered by the irregularities of the ground or sea or by those of the ionosphere itself. Side-scatter may cause deviations up to some 10 degrees off the great-circle propagation path. Ground side-scatter can be observed on multihop paths with the lowest path loss occurring for bearings towards an intersection of the skip distances around the transmitter and receiver because of skip distance focussing. Ionospheric scatter is produced by irregularities and tilts, especially near the auroral zone. The side--scatter modes of propagation can lead to an extension of time beyond which a given frequency can be used on the great-circle path and also lead to the use of frequencies higher than the MUF, e.g. within the skip zone of a transmitter. Communications may, however, be degraded by the fading produced by interfering signal components. The ground-scatter mode can maintain communications during disturbances; this mode often avoids propagation through regions of abnormally high absorption. A survey of ground and ionospheric side- and back-scatter can be found in [24].

4.5 Fading

Since the ionosphere is no stable medium, amplitude and phase of skywave signals fluctuate very distinctly with reference to time, space and frequency. These effects, collectively described as fading, have a decisive influence on the performance of radio communication systems. It is essential to know the fading characteristics to be able to specify transmitter power, protection ratios and the most effective type of modulation or coding.

4.5.1 General Causes Of Fading -

Fading may be caused by several different effects giving rise to different fading phenomena. Among the most common types of fading are:

- Absorption fading is caused by solar flare events (see sect.4.3.2) and affects especially the lower frequency bands. It is the slowest type of fading which may last for some minutes up to more than one hour.

- Focussing and defocussing: Ionospheric irregularities can produce wavelike structures in the normally stratified layers. If the curvature of the deformed layer is concave with respect to the ground, an incident bundle of rays will be focussed at a receiving location. If the curvature is convex, the rays will be dispersed. Motion of these structures gives rise to fading with periods up to some minutes.

- Polarisation fading is caused by Faraday rotation. As was shown in section 4.2.2.3 a plane polarised wave entering the ionosphere is split up by the earth's magnetic field into two elliptically polarised components. Both components can interfere to yield an elliptically polarised resulting wave. Because of random fluctuations in the electron density along the propagation paths there will be continuous changes in the direction of the major axis of the ellipse of polarisation. Since a linearly polarised antenna gives a maximum output voltage if the electric field vector is parallel and zero output if it is perpendicular to it, the receiver input voltage will change according to the spatial rotation of the ellipse of polarisation. The fading period is in the order of a fraction of a second to a few seconds.

- Interference fading: the deepest and most rapid fading occurs from beating between two or more signal components of comparable amplitude, propagating along different paths. This multipath fading may arise from the interference of multiple-mode and multiple-layer propagated rays, high- and low-angle modes,

ground and skywaves or scattered waves. Interferences between skywaves and the ground wave occur in the vicinity of the transmitter location where the amplitudes of both wave components can have the same order of magnitude. A regularly reflected signal together with a signal scattered from spread-F irregularities can give rise to so-called flutter fading with rates of about 10 Hz.

4.5.2 Fading Depth -

For the purpose of practical system design it is not always essential to identify the individual contributory effects. Instead it is more appropriate to characterise the fluctuations of signal levels in time as a random process. It is expedient to consider short term and long term fading separately.

The short term fading component includes phase interference between multipath propagation components, and rapid variations of signal strength caused by ionospheric irregularities. A sampling period of an hour has been found suitable for short term fading.

The long term fading arises from random changes in the short term median values of a signal, e.g. the variation in the median fieldstrength measured from day to day over a fixed hour. Diurnal, seasonal and solar-cycle variations are of systematic nature and are usually not associated with fading.

The fading depth of a continous wave signal can be described by the signal amplitude distribution functions. These functions obtained analytically to describe the envelope of a fading signal, differ according to the different assumptions made with respect to the structure of the contributory signals. Among the most frequently used models for multipath propagation is the Rice-Nakagami probability distribution function which assumes that the received signal is composed of a steady sinusoidal component and a random Rayleigh component with comparable amplitudes and a uniform phase probability density. When the steady component disappears the Rice-Nakagami function approaches the Rayleigh distribution, when the Rayleigh component becomes small it reduces to the Gaussian distribution function. The fading depth, defined as the difference (in dB) between the signal levels exceeded for 10% and 90% of the time, has been confirmed by measurements to be in the order of 13.4 dB expected for the Rayleigh distribution, although the distribution may differ from Rayleigh. This value does not appear to vary greatly with path length in the range 1500 - 6000 km, with time of day or with season. However, at high signal levels the fading range has been observed to fall below the Rayleigh value, possibly due to a strong constant component term arising from a specular reflection, and distributions of the Rice-Nakagami type will apply under these conditions.

In the case of long term variations, i.e. the variations of hourly median values over a month, the log-normal distribution gives a good approximation of the actual situation. It has been shown by measurements that the standard deviation of this distribution is between 5 and 10 dB, values exceeding 10 dB may be found in polar regions. Long term fading does not appear to vary greatly with path length, time of day, season or sunspot number. There is, however, a significant dependence on the ratio of wave frequency to monthly median MUF.

If the distribution of the instantaneous values about the median value during each individual time is Rayleigh and the daily median values are normally distributed, then the total probability distribution is given by Fig.11, showing the level below which the signal falls for given percentage of total time [25].

4.5.3 Rapidity Of Fading -

The rapidity of fading may be measured by the level crossing rate, i.e. the number of positive crossings per unit time through any specified level. The duration of a fading period is the time during which the receiving amplitude falls below a certain level relative to the mean receiving level. Fig.12 shows the number and Fig.13 the duration of fades measured on some specific circuits [26]. it can be seen that short term fades are much more frequent then long term fades. The time between two successive fades is generally several times longer than the fading period. Consequently, The transmission errors are not distributed stochastically in time, but periods with a high error density (error bursts) alternate with periods of low error density.

It seems that equatorial paths may show more rapid fading than paths confined to higher latitudes. Seasonal and sunspot-cycle variations occur but may differ with path orientation. Long distance paths seem to be affected more at the equinoxes and at sunspot maximum.

4.5.4 Fading Allowance For Planning Purposes -

For planning purposes the monthly median estimates of wanted and unwanted signals have to be augmented by appropriate fading allowances. These depend on

- the within-an-hour and day-to-day fading of the wanted signal
- the within-an-hour and day-to-day fading of the background (noise and/or interfering signals)
- the correlation between the wanted signal and background strength, and
- the fraction of the time for which the desired reception quality must be achieved.

According to CCIR [25] the fading allowance necessary to achieve the desired reception quality for 90% of the time is given by the root-sum-square of the following quantities:

- wanted signal lower decile deviation from the hourly median field strength arising from within-an-hour changes (dB),
- wanted signal lower decile deviation from the monthly median field strength arising from day-to-day changes (dB),
- background upper decile deviation from the hourly median field strength arising from within-an-hour changes (dB),
- background upper decile deviation from the monthly median field strength arising from day-to-day changes (dB),

where both the within-an-hour and day-to-day fading of the wanted signals and the background are assumed to be uncorrelated.

4.5.5 Fading Influence On Modulated Signals -

The rapidity and depth of fading as described in the prceding sections characterise the amplitude fluctuations of an unmodulated wave. Amplitude fading is accompagnied by associated fluctuations in group path and phase path, giving rise to time and frequency dispersion of modulated signals. When the transmitter or receiver is moving or there are ionospheric movements the received signal is also Doppler frequency shifted.

4.5.5.1 Time Dispersion -

In ionospheric propagation the main origin of time dispersion is multi-path propagation arising from differences in transit time between different propagation paths as was discussed in section 4.2.2. The multipath spread causes amplitude and phase variations in the signal spectrum due to interference of the multipath wave components. If these fluctuations are correlated within the signal bandwidth the fading is called non-selective or flat, because all spectral components of the signal are affected in nearly the same way. If the fluctuations are uncorrelated the fading is frequency selective. Time dispersion is characterised by its delay power spectrum or the frequency autocorrelation function in the frequency domain. A single parameter of these functions, the multipath delay spread or its inverse, the coherence or correlation bandwidth are used to distinguish between selective and nonselective fading.

In digital transmissions time dispersion results in intersymbol interference and in irreducible error rate, not improved by increasing signal-to-noise ratio. This effect arises when the spread in propagation delays becomes comparable with the duration of the digital data frame period. Therefore, simple data transmission systems are limited to a relatively low bit rate. The error probability of binary signals can be calculated from the probability distribution of the amplitude fluctuations in a multipath channel [27].

The maximum delay time between pulses corresponding to the individual propagation paths depends on the propagation conditions. Fig.14 shows the delay as a function of radio circuit length for two sunspot activity levels and a parabolic layer model. The delay times range between 0.5 and 4 ms and almost double in the transition from minimum to maximum sunspot activity [28].

The multipath delay time also depends strongly on the operating frequency and tends to zero when the frequency approaches the MUF. This is shown by the 'multipath reduction factor' (Fig.15) which is defined as the lowest percent of the MUF for which the range of multipath propagation times is less than a specified value [29]. It thus defines a frequency above which a specified minimum protection against multipath is provided.

4.5.5.2 Frequency Dispersion -

Frequency dispersion nearly always accompagnies time dispersion, but the converse is not necessarily true. Frequency dispersion arises because the reflecting properties of the ionosphere are varying with time. This is due to changes in the electron density caused by the variations of the sun's zenith angle with the time of the day but also by irregularities. The ionosphere may be thought of as an irregular reflecting surface that is drifting across the sky. Because of this drift, waves that arrive at

the receiving site are being reflected from elemental surfaces that are in motion; consequently, each reflected wave will suffer a different Doppler frequency shift. The interference of these frequency-shifted waves give rise to rapid fading, or, in digital transmissions, may affect the performance of frequency-shift keying systems. The system will make an error if the transmitted frequency is changed far enough to cross over into the wrong side of the discriminator and remains there for a period comparable to half the element length.

A reasonable model describing the fading signal is to assume it has the character of very narrowband Gaussian noise. For this case the fading signal bandwidth or Doppler spread is proportional to the fading rate, i.e. the mean number of times per second the signal passes through the median signal level with positive slope. The proportionality factor depends on the type of filter assumed. A more comprehensive description of the fading signal is obtained by the probability distribution function of the instantaneous frequency deviation. Fig.16 shows this function normalised to the fading rate. For a fading rate of 1 Hz, the instantaneous frequency will be within about 12 Hz of the carrier frequency for 99.9% of the time. The mean duration of the instantaneous frequency deviation is shown in Fig.17 with the fading rate as a parameter. It is interesting to note that it is practically independent of the fading rate [30].

As is the case with multipath spread, the Doppler spread also depends on the frequency-to-MUF ratio. Correspondingly, a 'Doppler reduction factor' can be calculated, defined as the fractions of the MUF below which one must operate in order to remain below a certain prespecified Doppler spread [31].

4.6 Measures Against Multipath

From measurements it was deduced that any attempt to transmit serial digital streams over long distances via the ionosphere, at a rate exceeding 100 to 200 pulses/second without specific anti-multipath measures, will tend to result in a severe intersymbol interference problem. Unfortunately, the effects of multipath can not be overcome by increasing the signal-to-noise ratio, e.g by increasing the transmitter power or the antenna gain. Instead, special techniques of modulation and reception have to be used which are less vulnerable to fading. Most widely known are multiple-receiver combining techniques classified as diversity reception. The principle of diversity consists in the combination of different signals, the fluctuations of which are uncorrelated. We distinguish

- Space (spaced antenna) diversity: The signals are received by at least two antennas separated in space. The correlation distance for skywave propagation is in the order of 100 to 600 meters.

- Angle of arrival diversity: This method makes use of the angle of arrival of different modes being different. This can be done by highly directive antennas pointing into different directions or by automatically adjustable antenna arrays, which discriminate against certain mode paths in favour of others.

- Polarisation diversity: Signals are received by differently polarised antennas. This method makes use of the statistical fluctuations of polarisation of the skywaves.

- Frequency diversity: The information is transmitted simultaneously on different frequencies. The selective frequency correlation radius is generally inversely proportional to the maximum delay between modes. On medium latitude paths up to 3000 km, the frequency correlation radius may range from 250 to 7500 Hz as a function of the ratio of wave frequency to MUF and reach a maximum at a value of between 0.7 and 0.8.

  Intersymbol interference can be reduced by transmitting multiple tones, each carrying digital data at relatively low baud rate. For example, a paralell tone technique using 16 data tones, each carrying data at a 75 baud rate permits data rates up to 2400 bit/s.

- Time (signal repetition) diversity: The signal is transmitted several times

- Multipath diversity: The system makes positive use of multipath signals by separating the different contributory modes.

In modern digital communication applications, diversity schemes are invariably employed in conjunction with a suitable coding of the information to be transmitted, the basic form of coding being the forward error correction code. Studies of error statistics in digital transmissions over operating HF radio radio circuits agree at the point that the observed error patterns are in general very different from random patterns, which would be expected with independent occurrence of binary errors. Instead, errors tend to cluster, and when the modulation technique includes frequency division multiplexing, periodicity of error pattern may sometimes be observed.

The effectiveness of short block codes with interleaving, diffuse convolutional codes and error detection and repeat schemes appear to be well established. However, procedures to select the best technique for a particular channel and particular desired

error rate have not yet been developed. Comparison between various studies found in literature is difficult as the parameters of the tests reported vary widely.

5 REFERENCES

[1] A.Sommerfeld, "The propagation of waves in wireless telegraphy", Ann. Physik, 28, 1909, 665.

[2] B.Van der Pol, H.Bremmer, Phil. Mag. 24, 1937, 141-176 and 825-864.

[3] K.A.Norton, "The propagation of radio waves over the surface of the earth and in the upper atmosphere", Proc. IRE, 24, 1936, 1367-1387 and 25, 1937, 1203-1236.

[4] J.Grosskopf, "Wellenausbreitung", Hochschul- Taschenbuch 539/539a, Bibliograph. Institut, Mannheim, 1970.

[5] CCIR "Groundwave propagation curves for frequencies between 10 kHz and 30 MHz", Rec.368-4, Intern. Telecomm. Union, Geneva, 1982.

[6] R.H.Ott, L.E.Vogler, G.A.Hufford "Ground-wave propagation over irregular inhomogeneous terrain", NTIA Rep.79-20, 1979, NTIS Access No. PB 298 668/AS Nat. Techn. Inf. Service, Springfield, USA

[7] CCIR "Ground-wave propagation in an exponential atmosphere", Rep.714-1, Intern. Telecomm. Union, Geneva, 1982.

[8] CCIR "Electrical characteristics of the surface of the earth", Rec.527-1, Intern. Telecomm. Union, Geneva, 1982.

[9] CCIR "Electrical characteristics of the surface of the earth", Rep.229-4, Intern. Telecomm. Union, Geneva, 1982.

[10] CCIR "World atlas of ground conductivities", Rep.717-1, Intern. Telecomm. Union, Geneva, 1982.

[11] CCIR "Propagation by diffraction", Rep.715-1, Intern. Telecomm. Union, Geneva, 1982.

[12] CCIR "Influence of terrain irregularities and vegetation on tropospheric propagation", Rep.236-5, Intern. Telecomm. Union, Geneva, 1982.

[13] A.Malaga, "Delay dispersion of wideband ground wave signals due to propagation over a rough ocean surface", Radio Science, Vol.19, No.5, Sept-Oct.1984, 1389-1398.

[14] CCIR "Long-distance ionospheric propagation without intermediate ground reflection", Rep.250-5, Intern. Telecomm. Union, Geneva, 1982.

[15] K.Rawer, "Optique geometrique de l'ionosphere", Revue scientifique, 86, 1948, 585-600.

[16] K.J.Hortenbach, F.Rogler, "On the propagation of short waves over very long distances: predictions and observations", Telecomm.J., VOL.46, VI/1979.

[17] P.A.Bradley, "Wave polarisation and its influence on the power available from a radio signal propagated through the ionosphere", Proc. IEE, 115(6), 777-781.

[18] CCIR "CCIR Atlas of ionospheric characteristics", Rep.340, Intern. Telecomm. Union, Geneva,1983.

[19] CCIR "CCIR Interim method for estimating sky-wave field strength and transmission loss at frequencies between the approximate limits of 2 and 30 MHz", Rep. 252-2, Intern. Telecomm. Union, Geneva, 1970.

[20] CCIR "Second CCIR computer-based interim method for estimating sky-wave field strength and transmission loss at frequencies between 2 and 30 MHz", Supplement to Rep. 252-2, Intern. Telecomm. Union, Geneva, 1980.

[21] CCIR "Propagation prediction methods for high frequency broadcasting", Rep. 894, Intern. Telecomm. Union, Geneva, 1982.

[22] J.D.Whitehead, "Production and prediction of soradic E", Rev. Geophys. Space Phys., 8, 1970, 65-144.

[23] CCIR "Method for calculating sporadic-E field strength", Rec.534-1, Intern. Telecomm. Union, Geneva, 1982.

[24] CCIR "Ground and ionospheric side- and backscatter", Rep.726, Intern. Telecomm. Union, Geneva, 1982.

[25] CCIR "Ionospheric propagation characteristics pertinent to terrestrial radiocommunication systems design", Rep.266-5, Intern. Telecomm. Union, Geneva, 1982.

[26] CCIR "Factors affecting the quality of performance of complete systems in the fixed service", Rep.197-4, Intern. Telecomm. Union, Geneva, 1982.

[27] A.I.Fomin, P.N.Serdyukov, V.V.Mezhevich, "Evaluation of the error probability for optimal incoherent reception of binary signals in a multipath channel", Radiotekhnika, No.8, 1984, 48-50.

[28] CCIR "Path time- delays and shifts caused by multipath propagation on radio circuits", Rep.203-1, Intern. Telecomm. Union, Geneva, 1982.

[29] D.K.Bailey, "The effect of multipath distortion on the choice of operating frequencies for high frequency communication circuits", IRE Trans. Ant. Prop. AP-7, 1959, 398.

[30] CCIR "Influence on long-distance HF communication using frequency-shift keying of frequency deviations associated with passage through the ionosphere", Rep.111, Intern. Telecomm. Union, Geneva, 1982.

[31] L.W.Pickering, "The calculation of ionospheric Doppler spread on HF communication channels", IEEE Trans. Comm., COM-23, No.5, May 1975, 526-537.

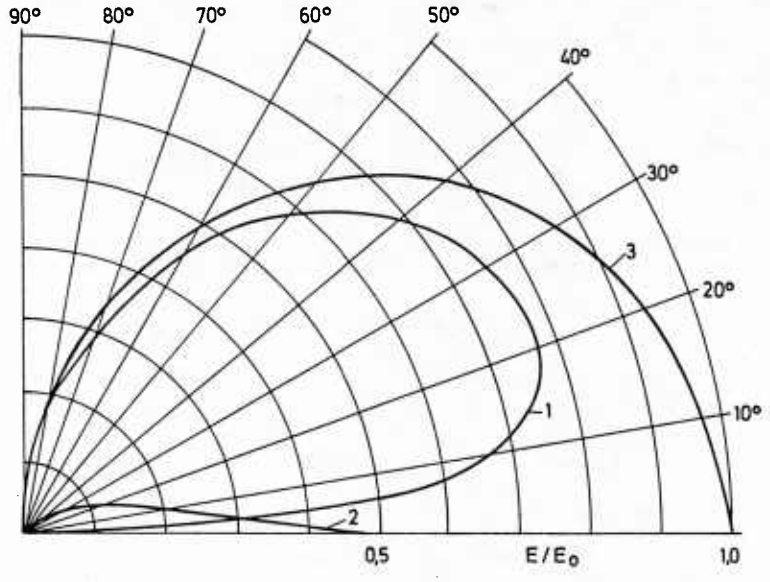


Fig.1 vertical radiation pattern of a short vertical antenna on ground:
1: space wave 2: ground wave 3: infinite ground conductivity

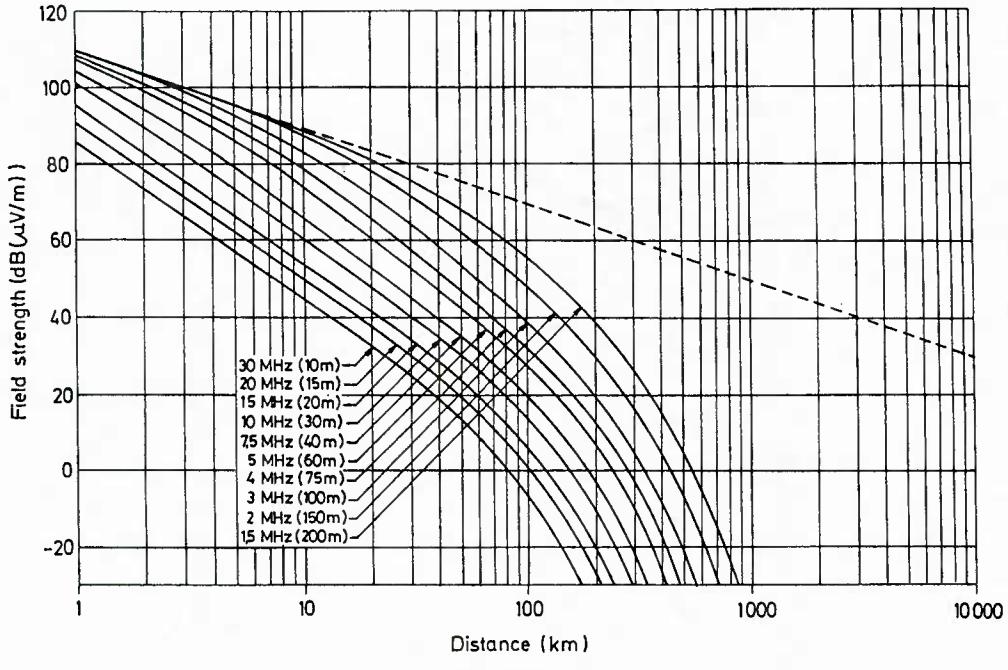


Fig.2 Ground wave propagation curves [5]
Land: σ = 0.03 S/m, ε = 30
---- inverse distance curve

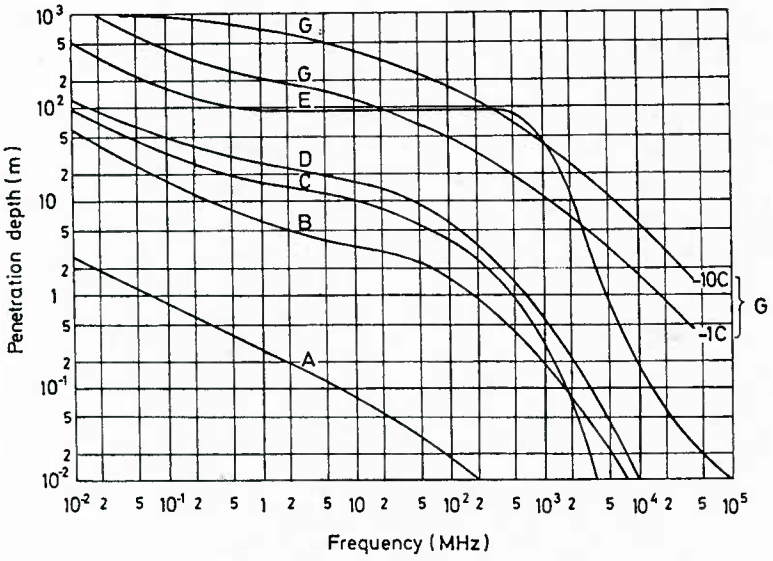


Fig.3 Penetration depth as a function of frequency [8]

A: sea water
B: wet ground
C: fresh water
D: medium dry ground
E: very dry ground
G: ice (fresh water)

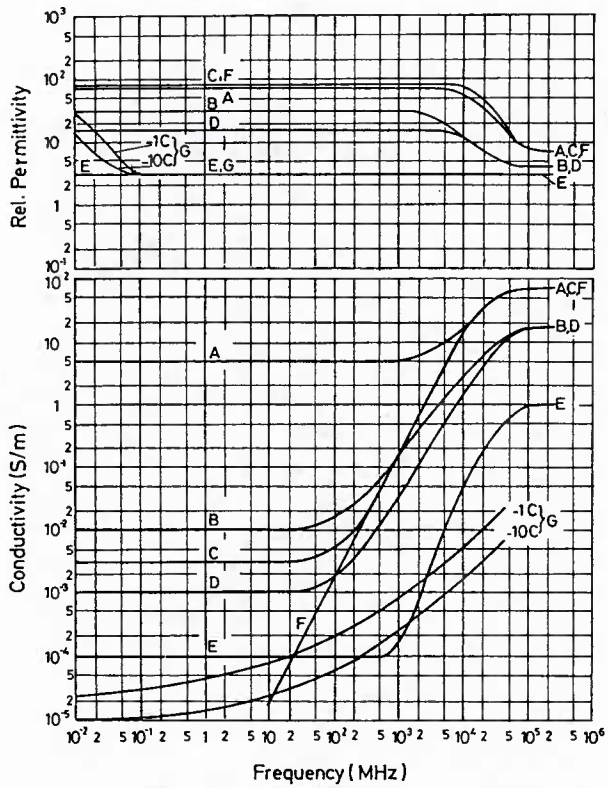


Fig.4 Relative permittivity and conductivity as a function of frequency [8]

A: sea water, 20°C
B: wet ground
C: fresh water, 20°C
D: medium dry ground
E: very dry ground
F: pure water, 20°C
G: ice (fresh water)

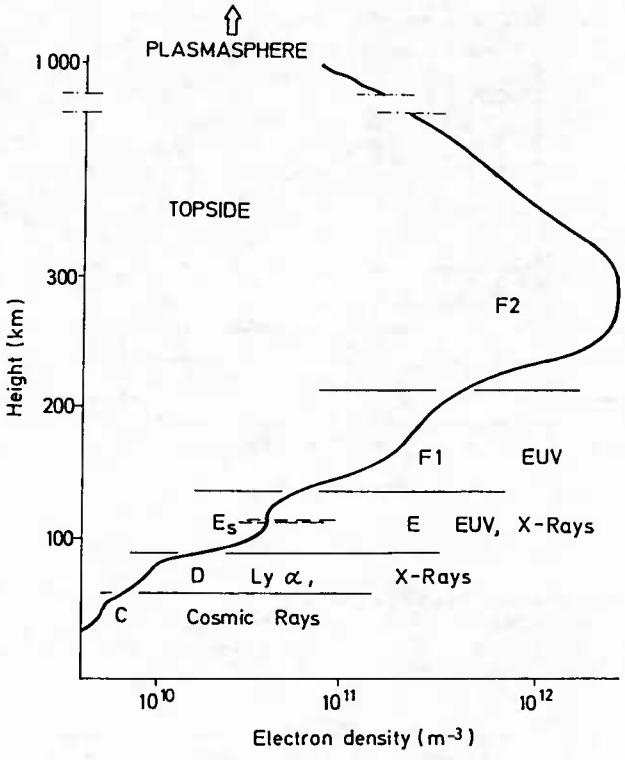


Fig.5 Vertical electron density profile of the ionosphere [after DAVIES]

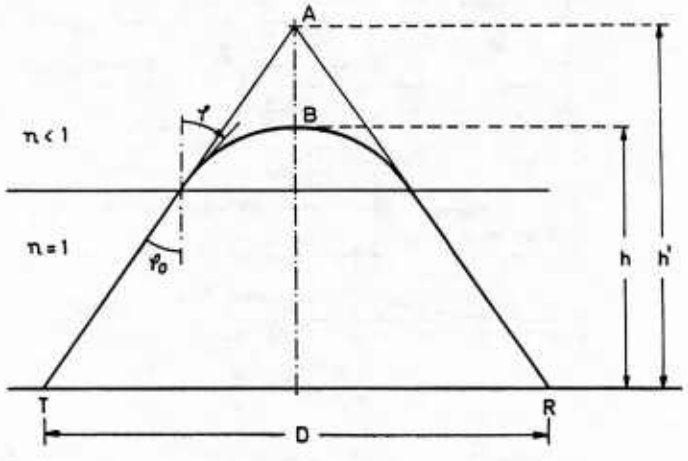


Fig.6 True and virtual propagation paths in the ionosphere

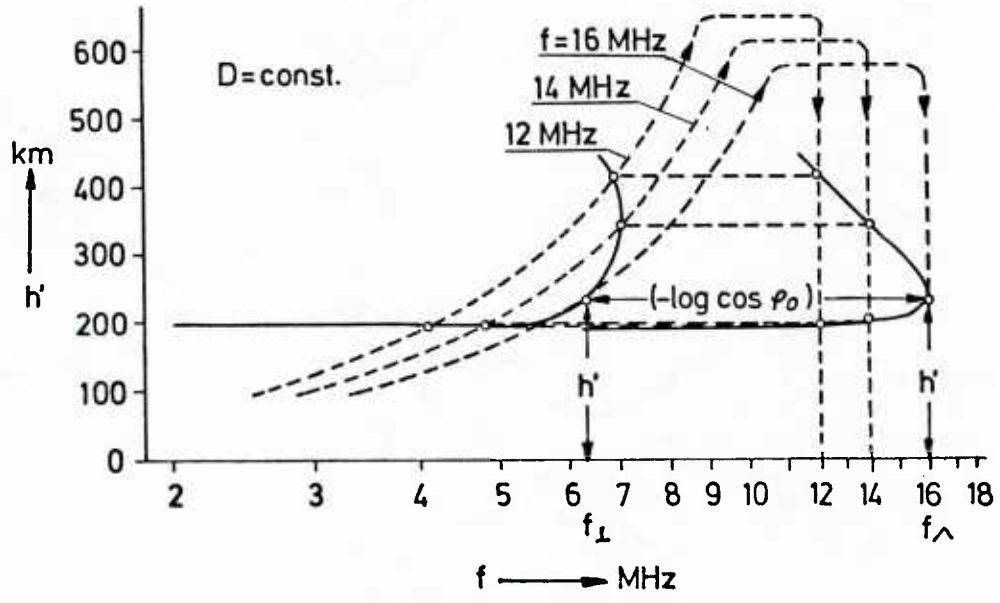


Fig.7 Construction of propagation curves h'(f) for oblique incidence from vertical incidence ionograms by use of the 'transmission curves' [after DIEMINGER]

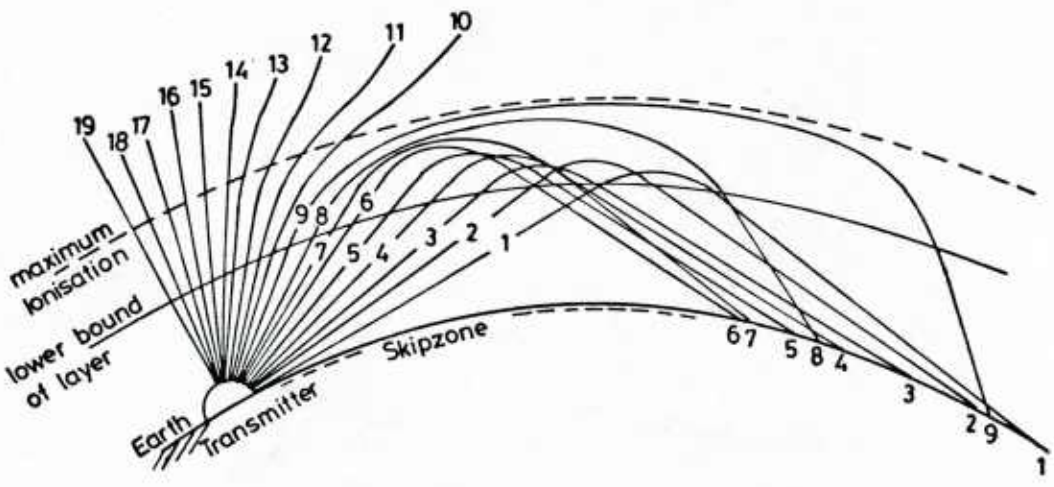


Fig.8 Raypaths in the ionosphere for different elevation angles and constant frequency

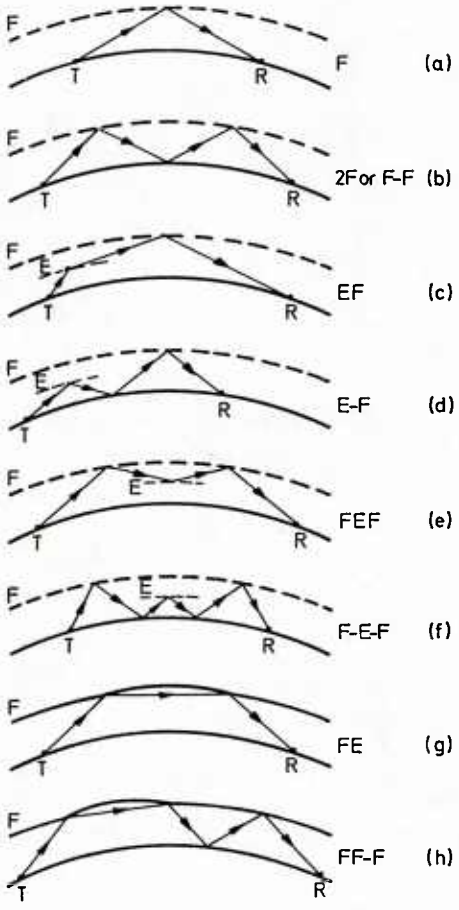


Fig.9 Examples of possible multihop-/multimode paths

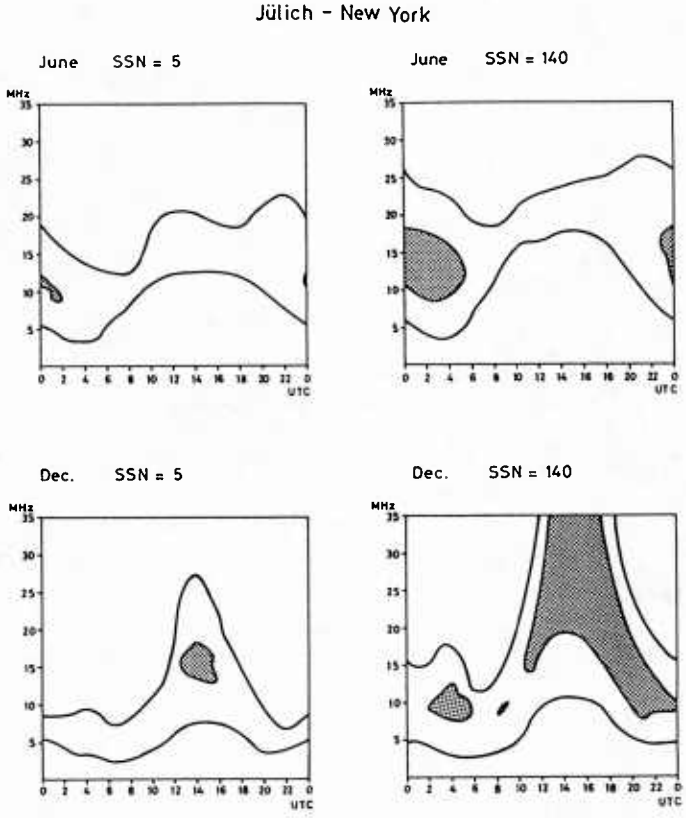


Fig.10 Variation of usable frequency range with time of day, season and sunspot number

(Within the areas limited by the upper and lower curves the circuit reliability exceeds 50%, within the shaded area 90%)

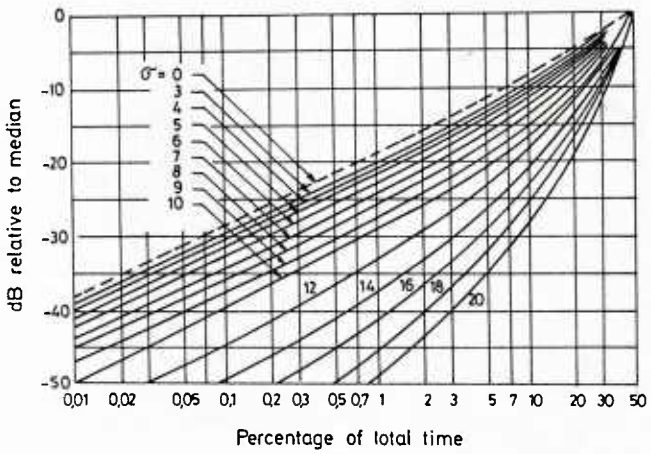


Fig.11 Level below which signal strength falls when instantaneous signal strength has a Rayleigh distribution but daily median values have a log-normal distribution with standard deviation σ dB [25]

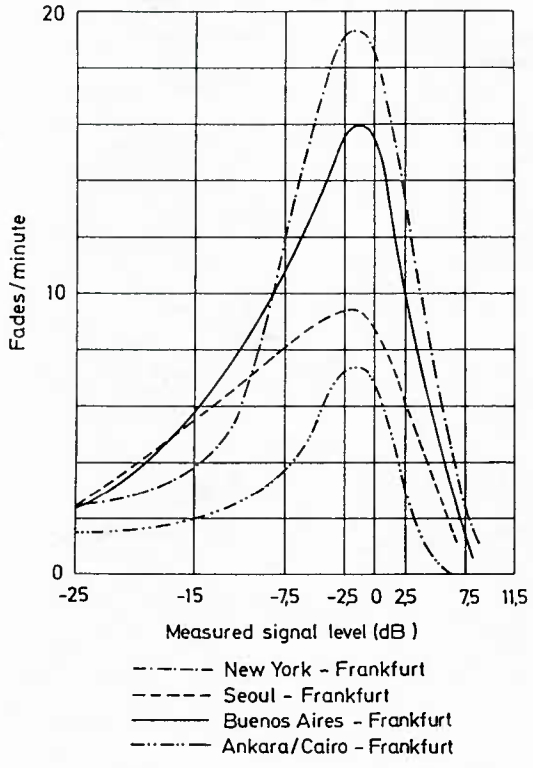


Fig.12 Number of fades per minute as a function of the signal level for various circuits [26]

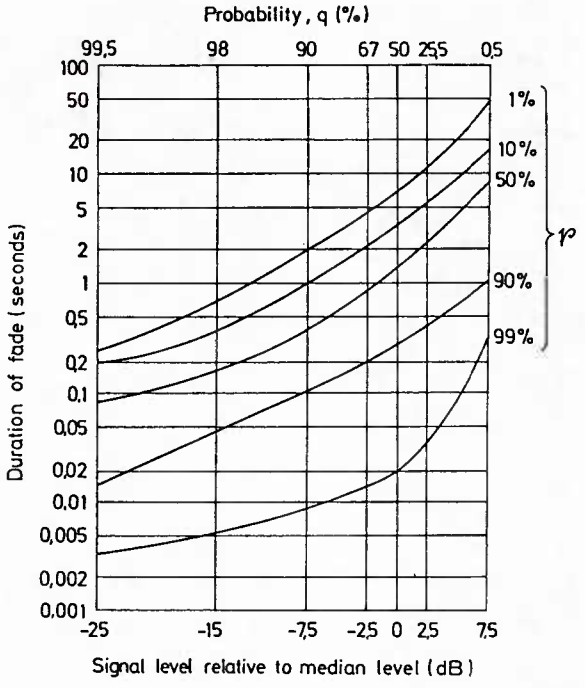


Fig.13 Duration of fades as a function of the signal level [26]
Circuit New-York - Frankfurt, 14-Sep-1961, 1000 UTC, 13.79 MHz

p: percentage of the number of fades for which a given duration of fade is exceeded
q: probability that signal level is exceeded

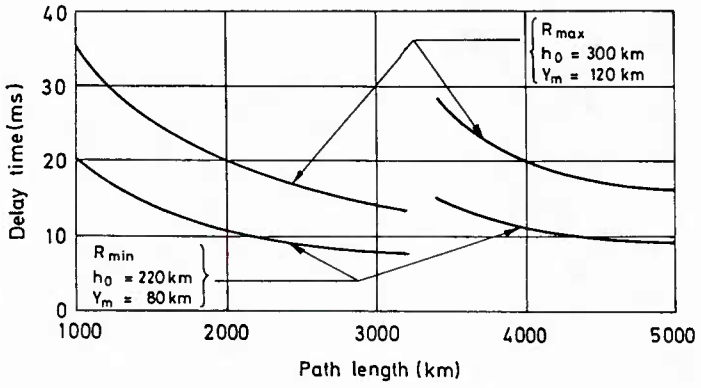


Fig.14 Multipath delay time for a parabolic layer model [28]

R_{min} : minimum sunspot activity
R_{max} : maximum sunspot activity
h_o : height of lower limit of ionospheric layer
Y_m : half-thickness of ionospheric layer

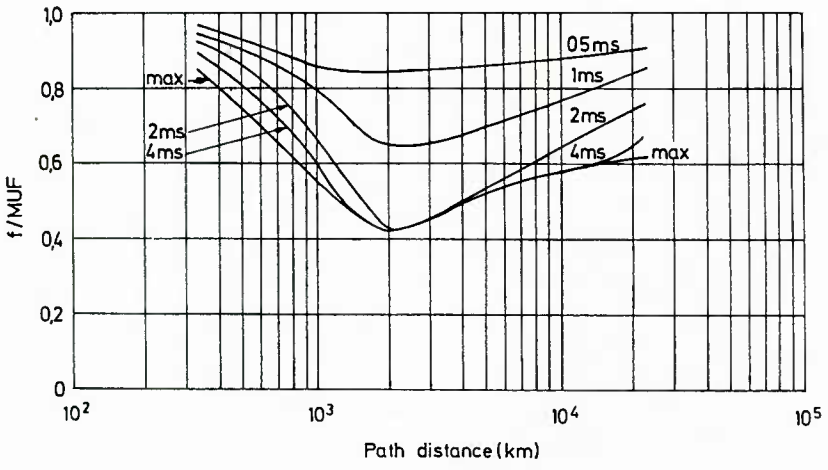


Fig.15 Multipath reduction factor [29]

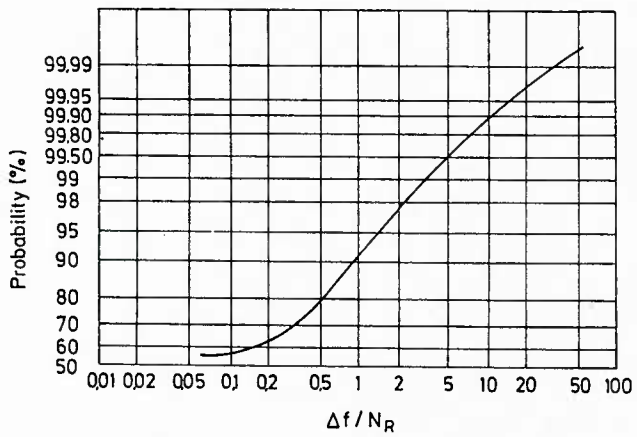


Fig.16 Probability that instantaneous frequency change is less than the abscissa [30]

Δf: Instantaneous departure from centre frequency (Hz)
N_R: Fading rate (fades per second)

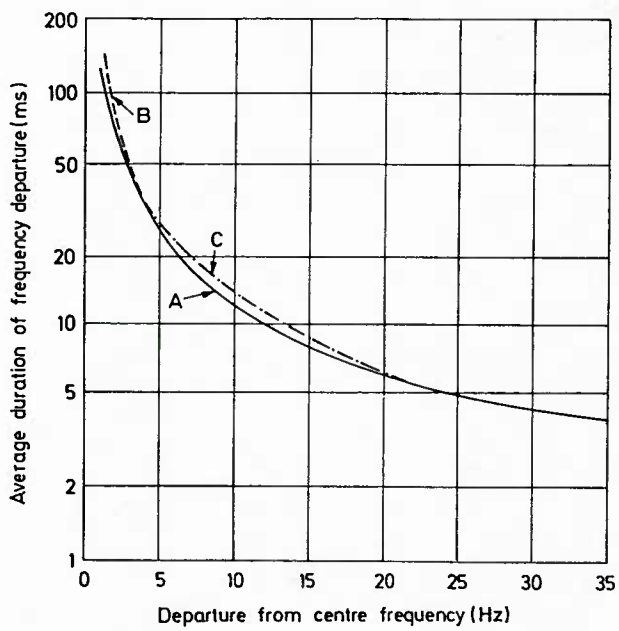


Fig.17 Average duration that instantaneous frequency change exceeds abscissa for fading rates 0.2 (A), 1.0 (B) and 5.0 (C) fades per second (Narrowband Gaussian noise model) [30]

HF SYSTEM DESIGN

M Darnell
Department of Electronic Engineering
University of Hull
Hull HU6 7RX
UK

SUMMARY

This lecture is intended to complement the companion lecture dealing with the propagation aspects of HF communications.

The major topics covered are

(i) the design of point-to-point systems;
(ii) the design of systems involving mobile terminals;
(iii) system control and frequency management;
(iv) resistance to interception and disruption.

The lecture attempts to show how the propagation effects discussed in the previous lecture on "HF groundwave and skywave propagation" are taken into account in the design of HF communications systems for various purposes.

1. INTRODUCTION

In this lecture, the design of static and mobile HF communication systems involving static and mobile terminals will be considered from the viewpoints of present practice and future trends (Darnell, 1986a). The purposes of the lecture are

(i) to introduce the problems and techniques associated with the design of both static and mobile HF communication systems, and with their control and operation;

(ii) to identify the techniques which appear to offer greatest promise of improved performance in future designs;

(iii) to identify the deficiencies in propagation and noise/interference databases requiring more basic measurements to be made.

Before the HF propagation medium can be used efficiently for communications purposes, its strengths and weaknesses must first be appreciated; the more important of these are listed below.

1.1 Strengths of the HF Medium

(a) The ionosphere is a robust propagation medium which recovers rapidly after major disruptions, eg polar cap events (PCE's), sudden ionospheric disturbances (SID's) and high altitude nuclear bursts.

(b) Long-term (monthly median) skywave propagation parameters are normally predictable with reasonable accuracy.

(c) Groundwave propagation parameters can be predicted with high precision;

(d) Each communication link exhibits unique characteristics, eg fade rates and depths, multipath structure, noise/interference levels, etc, which potentially can be used to isolate a given channel from the effects of other transmissions in the HF band.

(e) Only relatively simple equipment and operating procedures are required to achieve access to the medium, thus enabling it to be exploited by simple mobile terminals.

(f) Equipment costs are low in comparison with those of other types of long-range, beyond line-of-sight (LOS) communications systems.

1.2 Weaknesses of the HF Medium

(a) The ionosphere is subject to sudden and unpredictable disturbances such as the PCE's and SID's mentioned in the previous section.

(b) Although the long-term parameters of a skywave propagation path are reasonably predictable, significant departures from such predictions can be expected in the short-term from day-to-day.

(c) Levels of man-made co-channelinterference are high, particularly at night, and the nature of such interference is inadequately characterised.

(d) A high level of system availability and reliability requires considerable operator expertise in manually-controlled systems.

(e) The available capacity of a nominal 3 kHz bandwidth HF channel is limited to a maximum of a few kbits/s; data rates of, at most, a few hundreds of bits/s are more realistic if high levels of availability and reliability are required.

As will be shown in Sections 4 and 5 of this lecture, care must be taken to employ HF systems operationally for the types of traffic and services for which they are well matched, and not to attempt to impose "unnatural" requirements for which the medium is fundamentally unsuitable.

1.3 Mobile Problems

In addition to the inherent natural weaknesses of the HF medium noted above, systems incorporating mobile terminals suffer from other fundamental limitations. When contrasted with static (normally land-based) terminals, mobile installations have the following disadvantages (Darnell, 1985):

(a) Relatively low transmitter powers, limited by the electrical generation capacity of the mobile platform.

(b) Small, narrowband, and relatively inefficient transmitting antennas, possibly with minimal directivity.

(c) Electromagnetic compatibility (EMC) problems arising from the HF communications equipment being co-sited with other EM systems such as radars and communications transmitters using different frequency bands.

(d) Reliance upon a propagation path which, in addition to natural time variability, is also affected by significant variations in mobile position.

Land based stations, on the other hand, can use high powers and efficient and directive antennas, coupled with electrically quiet receiving locations. It is also relatively straightforward to provide interconnections between sites to give a capability for networking.

In military applications, mobile terminals may represent a particularly vital element in the operational strategy: as a consequence there may be additional requirements placed on the mobile to function in such a way that the probability of its transmissions being decoded, intercepted, located or jammed is minimised, ie it has low probability of intercept/anti-jamming (LPI/AJ) characteristics.

Mobile terminals can thus be considered as a class of "disadvantaged" users of the HF spectrum, who may nevertheless be required to communicate reliability and securely over a wide geographical region, possibly via an ionospheric skywave path exhibiting severe irregularities.

1.4 Design Philosophy

One of the main problems associated with the operation of current HF systems is that they are employed primarily to pass traffic at a constant rate - possibly up to 2.4 kbits/s and above - in the same way as with other less dispersive media, ie they are in a sense considered to be interchangeable with line, satcom and other LOS propagation media. However, as will be shown in Section 2, the capacity of an HF skywave link is constantly varying over a wide range which ideally requires adaptation of the signal parameters and/or signal processing procedures in accordance with the available capacity at any time. The design concepts discussed in this lecture attempt in some circumstances to mitigate, and in others to exploit, the natural variability of HF paths in various ways.

For both static and mobile systems, the basic requirement is for more effective and responsive control and adaptation procedures to be incorporated into the system design, rather than continuing to use the relatively inflexible frequency utilisation procedures and transmission formats employed by many HF systems at present. An essential pre-requisite for this improved control and adaptation is the availability of an appropriate channel model obtained by real-time channel evaluation (RTCE): see (Darnell, 1986b) and (Darnell, 1983).

Two basic communications scenarios which are encountered in practice are illustrated in Figs. 1(a) and 1(b). Fig. 1(a) shows an "open-loop" situation in which there is no feedback path between receiver and transmitter, whilst Fig. 1(b) shows a "closed-loop" situation in which such a path does exist. Ideally, the aim of the control procedure should be to made the transmitted signal, x(t), and the received signal, y(t), identical. In both cases, before any form of optimal or sub-optimal control can be applied to achieve the desired purpose of the system, it is necessary to characterise the propagation path in order to produce an appropriate model for use in the control algorithm, ie the path must be "identified".

For the open-loop system this can be achieved in three ways:

(a) By using a priori knowledge of the nature of the path, eg from previous measured data and via off-line propagation analysis programs.

(b) By multiplexing RTCE signals with the normal communications traffic signals so that the receiver can extract information to model the path and hence can adjust its parameters accordingly.

(c) By deriving RTCE information at the receiver from the "natural" operating and control signals of the system.

In the closed-loop system, the availability of a feedback link enables information extracted from the received signal(s) to be used to characterise the path and the data derived from this characterisation then to be passed back to the transmitter to allow its parameters to be varied adaptively.

To date, the design of many HF systems incorporating mobile terminals takes only limited account of the specific problems mentioned in Section 1.3. In Section 5 of this paper, the manner in which these problems can be minimised in systems incorporating mobile terminals is discussed. The design philosophy adopted exploits the strengths of fixed terminals and minimises the effects of any problems associated with mobile terminals, at the same time placing as few constraints as possible upon mobile operational flexibility. In essence, the complexity of the overall design resides in the fixed terminals of the system, with the mobile equipment being kept as simple as possible.

2. OPERATIONAL ENVIRONMENT & REQUIREMENTS

Currently, there seem to be a number of misconceptions about the nature and potential of the HF band as a vehicle for communication. These misconceptions arise from the following factors:

(a) A tendency to consider the problem of reliable HF transmission as one of exploiting Rayleigh-fading ionospheric skywave and/or steady-signal groundwave paths, with a background of Gaussian white noise (GWN). In practice, the propagation mechanisms may be more complex and the noise background is non-Gaussian in most instances.

(b) An assumption that constant transmission rates are appropriate for HF systems, whereas the available information transmission capacity of any given channel may vary widely over a short time interval.

(c) An assumption that HF communication can replace, or back-up, other types of less dispersive communication media and still maintain a similar level of reliability. In fact, there are classes of traffic, which can be carried by say a 3 kHz bandwidth telephone channel, for which a similar bandwidth HF channel is fundamentally unsuitable.

(d) An incomplete understanding of the medium in terms of the strengths and weaknesses listed in Sections 1.1 and 1.2.

In the following sections, it will be assumed that conventional frequency assignment and bandwidth limitations will apply in any HF system design, ie a typical communication system will have a limited set of assigned channels within which to operate, with each channel having a nominal maximum bandwidth of 3 kHz. Additionally, it will be assumed that there are no geographical restrictions upon the operational deployment of the systems.

2.1 Propagation, Noise & Interference Models

In the design phase of HF communication systems, it is convenient to use off-line propagation analysis programs, eg (Barghausen et al, 1969) and (Lloyd et al, 1981), to dimension the system in terms of required transmitter powers, antenna gains, optimum location of terminals, etc. There are several basic problems associated with the use of these programs:

(i) They make use of limited data bases.

(ii) They can only treat effects such as sporadic E-layer propagation in a rather imprecise manner.

(iii) The modelling of noise and co-channel interference effects tends to be rudimentary, and can lead to inappropriate system design criteria (Darnell, 1984).

(iv) Their outputs are typically in the form of relatively long-term, eg monthly median, predictions of path parameters at hourly intervals.

Even after allowing for the shortcomings listed above, off-line analysis can yield useful estimates of required system design parameters. Such analysis, however, it does not provide information on the short-term, ie minute-to-minute or second-to-second, variability of the propagation medium which would enable more detailed consideration to be given to the form of signal generation and processing procedures appropriate for a particular system.

To some extent, channel simulators overcome this deficiency by making use of a deterministic model of HF propagation [see for example (Watterson, 1975) which may, or may not, be representative of the conditions on a given path at a given time]. Howeve, and more importantly, simulators still suffer from the lack of an adequate model of co-channel interfering signals - a vital factor in determining the reliability of many communication systems. In most cases, channel noise and interference is modelled simplistically as a GWN process.

In order to assess the types of transmission formats and services for which the ionospheric medium is appropriate, it is first helpful to describe the propagation and interference environment in a way which is meaningful to a communication system designer. An information-theoretic model of propagation is postulated in (Darnell, 1982a); this describes ionospheric skywave propagation in terms of 3-dimensional "windows" with dimensions of frequency, time and distance, as shown in Fig. 2. The sides of the window cover the ranges

$$\left.\begin{array}{rl} & f_1 + \Delta f \\ & t_1 + \Delta t \\ \text{and} & d_1 + \Delta d \end{array}\right\} \qquad [1]$$

Within a given window, a specified signal-to-noise power ratio (SNR) will be maintained for the ranges of f, t and d shown. In general, the lower the required SNR, the greater will be the dimensions of the window; thus in Fig. 2, window 2 corresponds to a higher SNR than does window 1. For a given communication system, the "volume" of an individual window will depend upon the nature of the signal processing procedures used at the receiver, eg modulation and coding, since these affect the required SNR at the receiver; transmitting and receiving antenna characteristics, diversity processing, diurnal and seasonal variability, etc, may also cause the window dimensions to change.

One systematic study of the nature of co-channel interference in the HF band is being carried out by (Gott et al, 1983). Fig.3 is an example of the occupancy of the HF spectrum obtained from this programme, showing a high level of spectral congestion. Fig.4 illustrates an error rate profile for a typical HF channel, clearly demonstrating large variations over relatively short periods of time. In the same way as propagation windows can exist between communications transmitter and receiver, similar windows may also occur between interfering transmitters and the communications receiver; any overlap between communications and interference windows will tend to reduce the effective volume of the former, and hence the capability of the communication system. Fig. 5 illustrates this situation.

It is instructive to consider what a window model of ionospheric propagation might look like for sporadic E-layer and meteor-burst modes. When sporadic E is present, it can give rise to stable, single-mode propagation: however, a given region of ionisation tends to be usable only over a somewhat restricted geographical area. Thus, it might be expected that windows due to this mechanism would be relatively large in the f and t dimensions, but small in the d dimension. Because of the physical nature of meteor-burst communication paths, ie they are of short duration and wide bandwidth, and are virtually unique for given transmitter and receiver locations; thus, they have small t and d dimensions but are relatively large along the f dimension. Fig.6 illustrates typical windows for various ionospheric modes: although, for simplicity, regularly-shaped windows are shown, in practice they would tend to be irregular volumes.

2.2 Transmission Characteristics arising from the Window Model

From the discussion in the previous section, it can be seen that the lower the SNR required at the receiver, the greater will be the volume of any transmission windows available. Therefore, if a transmission wndow, associated with a signal-to-noise ratio SNR_1, has maximum dimensions Δf_1, Δt_1 and Δd_1, any window capable of supporting a ratio SNR_2 where

$$SNR_2 > SNR_1 \qquad [2]$$

and lying in the same ranges of f, t and d will normally have maximum dimensions Δf_2, Δt_2 and Δd_2 given by:

$$\left.\begin{array}{l} \Delta f_2 < \Delta f_1 \\ \Delta t_2 < \Delta t_1 \\ \Delta d_2 < \Delta d_1 \end{array}\right\} \qquad [3]$$

This implies that the SNR_2 volume is less than the SNR_1 volume, as indicated in Fig. 2.

Using the Shannon upper bound for error-free channel capacity C (bits/s), ie

$$C = \Delta f \log_2 [1 + SNR] \text{ bits/s} \qquad [4]$$

where Δf, the transmission bandwidth, and distance are assumed fixed, it is possible to view the variation of SNR with time as a variation of channel capacity, C, as shown in Fig. 7.

This variation in channel capacity is due to changes in the window structure with time and might stem from a number of physical factors, eg one mode becoming dominant in a multipath situation or a fading co-channel interfering signal. These effects are illustrated in Fig. 8. In the region Δt, one of the wanted propagation modes is dominant with respect to both the other mode and the interfering signal, thus giving rise to a high effective SNR over that interval with a correspondingly large capacity. On a given ionospheric link, Δt values may range from milliseconds to hours, dependent upon the required value of SNR at the receiver; similarly, capacity may vary rapidly between a few kbits/s and a few tens of bits/s within seconds.

Many existing constant-rate transmission schemes employ techniques which, to some extent, are capable of smoothing out SNR variations, eg diversity processing and adaptive equalisation. However, as is explained in (Darnell 1984), these techniques cannot overcome capacity changes due to the presence of co-channel interfering signals. Therefore, the basic problem that the communication system designer has to solve is one of making effective use of a propagation medium with a time-varying information transmission capacity.

From the preceding discussion, it is evident that systems operating at constant transmission rates will, in many cases, be fundamentally mismatched to the variable capacity medium. In Fig. 7, a system operating at a constant rate R would normally have

$$R < C_{min} \qquad [5]$$

for error-free reception over a given time interval - which would represent an inefficient utilisation of available capacity for most of the transmission interval. From this simplified consideration of ionospheric path characteristics, it would seem that a more natural mode of transmission would involve a variable rate, adapted in respose to the prevailing path conditions. Using the model of propagation and interference shown in Fig. 8, even in the presence of extremely high levels of interference there will be occasional windows when the unwanted signal is in a deep fade and one of the wanted modes is dominant; it should be noted that this may be the case even if the average SNR is very low. If these windows can be identified, it will be possible to maintain some residual channel capacity, even in the face of extremely high-level co-channel interfering or possibly jamming signals. It should be stressed, however, that in severe interference these windows may only persist for seconds or fractions of a second at a time.

The similarity between the above model of normal ionospheric skywave propagation with high levels of interference and the characteristics of meteor-burst modes (Bartholomé & Vogt, 1968) is obvious; both can be considered as intermittent paths. The situation with normal skywave propagation and relatively low levels of co-channel interference is rather more complex; instead of the path being completely intermittent, it gives rise to transmission windows of variable capacity, possibly with some residual low-level of capacity being available, even under worst-case conditions.

In addition to the normal propagation and interference environment discussed in the previous two sections, HF systems may be required to operate under extreme conditions of poor signal propagation or high manmade interference levels. These situations are examined in the following two sections.

2.3 High Latitude Effects

Propagation mechanisms can be very irregular in polar regions, with the major effects being (Thrane, 1983):

(a) polar cap absorption (PCA) events due to the influx of high energy protons and alpha particles penetrating to the lower ionosphere: such events may persist for several days, causing severe attenuation of skywave signals and an effective communication "blackout" over a wide geographical area;

(b) auroral absorption, caused by high energy electrons penetrating the ionospheric D-region: this source of signal attenuation typically persists for a few hours only and affects a more restricted geographical area than does a PCA;

(c) irregular behaviour of sporadic E-layer, spread F-layer and auroral oval propagation modes, giving rise to abnormal and variable multipath time spreads;

(d) abnormal doppler shifts and phase instability of received signals due to the variable nature of the high-latitude ionosphere, manifesting itself in the form of very rapid signal fading;

(e) large variability of maximum usable frequencies (MUF's) from very low to abnormally high values: the former may result in spectral congestion and hence high levels of co-channel interference, whilst the latter may allow HF type propagation to take place in an extended frequency range above 30 MHz.

Obviously, the propagation environment in polar regions may be relatively unstable and unpredictable, thus limiting the effectiveness of off-line propagation analysis programs. Also, the degree of adaptation required of systems operating in these

regions is likely to be greater than for those deployed in temperate latitudes where ionospheric predictability is greater.

2.4 Interception and Disruption

In Fig. 5, the interference window could arise from the activities of a jammer attempting to disrupt the operation of the communication system. Similarly, the paths from communications transmitter to interception receivers will also have a characteristic window structure.

The problem for the designer and operator of the cummunication system is thus to minimise the probability of coincidence between interfering/jamming/interception windows and those for the communication system. If partial or complete coincidence does occur, there is an opportunity for disruption or interception of the communications traffic. The design strategy in this environment should therefore be to

(i) Maximise the size of the communications windows, and

(ii) Minimise the coincidence probability between communications and interference/jamming/interception windows by improved system control.

In general, it can be assumed that a potential jammer will have a significant radiated power advantage relative to the communicator; also, it is probable that such a jammer will have available directive and steerable antennas, and be capable of operating in both narrowband and wideband modes. Thus, in Fig. 5, the window dimensions for the jammer will on average be rather larger than those for the communications system. It is instructive to consider the fundamental limitations on ECM effectiveness and then to identify those techniques which can exploit those limitations, together with the strengths of the medium listed in Section 1.1, to the benefit of the communicator. Attention should not be confined solely to the conventional HF band; the possibilities offered by the use of higher frequencies should also be considered.

2.5 Fundamental Limitations on ECM Effectiveness (Darnell, 1982b)

As illustrated in the simple models discussed in Section 2.1, each HF path is, to some extent, unique in its characteristics. The probability of complete overlap of communications and ECM windows can therefore, in principle, be made low.

2.5.1 Propagation of Analysis Time Limitations

Consider the arrangement illustrated in Fig. 9: the network shown comprises a communications transmitter and receiver, together with a jamming and interception site. The propagation times in the network are:

$$\left.\begin{array}{lll} \text{(i)} & \text{Communications Transmitter} \longrightarrow \text{Communications Receiver} & = \tau_{TR} \\ \text{(ii)} & \text{Communications Transmitter} \longrightarrow \text{Jamming and Interception Site} & = \tau_{TJ} \\ \text{(iii)} & \text{Jamming and Interception Site} \longrightarrow \text{Communications Receiver} & = \tau_{JR} \end{array}\right\} \quad [6]$$

Assuming that the jammer cannot react to changes in communications system parameters before its associated interception site has monitored these changes, the fundamental propagation time advantage τ_A for the communicator over the jammer is:

$$\tau_A = [\tau_{TJ} + \tau_{JR}] - \tau_{TR} \qquad [7]$$

In addition, however, the interceptor will require a further time, τ_{AN}, in order to analyse changes in the nature of the communications traffic before he can command the jammer appropriately. Thus, the overall time advantage now becomes:

$$\tau_A = [\tau_{TJ} + \tau_{JR} + \tau_{AN}] - \tau_{TR} \qquad [8]$$

This represents a fundamental limitation on jammer effectiveness which could only be overcome to some extent, and possibly at inadmissible cost, by continuous, wideband, "blanket" jamming.

Practically, it is difficult for co-located interception facilities to look through a high power jammer, and therefore the intercept location may well be separated physically from the jammer. In this case, an additional propagation time between interceptor and jammer must be included in expression [8].

Typically, with modern intercept and jamming systems τ_A could be expected to have a minimum value of a few tens of milliseconds - the majority of that time being taken up by signal analysis. Two approaches suggest themselves as being able to exploit this time limitation on ECM effectiveness:

(i) The use of a communications system employing rapid frequency agility under cryptographic control such that the time spent on each frequency is τ_A. This would inevitably require high signal levels which would be relatively simple to monitor and track.

(ii) The use of a transmission format which has no unique characteristics and is thus difficult to distinguish from other HF traffic in which the jammer has no interest. In this case, the "sorting" problem faced by the interceptor is increased with a consequent lengthening of analysis time, τ_{AN}.

2.5.2 Corrupted Reference Limitations

A further advantage held by the communicator over the jammer/interceptor is the fact that the communications receiver will always be in the position of knowing the exact range of possibilities for the transmitted signal prior to reception. Therefore, the actual received signal can be compared with a set of perfect reference signals in a maximum likelihood detection scheme. The interception site, on the other hand, should at best only have available a corrupted estimate of the set of possible transmitted signals. Potentially, this means that the communications system has at least a 3dB advantage over the interceptor (eg the difference between coherent and differentially-coherent phase-shift keying). This advantage can be further enhanced by means of techniques such as correlation reception, quenched filtering, accurately-controlled tone filtering, time adaptation, frequency agility, encryption, signal format variation, etc, which will be discussed further in the later sections.

2.5.3 Data Base Limitations

Related to the analysis time and noisy reference limitations identified above is a more general limitation of the data base available to the interceptor and jammer. An effective jamming control strategy requires detailed knowledge of communication network configuration and of which are the key links. Propagation and analysis time limitations may well mean that it will take an ECM controller a considerable period to build-up such a data base. A communicator can accentuate this problem by making his communication system time-variable to the maximum extent, eg by altering frequency plans, traffic flows, network topology, cryptographic procedures and signal formats.

Direction finding (DF) can, and will, be employed by the interceptor as a vital sorting parameter: in many cases, it may well be the only available means of signal propagating to the interception site and short transmission durations can be used to increase the errors associated with DF.

As a general principle, the amount of "real" traffic passed by a communication system should be minimised. This then affords the option for the system transmission rates to be contracted by the use of more powerful source and channel coding techniques in the face of an ECM threat, whilst still retaining sufficient capacity to pass essential information. Hence, the overall system should not be organised in such a way that the peacetime traffic levels must be maintained in order to ensure its effective operation in time of stress.

In the following three sections, the effect of the factors discussed in Sections 1 and 2 will be considered in the context of a generalised HF system design methodology.

3. SYSTEM DESIGN: CURRENT STATUS

In a generalised HF communication system incorporating both static and/or mobile terminals, there may be a requirement for any of the following:

(a) static terminal-to-static terminal communication;

(b) static terminal-to-mobile terminal communication;

(c) mobile terminal-to-static terminal communication;

(d) mobile terminal-to-mobile terminal communication.

The manner in which each of the links above is currently implemented will now be outlined briefly.

3.1 Static Terminal-to-Static Terminal

Because relatively high transmitter powers and directive antennas can be installed at fixed sites, the engineering of a static terminal-to-static terminal link can normally be accomplished to give a reasonable communications reliability - at least for simple forms of traffic, eg low-speed telegraphy data and analogue speech. However, digital data, or digital speech, at say 2.4 kbits/s will present a greater problem, and circuit reliabilities for this form of traffic will be relatively low.

If a static link operates with a defined frequency complement on a regular schedule, there will be a significant frequency "airing" effect which will tend to deter other

users of the spectrum from encroaching upon those frequencies, thus minimising the levels of co-channel interference.

3.2 Static Terminal-to-Mobile Terminal

Static terminal-to-mobile terminal communication is frequently implemented via a broadcast-type system; this again allows high transmitter powers and directive antennas to be used at a fixed site in order to radiate signals simultaneously to multiple mobile terminals, possibly situated anywhere within a wide geographical region.

A basic broadcast system operates on the principle of long-term frequency selection diversity; Fig.10 illustrates that principle. It is seen that at any time of day, the same traffic signal will be radiated on a number of distinct frequencies or components, with the assumption being that at least one of the components will propagate effectively to any given mobile. The frequencies must therefore be chosen to provide a compromise coverage over the complete mobile operational area for both day and night conditions.

Typically, each component will comprise multiple sub-channels arranged within a nominal channel bandwidth, as shown in Fig. 11. The sub-channels correspond to traffic signals originating from various message sources; a given mobile will only extract those sub-channels of interest to it from the multiple sub-channel raster.

If all sub-channels are not assigned, it is possible to transmit the same information on more than one sub-channel to provide in-band frequency diversity. In Fig.11, the i^{th} sub-channel is employed for engineering order wire (EOW), or system control, purposes.

The system thus has the potential for frequency diversity at two levels: a given mobile selects the best component in terms of received SNR and, within that component, may also be able to apply in-band frequency diversity combining.

It is also possible for the same broadcast traffic to be radiated simultaneously from different static sites, giving path diversity in addition to frequency diversity.

3.3 Mobile Terminal-to-Static Terminal

The mobile terminal-to-static terminal situation is, in many respects, the complement of that discussed in Section 3.2. Two basic options exist as indicated in Fig. 12: the first where the mobile must access a single static terminal; the second where a number of alternative static terminal accesses are available. For the situation shown in Fig. 12(a), there may also be occasions when a bi-directional, full duplex link must be engineered between the two terminals. This can be difficult to maintain because of changes in mobile antenna characteristics as it changes position and orientation, coupled with the need for co-ordinated frequency changes.

In Fig. 12(b), the mobile has a choice of static terminals with which to communicate. If the static terminals are widely separated, there will be a high probability that both propagation and co-channel interference effects will be largely decorrelated at the various locations. Thus, it can be expected that at least one of the static terminals will be able to receive the mobile traffic in a reasonably optimum manner over the complete ranges of time and position; this traffic can then be passed to the desired recipient by HF or non-HF communications links from the static terminals.

3.4 Mobile Terminal-to-Mobile Terminal

Fig. 13 shows the two basic methods by which a mobile terminal can communicate with another mobile terminal. In Fig. 13(a), the mobiles communicate directly via skywave or surface wave HF propagation. In the latter case, vertically-polarised antennas mounted on ships allow links to be established over distances up to a few hundreds of kilometres. Skywave links over comparable ranges necessitate high-angle propagation via horizontally-polarised antennas, which may be difficult to mount on ships. Both terminals suffer from the fundamental limitations listed in Section 1.3, and thus direct mobile links are difficult to operate reliably over a wide range of distances.

Fig. 13(b) illustrates an alternative arrangement in which traffic is routed from one mobile to another via a static "relay" terminal. In this case, the advantages of the static station can be exploited in its relay role and communication between mobiles can be maintained over a wider range of distances with greater reliability - at the expense of more sophisticated control procedures.

As stated previously, HF links of various types are normally controlled by using data obtained from off-line propagation analysis programs. More recently, various forms of real-time channel evaluation (RTCE) are being used to aid frequency selection and system control.

Sections 4 and 5 respectively discuss design trends in HF communication systems

(a) involving static terminals alone;

(b) in which mobile terminals are present, possibly in addition to static terminals.

Since (b) presents a more significant and important practical problem, attention will be concentrated in this area. Prior to this discussion, however, the topic of diversity processing will be reviewed briefly because potentially it can provide a substantial improvement in performance for all types of HF communication systems at reasonable cost.

3.5 The Value of Diversity Processing in HF Communications

It has long been recognised that diversity processing of various types can be applied with advantage to HF communications. The most widely-used types are:

(a) spaced receiving antenna diversity;

(b) diversity obtained from receiving antennas of different polarisations;

(c) frequency diversity, either within a given assigned channel or over different assigned channels;

(d) time diversity, in which the same signal is repeated atdifferent time offsets;

(e) diversity arising from the availability of independent multipath components.

All forms of diversity combining require there to be available two or more independently-derived received versions of the wanted signal. Under these circumstances, fades or interference affecting one version may be decorrelated with similar mechanisms affecting the other version(s), thus allowing a better overall signal estimate to be derived by using information from all versions (Stein & Jones, 1967).

Fig. 14 shows the reduction in received SNR admissible, under Gaussian White Noise (GWN) and Rayleigh fading conditions, when 2nd and 4th-order selection diversity combining is used rather than non-diversity reception. It is seen that for dual-diversity combining, an SNR reduction of about 10dB can be expected for an error probability of 1 in a 1000 - equivalent to a ten-fold increase in transmitter power, or the difference between an omnidirectional and a highly directional receiving antenna.

3.6 Speech Transmission at HF

One of the major technical problems associated with HF communication, which has yet to be solved satisfactorily, is that of transmitting digitised, and hence secure, speech reliably over such a time-variable channel. Existing vocoders typically operate at data rates of between 1.2 and 2.4 kbits/s: at such rates, reliable HF transmission cannot be guaranteed and, indeed, in spectrally congested environments such as the central European region, it is improbable that a satisfactory grade of service can ever be achieved - particularly at night.

Two options present themselves: first, users must accept that secure/digitised speech cannot be transmitted reliably over many types of HF channel - a situation which is operationally unacceptable; secondly, more sophisticated speech source encoding and system management schemes must be developed to give an average transmission rate not exceeding a few hundreds of bits/s.

One such system, involving a combination of speech-recognition and synthesis, coupled with user interaction with visual or audio prompts, has been demonstrated practically (Chesmore and Darnell, 1985). The information rate for this syystem is approximately 25 bits/s, making it suitable for use with simple, low-power, mobile terminals.

4. DESIGN TRENDS IN STATIC SYSTEMS

To date, as indicated in Section 3.1, static HF systems have tended to exploit high transmitter powers and directive antennas for both transmission and reception purposes. However, the majority of developments in land-based static networks are now directed towards the design of systems with powers ranging from a few hundreds of W to a few kW; also, since such systems are often required to be redeployable, the antennas used must be of types which can be dismantled and re-erected rapidly. As a consequence, more emphasis has to be placed upon system control and channel coding/decoding procedures in order to achieve the desired level of performance (Darnell, 1983a).

4.1 Real-Time Channel Evaluation

Referring to the window model of HF propagation introduced in Section 2.1: for a static system, the position of the terminals is determined by operational considerations; thus, the problem of system control reduces to one of selecting

values of frequency and transmission time for the given path. This selection procedure can be implemented by employing some form of RTCE. RTCE procedures have been developed for both the open loop and closed loop scenarios introduced in Section 1.4. At present, however, RTCE techniques are only designed to monitor and select the best of a set of alternative transmission channels at a specified time; in general, they are not well matched to the task of monitoring the short-term time variability of those channels. Thus, in the context of the simplified HF channel model postulated in Section 2.1, current RTCE algorithms search for transmission windows which are likely to persist for the complete duration of a fixed constant rate transmission; at the same time, they attempt to minimise the long-term overlap between communication and noise/interference windows.

In future, since the utilisation of the HF medium will remain at a relatively high level, or even increase, more emphasis should be given to the development of RTCE systems which enable the available capacity of the medium to be used more effectively by monitoring the shorter-term time fluctuations of a set of assigned HF channnels. This RTCE data could then be used in the control of an adaptive transmission rate system in which the transmission rate at any time could be matched more accurately to the available capacity. To implement this form of channel monitoring economically, it would be necessary to time/frequency multiplex RTCE probing signals and measurement periods with the traffic signals, possibly with the input data stream being buffered during the probing/measurement intervals.

One of the main disadvantages of many RTCE systems as currently implemented is that they are separate from the communication systems which they are designed to support; they require dedicated and expensive units which are comparable in cost to the units of the communication system itself. The future trend should be for RTCE to be integrated (or embedded) into the communication system design and hence to employ the same basic RF and signal processing equipment (Darnell, 1986b). Also RTCE in the form of ionospheric sounding, in which energy is radiated across a major portion of the HF spectrum, should be restricted to the purposes of ionospheric research and not employed to support communication systems. Rather, integrated communication and RTCE systems operating only in assigned channels should be developed. In this way, the spectral pollution associated with dedicated RTCE systems would be minimised.

4.2 System Control

It should again be emphasised that, in the system control context, RTCE represents the process of identification, or modelling, which must take place before optimal control can be applied. Clearly, the path parameters are time varying and thus any control algorithm needs to be adaptive in response to such changes. If manual control procedures are used, response times will be at best a few tens of seconds and there will be no chance of the system being able to utilise relatively short duration, high capacity transmission windows (analogous to the meteor burst situation). RTCE, however, provides the essential information to enable the system control procedures to be automated - hence potentially providing a greatly improved response time.

The majority of HF systems operate in a closed loop, or 2-way mode. In addition to RTCE, the control of this type of system requires the ability to be able to transfer information concerning reception conditions between receiver and transmitter sites, ie the provision of a high integrity engineering order wire (EOW) facility. It is a fundamental requirement of such an EOW that it should be capable of passing the essential control data reliably for a short interval after the quality of the received traffic has deteriorated to an unacceptably low level due to changes in channel conditions. The transmission rate required of an EOW normally will be considerably lower than that of the traffic channel, thus enabling robust channel encoding (modulation and coding) procedures to be adopted for protection of the system control data.

The system control procedures should also allow for:

(i) Continuous checking of traffic quality: an automatic repeat request (ARQ) technique would be suitable for this purpose.

(ii) Automatic re-establishment of the circuit should contact be lost: probably based upon the use of back-up accurate time and frequency transmission schedules.

In the case of an open loop configuration with no EOW available, eg a broadcast system, the potential for adaptive operation is relatively limited. All adaptation must be carried out at the receiver in response to RTCE data embedded in the transmitted signal. Receiver processing techniques such as equalisation, adaptive filtering, diversity combination of two or more independent versions of the transmitted signal, antenna null and beam steering, etc are relevant in this situation.

From the preceding discussion, it is evident that automatic HF system control requires the application of considerable information processing power. Therefore, future HF communication systems will inevitably be processor based, with manual intervention only in exceptional circumstances.

4.3 Channel Encoding/Decoding

The purpose of the channel encoding procedure in a communication system is to condition the information to be transmitted in such a way that it is resistant to the types of noise and other perturbations likely to be encountered during its passage over the channel. The key aspects of channel encoding are therefore modulation and error control coding.

The range of channel encoding/decoding techniques, which offer promise of economic and effective implementation with variable capacity HF systems, has expanded considerably with the advent of increased signal processing power.

As has been indicated previously, signal processing in existing HF systems tends to be relatively simple and based upon the premise of non-adaptive, constant rate operation. If variable rate operation is contemplated, a fresh consideration of potential signal processing techniques is necessary. Those procedures which appear to offer promise are now outlined briefly.

(a) Adaptive filtering at the receiver: in which the parameters of baseband or IF receiver filters are modified to counter changes in the nature of interfering signals in the communication channel.

(b) Quenched filtering: applied to a synchronous transmission system enables a receiver filter to be opened to receive a wanted signal only over the intervals when that signal is expected to be present. In this way, the filter is not initialised by noise or interference and the detection decision is made when the detector output SNR can be expected to be a maximum. This type of detection scheme is used in the PICCOLO system (Bayley & Ralphs, 1972). For a given form of modulation, if synchronous or coherent detection can be applied, a significant improvement in system performance can be achieved.

(c) Soft-decision decoding implies that a "hard" detection decision, say 1 or 0, is augmented by information concerning the confidence level associated with that decision, eg based upon signal amplitude, phase margin, etc. When coupled with error protection coding, soft-decision decoding can enhance the performance of a basic detection system substantially (Chase, 1973).

(d) Possibly one of the most powerful techniques for use in a high interference environment is that of correlation reception. For this, sets of relatively long sequences with approximately impulsive autocorrelation functions are required so that matched filtering can be applied at the receiver. Sequences with suitable autocorrelation properties also allow RTCE data, say in the form of a channel impulse response, to be extracted conveniently.

(e) Bearing in mind the variable nature of HF channel capacity discussed previously, it would appear that some form of "embedded" data encoding would be appropriate. Here, several versions of the source data, each keyed at a different rate, would be combined into a single transmitted signal format. At the receiver, data would be decoded at the highest rate which the channel capacity at that time would allow. System control and re-alignment would be accomplished at intervals by an ARQ arrangement operating via the EOW (Darnell, 1983a).

(f) With the advent of cheap and powerful computing capacity, it is now possible to consider the introduction of post reception processing at the receiver. The unprocessed received signal could be stored, either at a low IF or baseband, and then processed rapidly off-line using different signal processing parameters and algorithms until a best estimate of the transmitted signal is obtained.

4.4 Radiated Power Levels

In the HF band, high radiated power levels are undesirable since they tend to cause spectral pollution and an increase in transmitter and antenna sizes - and hence costs. The use of RTCE in optimising the primary parameters of an HF communication system will tend to reduce the necessity for higher radiated powers by selecting frequencies and transmission times for which the received SNR is maximised and interference avoided. As a general principle, system availability and reliability should be improved by the use of RTCE and more effective signal processing, rather than by transmission at higher power levels; the latter should be held in reserve for particularly vital links.

The RF units of the system should be capable of rapid frequency agility so that full advantage can be taken of automatic and adaptive system control procedures. The ability to change frequency rapidly will also facilitate the RTCE process of finding a channel with an acceptable SNR. This will eliminate the "resistance" of system operators to changing frequency except where absolutely essential and thus lead to improved flexibility and lower radiated power levels.

4.5 Antenna Characteristics

Diversity using spaced receiving antennas is an effective method of improving HF system performance and should, where possible, be incorporated into a system design. Also, a simple 2-element receiving antenna system can be used to place a null in the

direction of an interfering signal. The depth of the null will depend upon the instantaneous propagation conditions, but an average depth of a few dB may well be achievable for skywave signals. Against more stable groundwave interfering signals, the nulling procedure will be considerably more efficient.

Again, the systematic use of RTCE and improved system control procedures will tend to reduce dependence upon antenna characteristics.

4.6 System Design with Embedded RTCE

It is clear that, particularly in a high-interference environment, a constant rate transmission system will provide a variable level of performance, and at rates of 2.4 kbits/s is likely to have a low overall reliability. Many regions of the world have high interfering signal densities such that the probability of finding a clear frequency slot in which to place a 2.4 kbits/s transmission is extremely low (Gott et al, 1983). Thus, it can be inferred that requirements for constant data rate transmissions of this form are in many cases incompatible with the fundamental nature of the propagation medium when its properties are considered on a world-wide basis. This has a particularly serious implication for digital speech links, which normally operate at rates in the range 1.2 - 2.4 kbits/s, since in regions where the HF spectrum is congested, eg in central Europe, they can rarely be expected to provide a satisfactory grade of servie over extended periods. The use of real-time channel evaluation (RTCE) to select the optimum assigned channel at any instant will only be of limited effectiveness in such an environment because a channel with the required transmission characteristics will frequently not be available.

There are certain techniques which can smooth out the variations in channel capacity as seen by a non-adaptive communication system (Darnell, 1977). In the context of the window model described earlier, this corresponds to a reduction in required SNR with a consequent expansion of window dimensions. These techniques include:

(a) diversity processing;

(b) adaptive equalisation;

(c) error control coding;

(d) antenna beam-forming.

In practice, RTCE can be used to select a frequency channel having suitable propagation characteristics and clear of significant levels of co-channel interference. This then enables techniques such as those listed above to be applied more effectively to increase the transmission window volume.

The potential of the ionospheric medium to transfer certain forms of traffic reliably, even in the face of high levels of interference or disruption, is of great operational importance. However, in order to realise this potential fully, HF communication systems must be designed and operated in a rather different manner from that associated with the majority of systems currently in service. Fig. 15 shows the elements of a duplex ionospheric communication system incorporating many of the features discussed previously. It is assumed that the system will use any appropriate mode of propagation. The function of the RTCE sub-system is to select the channel and characterise it in terms of its window structure as a function of required SNR at the receiver; this then allows the parameters of the signal processing unit to be adjusted to match the nature of the available windows, eg in terms of block length, modulation type, etc. Data is buffered until the transmission system is ready to accept it, with simultaneous transmission in each direction being based upon an ARQ mode of operation where control data and traffic signals are interleaved. At the receiver, the incoming data blocks are re-assembled in another buffer prior to output. It should be noted that both data and RTCE signals are multiplexed together and thus make use of single transmitters and receivers at each site, with no dedicated RTCE RF equipment, ie an embedded RTCE system. Receivers would normally be of the diversity type.

5. DESIGN TRENDS IN SYSTEMS INCORPORATING MOBILE TERMINALS

As explained previously, mobile terminals represent an important class of disadvantaged users of the HF spectrum, and it is important that the energy they do radiate is received in the most efficient manner possible. The technique offering most promise of achieving this may be termed "geographical diversity reception".

In Section 3.3, a network of widely separated static terminals for the reception of mobile transmissions was introduced. This concept will now be extended to that of a formal geographical diversity architecture in which the data from a number of spaced receiving sites can be combined in a systematic manner to provide an improved estimate of a desired mobile transmission. The value of such path diversity in improving communication availability has been demonstrated experimentally (Rogers & Turner, 1985).

Fig.16 is a schematic diagram of a geographical diversity network involving multiple static terminals interconnected by means to be described in Section 5.1.

The basis of operation of the network is that transmissions emanating from any given mobile terminal will be received at static terminals via propagation paths which will normally be skywave, but may also be groundwave; these paths can be considered as independent if the separation of the static terminals is chosen appropriately. Similarly, interfering transmissions will also be received via independent paths. Thus, taking the network in total, there will be a high probability that the signal and interference conditions on those links which are propagating between the mobile and the static terminals will be uncorrelated. This situation lends itself to some form of diversity combining in which all received versions of the mobile signal are brought together at a central control and processing site to allow an improved signal estimate to be made. There are several possible combination algorithms which could be considered, eg:

(a) selection of the path giving the greatest SNR at any time, ie selection diversity combining;

(b) alternative classical forms of diversity combining such as maximal ratio (Stein & Jones, 1967);

(c) operation on the demodulated and shaped versions of the propagating received signals by say majority voting;

(d) more elaborate combining techniques applied to demodulated and shaped signals, taking account of soft-decision data associated with the received symbols.

The most important considerations relevant to the establishment of a geographical diversity network will be reviewed in the following sections.

5.1 Interconnection Considerations

The static terminal interconnections shown in Fig.16 are clearly vital to the operation of the system; they must therefore be robust and reliable. The following media may be considered for interconnection purposes:

(a) HF skywave and groundwave point-to-point links;

(b) meteor-burst links;

(c) ionospheric scatter or tropospheric scatter links;

(d) satellite links;

(e) some form of common-user telephone/telecommunication system, eg such as those implemented by PTT's.

Links of types (a), (b) and (c) above are essentially "stand-alone" point-to-point interconnections, whilst (d) and (e) require the availability of separate communication systems, possibly controlled and operated by independent organisations. Hence, from the viewpoint of survivability in the face of physical or electromagnetic attack, (a), (b) and (c) are to be preferred. The various options will now be considered in more depth.

5.1.1 HF Interconnection

In a geographical diversity network of the form proposed, fixed receiving terminals may be separated by distances from say a few hundreds of kilometres to several thousands of kilometres. Interconnection via HF links will therefore involve skywave and, possibly, groundwave paths exhibiting a wide variety of propagation and interference characteristics. There is a need for data transfer between terminals to be effected with high integrity and resistance to disruption; this will require links with sophisticated embedded RTCE (Dawson & Darnell, 1985), together with adaptive error control (Hellen, 1985). Inevitably, the data rates which can be sustained reliably will be relatively low.

5.1.2 Meteor-Burst Interconnection

Meteor-burst links (Bartholomé & Vogt, 1968) make use of radio wave energy reflected from the short-duration ionised trails caused by the passage of micrometeorites through the lower ionosphere. Such links are naturally intermittent, although the statistics of meteor trail occurrence are relatively predictable. It has been demonstrated experimentally that average data rates of a few hundreds of bits/s can be sustained over a complete 24-hour period; this would normally be adequate for the role proposed here. The diurnal and seasonal variability of meteor-burst links is less than for corresponding HF skywave links; they are also less susceptible to excess absorption effects following severe ionospheric disturbances.

The most effective frequency range for meteor-burst operation is from about 25 to 100 MHz which allows compact, efficient and directive antennas to be used.

5.1.3 Ionospheric Scatter or Tropospheric Scatter Interconnection

Ionospheric scatter propagation makes use of radio wave energy scattered by the lower

ionosphere (Bartholomé & Vogt, 1965). Tropospheric scatter propagation employs energy scattered by irregularities in the tropospheric region of the earth's atmosphere (Yeh, 1960). Since the troposphere is at a much lower altitude than the ionosphere, the ranges achievable via this scattering mechanism are considerably less than via ionospheric scatter. In both cases, the scatter loss is large thus necessitating the use of high transmitter powers and directive antennas to illuminate the "scattering volume" efficiently.

Propagation using these scattering mechanisms is relatively reliable, although significant fading and time dispersion may occur.

5.1.4 Satellite Interconnection

Inexpensive and reliable satellite links are becoming available, although there may be coverage limitations at high latitudes. In the military context, the overall HF system should function as an independent entity under threat conditions, which might include the satellite being disrupted. For this reason, such links are not considered suitable for interconnection purposes, although they may be useful under benign operating conditions.

5.1.5 Common-User System Interconnection

Possibly the cheapest form of interconnection is to employ a PTT-type of network in which the static HF terminals, control sites and message originators and recipients are subscribers. Again, this medium has the disadvantage that it can be disrupted by physical or electromagnetic attack. However, the diversity of propagation media and routes makes it inherently more robust than a satellite system alone.

Practically a combination of interconnection media would appear to offer the best prospect of reliable operation over the whole range of possible operational conditions. Transmission systems covering say the frequency range from 2 to 60 MHz, having reasonably high radiated powers (a few kW) and directive antennas, could form the primary basis for interconnection; they could operate in HF skywave, HF groundwave, meteor-burst or ionospheric scatter modes, as available and appropriate. Also, a parallel interconnection might be established via a PTT network.

Obviously, some delay will be introduced into the process of mobile signal reception by the need to transfer data around the interconnection network. This delay will tend to increase as the number of mobile terminals accessing the network simultaneously increases. The longer the delay that can be tolerated, the greater is the potential for accurate reception. In general, it can be expected that the static terminal interconnections will be able to sustain a higher average data rate than that which will be associated with the mobile-to-static terminal transmissions. The effect of delays can be minimised by the signal design and network control strategies to be discussed in the following sections.

5.2 Signal Design Considerations

From the discussion in Sections 1 and 2, it is seen that the ionosphere can be subject to severe natural disturbances giving rise to multipath delay, signal fading, low signal levels and possibly abnormal doppler spreads or extreme phase instability in polar regions. Under these circumstances, it is necessary to use robust transmission formats that are resilient to such perturbations.

5.2.1 Mobile Signal Formats

Normally, it will be necessary for a geographical diversity network to handle signals originating from several distinct mobiles. A block transmission format would be convenient since it would allow simple time and frequency multiplexing to take place, with only one mobile using a particular frequency at a specific time. To allow this, accurate time and frequency standards would have to be available at all terminals in the system, but this can now be accomplished relatively simply and economically.

The major requirements for a robust mobile transmission can be summarised as follows:

(a) that its time of occurrence and frequency of propagation are accurately controlled;

(b) that the SNR required at the receiver for a specified error probability is minimised;

(c) that no elaborate synchronisation techniques are required;

(d) that long multipath delays can be tolerated;

(e) that high levels of phase instability can be tolerated;

(f) that moderate doppler frequency offsets can be tolerated;

(g) an ability to operate in any of a limited number of modes which can become progressively more robust as propagation conditions deteriorate;

(h) that signals should be compatible with digital encryption techniques.

A possible signal format on which the mobile transmission could be based is basic wide-shift FSK, in which detection is accomplished by a Law assessor arrangement (Alnatt et al, 1957) giving the effect of dual, in-band diversity under frequency-selective fading conditions. Simple block error control codes could be applied to the data blocks, with more powerful (ie more redundant) coding being introduced as channel conditions worsened. In extreme cases, the mobile might simply repeat its message many times to allow signal averaging to be employed at the receivers; note that, under additive GWN conditions, signal averaging improves the effective SNR by a factor proportional to the square root of the averaging time (Hewlett-Packard J, 1968).

Under conditions of poor SNR, a trade-off must be made between SNR and transmission time; ie as the SNR reduces, the user must allow more time for the transmission of a given message if it is to be received with a specified level of fidelity. In essence, some form of variable-redundancy error control coding is required (Hellen, 1985).

In circumstances where frequency offsets are negligible and short-term phase stability reasonable, systems of the PICCOLO type (Bayley & Ralphs, 1972), or reduced versions in the form of multiple FSK systems, would provide robust communication.

In general, transmissions should be designed so that soft-decision, or confidence, information can be extracted readily at the receiver.

5.2.2 Interconnection Signal Formats

Most of the comments made above in the context of mobile signals also apply to signals transferred over the interconnection links between static terminals.

Labelled data blocks describing signals from specific mobiles would be a convenient vehicle for passing data between the static nodes and the control/processing site.

As also indicated previously, the design and operation of reliable point-to-point links is a simpler problem than is the design of a mobile-to-static link because more efficient antennas and higher transmitter powers are available. Again, an ARQ mode of operation would seem to be appropriate (Dawson & Darnell, 1985), since this would be well matched to the characteristics of the interconnection media proposed, ie HF skywave and groundwave, meteor-burst or ionospheric scatter.

5.3 System Control Considerations

From the previous section, one of the potential problems that can be identified is that the maximum capacity required for the interconnection links is considerably greater than for the mobile-to-static links. This comes about because the interconnection links may be simultaneously carrying signals relating to:

(a) a number of distinct mobile transmissions;

(b) system control data for the network.

These will now be discussed in greater detail, together with other considerations affecting the control and management of the network.

5.3.1 Data describing Mobile Transmissions

One mode of operation for a geographical diversity system is where the various versions of the mobile signal received at the separated static receiving terminals are demodulated and reconstituted at those terminals, with the reconstituted data being forwarded to the site where combination is to be carried out. A further refinement of this process might be the transmission of soft-decision data, relating to each of the demodulated symbols, to the combination site. Alternatively, the complete analogue baseband signal from each static receiving terminal could be forwarded to the combination site. In the latter two cases, the information carried by the interconnecting links would be considerably expanded in comparison with that contained in the original signal transmitted by the mobile. However, since the mobile transmissions would normally be of short duration, this data expansion would be acceptable if some decoding delay could be tolerated.

5.3.2 System Control Data

It will also be necessary for system control data to be transferred on a regular basis between static terminals, simply to maintain the integrity of the network. This system control data would relate to frequency selection options, available data rates, numbers of mobile accesses, changes in coding levels, etc. It is envisaged that, in an ARQ-based system, the capacity overhead associated with these forms of system control data would not be excessive.

5.3.3 Frequency Management

Frequency management in a geographical diversity system is required at two levels:

(a) on a link-by-link basis;

(b) on a network-wide basis.

In the former case, each of the interconnection links would have its own embedded RTCE and protocols (Darnell, 1986b) operating over the frequency range from 2 to approximately 60 MHz. The propagating mechanism giving the largest predicted SNR, and hence reliability, would normally be selected.

At the network level, there is a need for a more global RTCE procedure which is capable of assessing the general state of propagation over the operational region encompassing both the static and mobile terminals. This could be accomplished by either a vertical-incidence ionosonde placed at the eastern edge of the region, or by oblique-incidence sounding carried out on an opportunity basis between the various static terminals of the network. Measurements of this type would be co-ordinated by a designated control site and might, for example, be used to detect the onset of irregular ionospheric disturbances and to indicate their geographical extent.

An additional function of the overall frequency management procedures would be to ascertain the degree to which frequencies could be shared or re-used by different static terminal interconnection links.

5.3.4 Other Control Functions

In military operations, the control of a communication system must, at least in part, be directed towards maintaining the security of the mobile traffic. Thus the mobile would, for example, be assigned transmission frequencies which would have the minimum probability of being intercepted or disrupted under the prevailing propagation conditions. Also, the mobile would exercise radiated power control (RPC) in response to recommendations sent to it by the overall system controller.

In addition, the system control procedures would seek to make maximum use of relatively local and transient propagation phenomena as they became available. There would also be a need to reconfigure the static terminal interconnections in order to avoid localised ionospheric disturbances. To illustrate this, Figs.17 and 18 show the benefits to be gained from geographical diversity in this respect. Fig.17 is a situation in which an ionospheric disturbance affects the skywave path between a mobile (M) and a static terminal (F1); the skywave path to the static terminal (F2), at a greater distance from M, avoids the disturbed region completely, as does the direct interconnecting path between F1 and F2. Fig.18 represents a more complex situation where the disturbed region affects both the mobile-to-F1 path, together with the interconnections F1-to-F3 and F1-to-F4. In this latter example, the interconnecting links must be re-routed as shown to maintain network connectivity.

Clearly, similar reconfiguration procedures could be employed to counteract any deliberate attempts to disrupt transmissions from a mobile to the static terminals.

5.4 System Design

An attempt will now be made to coalesce many of the concepts outlined in the previous sections of this lecture into an overall system design to provide improved HF communications reliability and survivability.

Fig.19 illustrates a possible format for such a design in which mobile-to-static transmissions are received by a geographical diversity network, whilst static-to-mobile transmission is carried out via a single broadcast system sited at a convenient location.

System control, together with signal generation and combination, are carried out at a designated control and processing site. Interconnections are implemented by 2 to 60 MHz ionospheric links, or via a PTT-type network, as appropriate. Broadcast traffic for all mobiles within the operational area is transmitted using a multiple sub-channel format, from which any given mobile can extract the sub-channel(s) addressed to it. The EOW sub-channel of the broadcast is used to request repeats of mobile data blocks received in corrupted form, and also to disseminate frequency management data to all mobiles wishing to access the system.

Data transfer between the HF system and the message originators and recipients can be effected via a PTT network, or possibly via other dedicated HF or non-HF links.

Fig.20 shows an alternative system format where the centralised broadcast system of Fig.19 is replaced by a distributed network of lower power transmitters, possibly co-sited with the static receiving terminals. Thus, in principle, for any specified mobile position, a broadcast transmitter site having maximum probability of successfully propagating to the mobile can be selected by the system frequency management procedures - as illustrated in Fig.20.

6. CONCLUDING COMMENTS

This lecture is intended to complement the discussion of "HF groundwave and skywave propagation" contained in the previous lecture, and also to relate the effects described in that lecture to the problems of practical HF communication system design.

6.1 General Conclusions

From the discussion in this paper, the following points are worthy of emphasis:

(a) That the variable capacity nature of the HF channel should be recognised and its characteristics exploited to advantage in the design of HF communication systems.

(b) That improved system performance should be achieved primarily by the application of more advanced signal generation/processing and system control procedures, rather than by the application of higher transmitter powers and large, directive antenna arrangements.

(c) That the key to improved performance is the effective integration and use of RTCE techniques in an HF system design, ie embedded RTCE.

(d) That HF systems should have a capability of operating in a frequency range above the traditional upper limit of about 30 MHz. This would provide additional flexibility in the choice of propagation mechanism, and also resistance to ECM/ESM activity.

(e) That the most effective method of enhancing the communications reliability of mobile terminals would appear to be via the use of geographical diversity network for reception purposes. A distributed network of the type illustrated in Figs.19 and 20 poses additional control problems; however, the resultant increase in operational flexibility and robustness potentially available from such a system would seem to outweigh possible disadvantages. In particular, an ability to reconfigure in the face of physical attack, electromagnetic attack, or natural ionospheric disturbances, would be operationally valuable.

(f) That the price paid by an interceptor or jammer for his activities should be maximised - both from the viewpoint of cost and that of the complexity of analysis he must employ.

(g) That the HF propagation medium should be used for the types of communication traffic which are compatible with its fundamental properties - see Section 6.2 below.

6.2 Natural Roles for HF Communication Systems

Following from point (g) of the previous section, it would appear that the forms of traffic which are best matched to the characteristics of the HF propagation medium include:

(a) Transmissions involving a low constant data rate, R, eg low-speed telegraphy, where R conforms to expression [5].

(b) Transmissions involving a low average data rate, where the information can be block-formatted and transmitted in packet form - possibly over a wide range of rates which may vary from block to block.

(c) Transmissions in which it is important to transfer a limited amount of information reliably, perhaps in the face of deliberate attempts at interception and disruption.

The transmission of digitised speech at constant rates between 1.2 and 2.4 kbits/s will always be difficult to maintain with an acceptable level of reliability; this is especially true of regions such as Europe where spectral congestion is high. Other techniques, such as the combination of speech recognition and synthesis introduced in Section 3.6, offer promise of digital speech message transmission at very low rates.

6.3 Aspects of System Design Requiring Further Investigation

From the viewpoint of HF system design, the following specific topics require further investigation and experimentation:

(a) The problems of networking at HF, and particularly the control and management system capacity "overheads" involved.

(b) The design and implementation of efficient embedded RTCE procedures, probably in conjunction with software modems and codecs.

(c) The design of adaptive signal generation and processing equipments.

(d) The provision of more realistic simulation facilities to aid HF system

development, especially in respect of co-channel noise/interference models and the effects of RF equipment characteristics, eg non-linearity and AGC.

6.4 Basic Propagation Measurements/Data Still Required

In the context of this Special Course, it is important to identify the specific forms of propagation measurements/data which, if available, would ease the problems of the HF communication system designer and allow more realistic simulation of the HF environment (see item (d) in the previous section) in the design phase of a system. The main deficiencies are seen to be in the characterisation of:

(a) Co-channel noise/interference in the various regions of the earth's surface in terms of probabilistic models describing

- levels
- bandwidths occupied
- decorrelation with distance
- persistence intervals
- variations with diurnal, seasonal and sunspot cycles.

(b) Propagation in distributed networks involving geographical diversity combining in terms of

- long-term (hour-by-hour) similarity of paths from a given transmitter to a number of separated receiving sites
- shorter-term (second-by-second) similarity of the same types of path
- optimum location of separated terminals to meet specified operational requirements.

(c) Data obtained from RTCE systems in a form that can be used to enhance the accuracy of models used in off-line propagation analysis programs.

(d) Simplified propagation analysis models of a form which can be implemented on a microcomputer system and hence embedded within the communication system equipment, eg to initialise an RTCE system.

(e) Time correlation between meteor-burst, ionospheric scatter and conventional HF skywave propagation modes.

(f) Sporadic E-layer propagation with a view to producing more precise predictive models.

(g) Situations in which the probability of propagation to a defined third location, geographically separated from the communications transmitter and receiver, is minimised.

7. REFERENCES

1. Darnell, M., 1986a: "The design of static and mobile HF communication systems", AGARD Lecture Series No.145 on "Propagation impact on modern HF communication system design".

2. Darnell, M., 1985 : "Techniques for improving the reliability of mobile HF communications over high-absorption paths", AGARD EPP Symposium on "Propagation effects on military systems in the high-latitude region", AGARD CP-382.

3. Darnell, M., 1983: "Real-time channel eveluation", AGARD Lecture Series No.127 on "Modern HF communications".

4. Barghausen, A.F., Finney, J.W., Proctor, L.L. and Schultz, L.D., 1969: "Predicting the long-term operational parameters of high-frequency sky-wave telecommunication systems using the ionospheric transmission channel", ITS Report, Boulder, USA.

5. Lloyd, J.L., Haydon, G.W., Lucas, D.L. and Teters, L.R., 1981: "Estimating the performance of telecommunication systems using the ionospheric transmission channel", ITS Report, Boulder, USA.

6. Darnell, M., 1984: "Interference characterisation and its importance in HF communications", Ionospheric Effects Symposium - 84, Washington D.C., USA.

7. Watterson, C.L., 1975: "A review of channel simulation techniques at the Institute for Telecommunication Sciences", ITS Report, Boulder, USA.

8. Darnell, M., 1982a: "Future HF system architecture", Proc. of IEE Internat. Conf. on "HF communication systems and techniques", London.

9. Gott, G.F., Dutta, S. and Doany, P., 1983: "Analysis of HF interference with applications to digital communications", Proc. IEE (Part F), Vol. 130. No. 5, August.

10. Bartholomé, P.J. and Vogt, I.M., 1968: "COMET - a new meteor-burst system incorporating ARQ and diversity reception", IEEE Trans., Vol. COM-16, April.

11. Thrane, E.V., 1983: "Problems in HF propagation", as (3).

12. Darnell, M., 1982b: "Limitations on ECM effectiveness in HF communication", AGARD EPP Symposium on "Propagation effects of ECM-resistant systems in communication and navigation", AGARD CP-331.

13. Stein, S. and Jones, J.J., 1967: "Modern communication principles", McGraw-Hill.

14. Chesmore, E.D. and Darnell, M., 1985: "A low-rate digital speech message transmission system for HF engineering order wire (EOW) applications", Proc. IEE Internat. Conf. on "HF communication systems and techniques", IEE CP-245, London.

15. Darnell, M., 1986b: "Embedded real-time channel eveluation techniques", as (1).

16. Bayley, D. and Ralphs, J.D., 1972: "PICCOLO 32-tone telegraph system in diplomatic communication", Proc. IEE, Vol. 119, No. 9.

17. Darnell, M., 1977: "Medium-speed digital data transmission over HF channels", Proc. IERE Internat. Conf. on "Digital processing of signals in communications", Loughborough, UK.

18. Rogers, D.C. and Turner, B.J., 1985: "Connectivity improvement through path and frequency diversity", as (14).

19. Dawson, J.F. and Darnell, M., 1985: "An HF system design with embedded channel evaluation", as (14).

20. Hellen, P., 1985: "The provision of computer quality data transmission on HF", as (14).

21. Bartholomé, P.J. and Vogt, I.M., 1965: "Ionoscatter communications: new design concepts and experimental results", SHAPE Technical Centre Report TR-83.

22. Yeh, L.P., 1960: "Simple methods for designing troposcatter circuits", Trans. IRE, Vol. CS-8, No. 3.

23. Alnatt, J.W., Jones, E.D.J. and Law, H.B., 1957: "Frequency diversity in the reception of selectively fading binary frequency-modulated signals", Proc. IEE, Vol. 104 (Part B).

24. Hewlett-Packard Journal, 1968: Special issue on "Signal averaging", April.

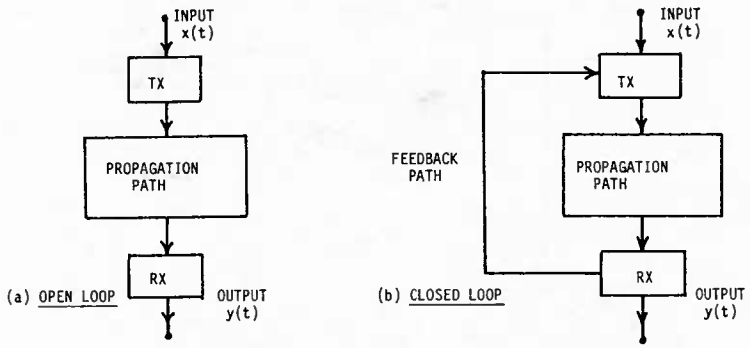


FIG. 1 BASIC COMMUNICATION SCENARIOS

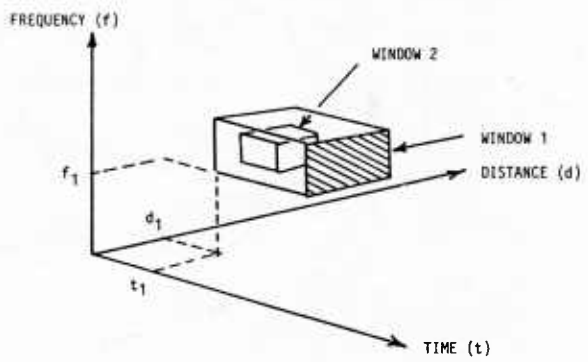


FIG. 2 WINDOW MODEL OF IONOSPHERIC PROPAGATION

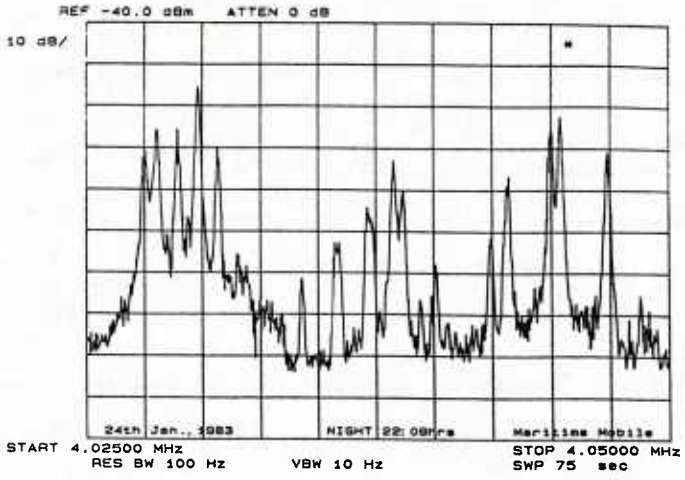


FIG. 3 EXAMPLE OF HF SPECTRUM OCCUPANCY

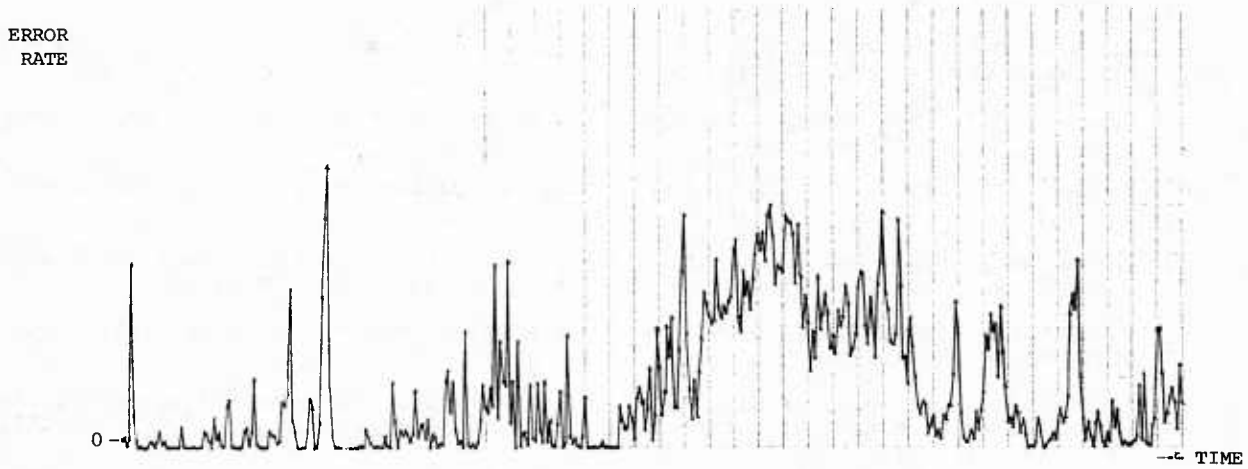


FIG. 4 ERROR RATE PROFILE FOR AN HF CHANNEL

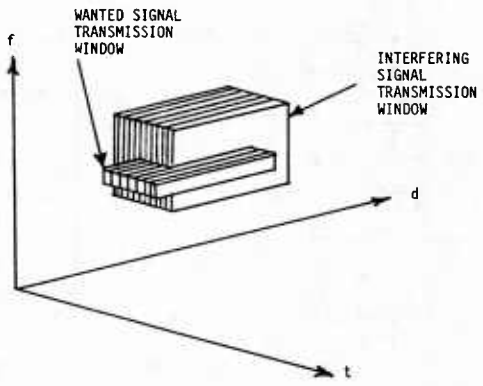


FIG. 5 EFFECT OF INTERFERENCE IN THE WINDOW MODEL

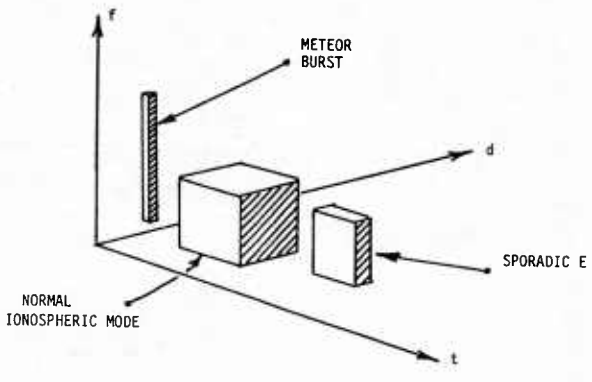


FIG. 6 WINDOWS FOR DIFFERENT PROPAGATION MECHANISMS

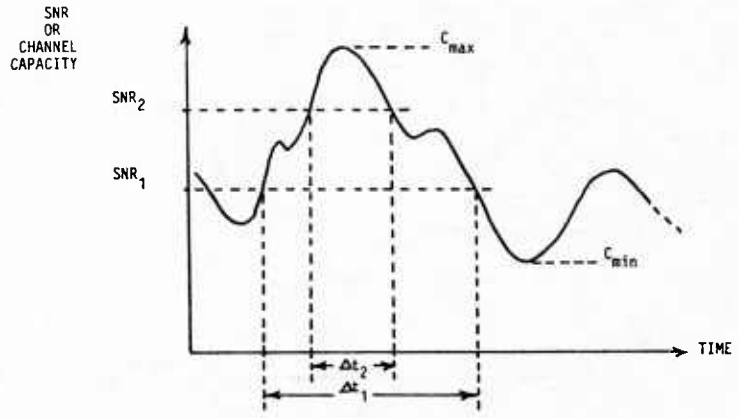


FIG. 7 VARIATION OF SNR/CHANNEL CAPACITY WITH TIME

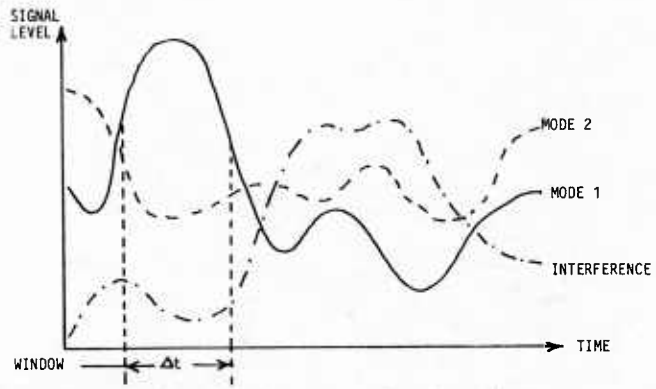


FIG. 8 EXAMPLE OF A PHYSICAL MECHANISM GIVING RISE TO IONOSPHERIC WINDOWS

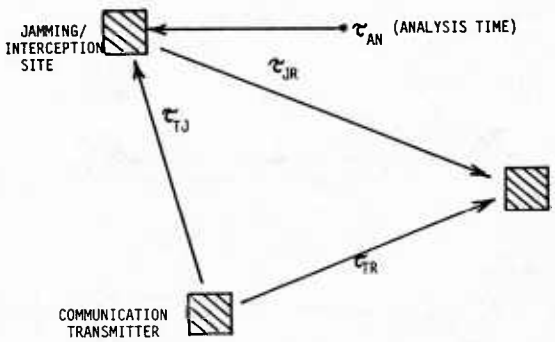


FIG. 9 PROPAGATION AND ANALYSIS TIME LIMITATION

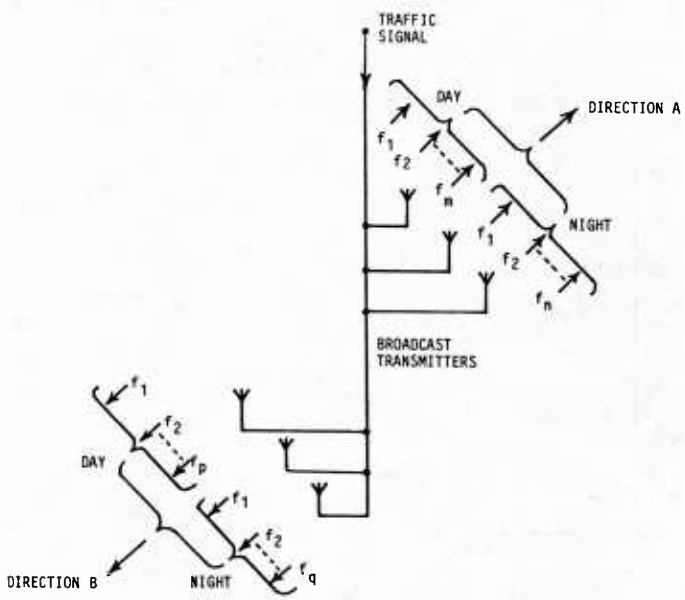


FIG. 10 FIXED-TO-MOBILE BROADCAST SYSTEM

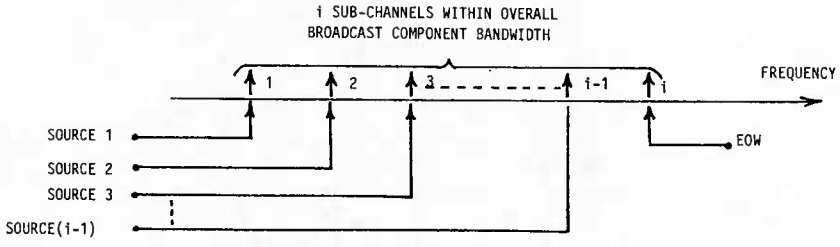


FIG. 11 BROADCAST SUB-CHANNEL RASTER

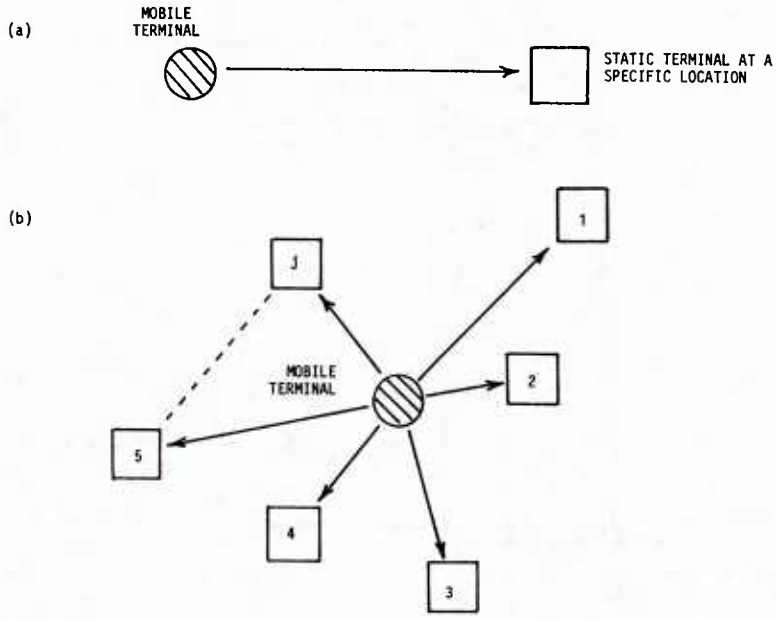


FIG.12 OPTIONS FOR MOBILE TERMINAL-TO-STATIC TERMINAL COMMUNICATION

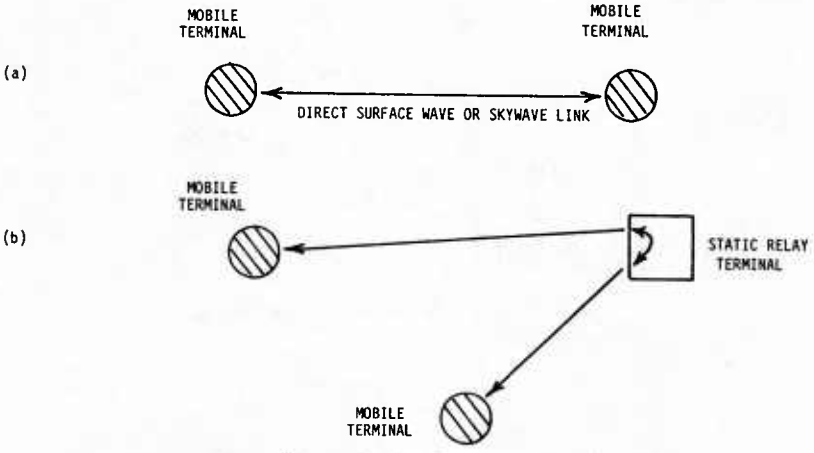


FIG. 13 OPTIONS FOR MOBILE TERMINAL-TO-MOBILE TERMINAL COMMUNICATION

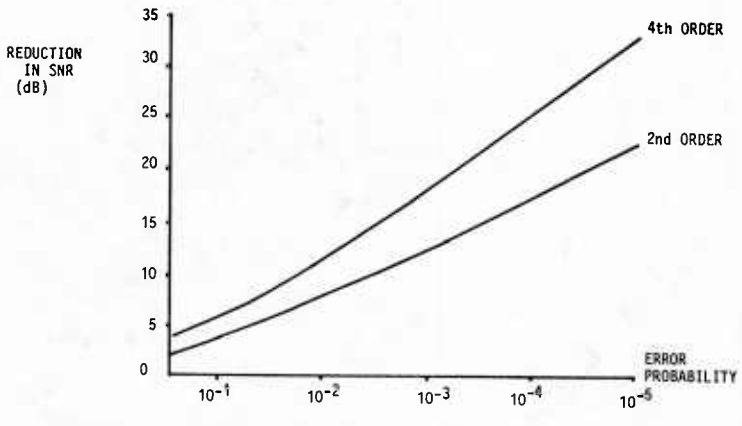


FIG. 14 EFFECT OF 2nd AND 4th ORDER DIVERSITY COMBINING

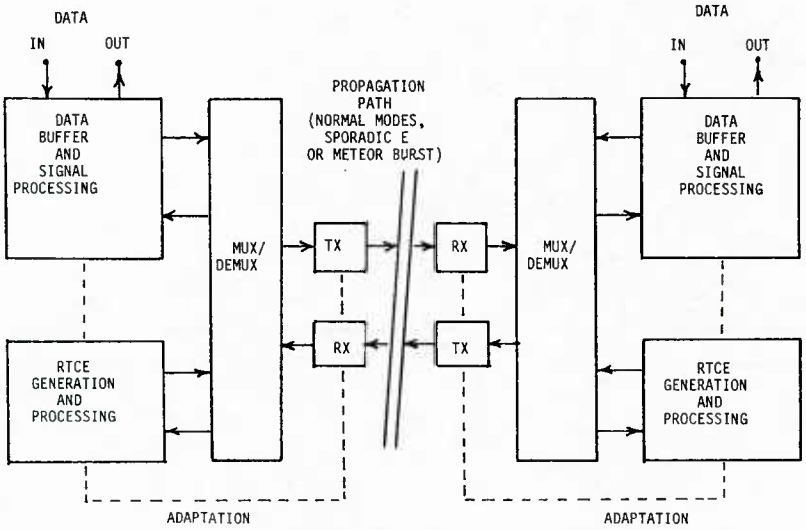


FIG. 15 SIMPLIFIED BLOCK DIAGRAM OF IONOSPHERIC COMMUNICATION SYSTEM

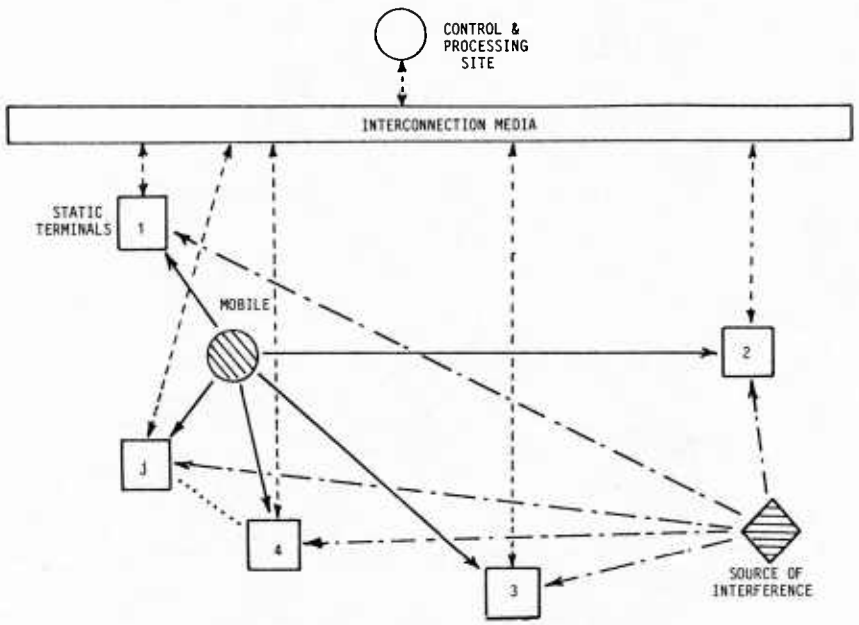


FIG. 16 SCHEMATIC DIAGRAM OF GEOGRAPHICAL DIVERSITY SYSTEM

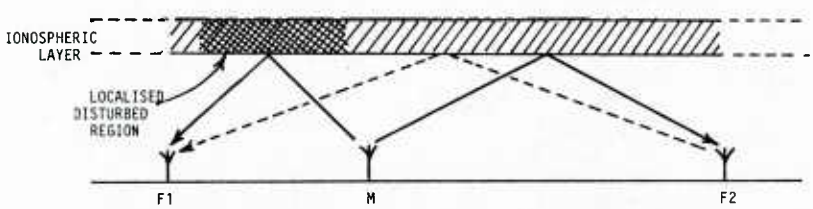


FIG. 17 AVOIDANCE OF A REGION OF LOCALISED IONOSPHERIC DISTURBANCE

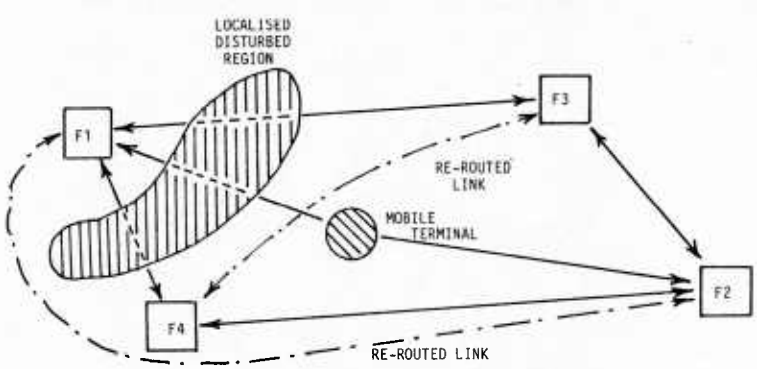


FIG. 18 RECONFIGURATION OF A GEOGRAPHICAL DIVERSITY NETWORK

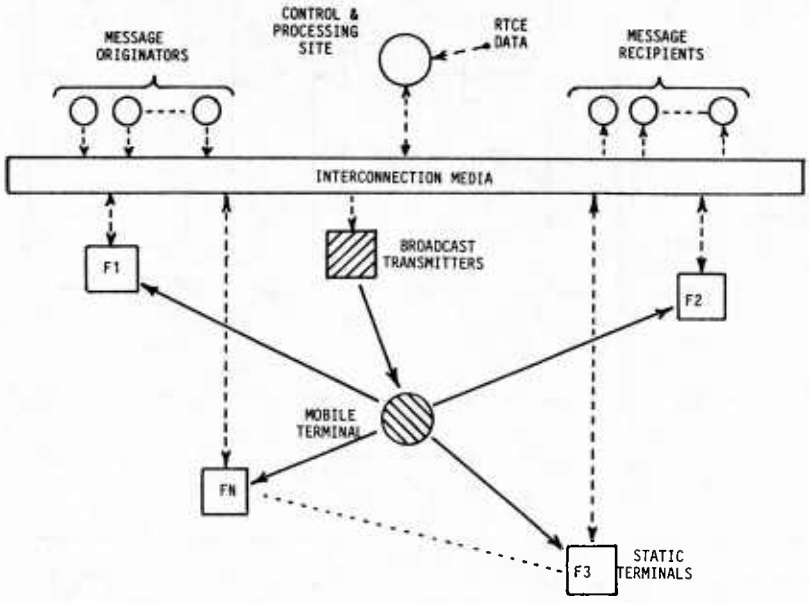


FIG. 19 COMMUNICATION SYSTEM FORMAT 1: CENTRALISED BROADCAST

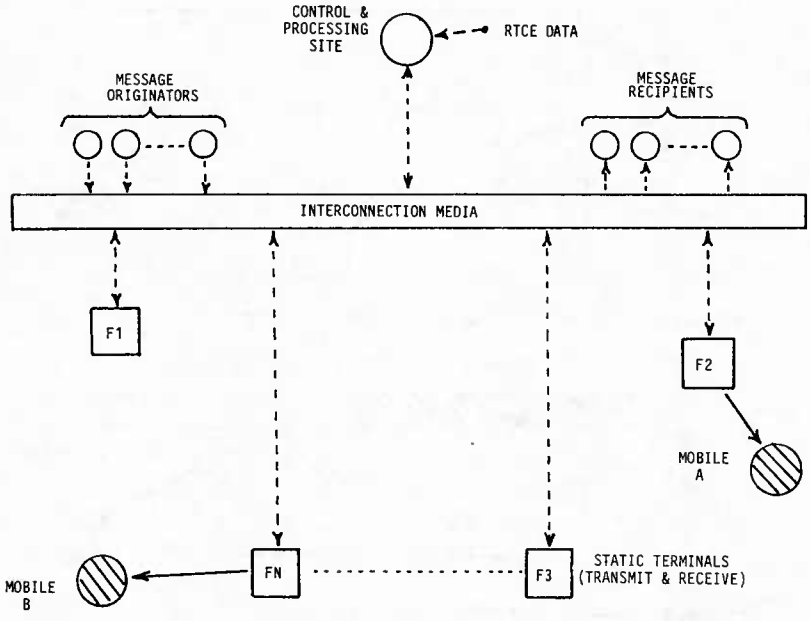


FIG. 20 COMMUNICATION SYSTEM FORMAT 11: DISTRIBUTED BROADCAST

METEOR BURST PROPAGATION AND SYSTEM DESIGN

Jens Østergaard
ElektronikCentralen
Hørsholm, Denmark.

BACKGROUND

The first truly operational meteor scatter communication system, JANET (Forsyth et al, 1957), was operated in Canada in the mid 1950's at approximately 40 MHz. However the system was severely impaired during a Solar Protron Event during testing between Yellowknife and Edmonton (Crysdale,1960), and experimental interest in meteor scatter propagation essentially vanished. This was partially due to the failure to operate under disturbed conditions, and partially because the technology needed to handle the handshaking and intermittent data storage requirements was not available in the 1950's.

SHAPE Technical Centre developed and tested a new meteor scatter communication system COMET (Brown and Williams, 1977), in the mid 1960's in Central Europe and in Norway. The system worked well although it had a fairly low capacity (1-4 telegraph channels). The system was tested both at 40 MHz and at 100 MHz to explore the frequency dependent properties of meteor scatter propagation. The advent of satellite communication diverted interest away from meteor scatter propagation for obvious reasons of stable propagation and large bandwidth. Since then, Meteor scatter communication systems have been constructed by the U.S. Dept. of Agriculture and a number of private corporations in the Western U.S (SNOTEL), and in Alaska (MBCS). These systems are currently in operation using fixed data rate signaling and a half duplex handshake configuration.

Recently interest in meteor scatter communication has been revitalized as a result of rapid advances in microcomputer technology and digital communication techniques. Some suggested uses for modern meteor scatter communication systems include automatic data collection from unmanned sites, thin-line digital radio links, where privacy and exclusive ownership of the communication terminals are essential. These be either fixed or mobile services, possibly incorporating some form of vehicle tracking. Meteor scatter could also be a candidate mode for operation of digital HF communication systems with extended frequency range of operation into the VHF band.

The aim of this paper is to present an overview of the meteor scatter propagation mechanism, and the ongoing research efforts which are directed towards the exploration of the potential of meteor scatter propagation as a communication channel.

1. INTRODUCTION TO METEOR SCATTER AND SYSTEM DESIGN

Normally the ionosphere is not sufficiently dense, nor are the ionization gradients large enough to efficiently reflect or scatter VHF signals back to earth. However, from time to time electron densities produced by micro meteorites entering the earthe atmosphere provide a suitable scattering mechanism. Such meteor trails are typically only a meter wide, but can have lenghts of 25 km, and more. The ionization that scatters the VHF signals is in the 85 - 120 km altitude range, and it can be used to establish communication over a range from about 300 to 2000 km. Figure 1 presents the geometry of meteor trail scattering.

To the communication system designer the meteor scatter channel is a very complex channel. The channel availability is, dependent on the transmitter power used a few percent. It is not known beforehand when the channel is open, the availability is a stochastic process which can be described statistically. The meteor scatter channel is therefore not suited for voice communication, but is inherently a digital channel. It is necessary to have available a description of the channel in order to design communication systems with predictable performances, to enable the selection of suitable types of modulation and signaling schemes. Such a description must contain as much detail as possible if the channel capacity is to be efficiently utilized. The description of the meteor scatter channel contains two main groups of information:

The Macroscopic time properties of the channel, which are descriptions of the astrophysical mechanisms behind the formation of ionized meteor trails and the statistics of trail availability. That is: when is the channel open? how long do the openings last? Are there any effects of season and time of day on the availability of the channel?

The Microscopic time properties of the channel, which are the description of the properties related to the quality of the channel such as the signal strength or path loss, the available bandwidth for signalling and the channel distortion. These are needed to select or construct the modulations and signaling schemes to be used with the channel, and to predict the performance of the communication system. The most important properties are:
Path loss; system noise and interference, time and frequency dispersion of the channel, depolarization of the radio waves by the channel.

2. METEOR TRAIL AVAILABILITY, MACROSCOPIC PROPERTIES

The availability of meteor trails show both a diurnal and a seasonal variation caused by the Earths rotation and the inclination of the Earths axis relative to the Ecliptic plane. Most meteorites orbit the sun in the same direction as the Earth, and the majority of the orbits are within $+-10^{\circ}$ of the Ecliptic plane. The escape velocity for a meteor at

Earths distance from the sun is 43 km/s. The velocity of the Earth is 29 km/s, and the gravity acceleration of a meteor moving with the Earth is 11 km/s. The resulting range of meteor velocities when entering the atmosphere is then 11 - 72 km/s. Approximately 10^{12} meteorites are swept up by the Earth every day, and a a given location the Earths rotation creates a diurnal variation of the number of meteor trails of ionization generated when the meteorites burn up in the atmosphere. There is a maximum of meteor trails at 06 hours local time when the position is near the Earths apex, and a minimum at 18 hours local time when at the Earths antapex. A seasonal variation is introduced due to the inclinatrion of the Earths axis with the Ecliptic plane. At northern latitudes, a seasonal maximum is encountered in July - August when the location is closest to the Earths apex, and similarly a minimum is found in January - February. At northern mid latitudes a seasonal variation of 1: 3-4, and a diurnal variation of 1: 5-6 is observed, and a total variation over the year can reach 1: 20. At higher latitudes a smaller diurnal variation and a smaller seasonal variation should be expected. The meteorite orbits are fairly uniformly distributed along the Earths orbit, but at certain periods the Earth orbit crosses orbits of meteor swarms,which create a substantial incresae of the amount of meteor trails for a few days. These are termed meteor showers, and the most well known shower is the Perseids which occur in the beginning of August. Figure 2 presents statistics of meteor trail returns obtained with the STC meteor scatter link in Central Europe.

3. SCATTER MECHANISMS.

Not all meteor trails within the common scattering volume of a pair of communication terminals will result in a channel opening. The meteor trail axis must be tangent to one of a family of confocal spheroids with the foci at the positions of the terminals to scatter RF energy transmitted from one terminal so part of the scattered energy reaches the other terminal. This feature implies a certain privacy in meteor scatter communication systems, as the trails which link a pair of terminals are not necessary the same trails which link other terminals together. This means that a single frequency can be used for a network with many stations, and that it is difficult to disturb a meteor scatter link with ground based transmitters which are separated some distance from the meteor scatter terminals. However, no protection is inherent in the propagation mechanism for interference from other users of the spectrum within view of one of the terminals.

Not all parts of the common scattering volume are of equal importance for the scattering of signals by meteor trails between two terminals. Two volumes termed the hot spots placed on either side of the great circle line between the terminals approximately at mid path are of special importance. Figure 3 presents the hot spots calculated for a 1000 km link. The importance of the individual hot spot varies with both the geographical positions of the terminals and with the time of day.

The scatter mechanism as described by Eshleman (1957) contain two different main modes of scattering dependent on the line electron density of the ionized meteor trail. One is called specular underdense scattering. This is modelled as scattering from the individual electrons of the meteor trail. The amplitude of the scattered wave rises sharply in approximately 100 ms as the trail is formed, and is strongest just after the formation of the trail. In course of time, the trail diffuses, and the amplitude of the signal decreases exponentially. The mean duration of the return from underdense trails is approximately half a second. The other main scattering mode is specular scattering from uverdense meteor trails. These are modelled as cylinders of ionization for which the electron density is large enough to completely reflect the wave transmitted from one terminal. The signal strength rises sharply in approximately 100ms as the trail is formed As the trail diffuses after its formation, the reflecting surface is enlarged, and the signal strength continues to increase for a while until the electron density is decreased enough not to completely reflect the wave. The signal then decreases exponentilally as the trail diffuses further. The returns from overdense meteor trails tend to last longer and provide larger signal strenghts than do the returns from underdense trails.

Other scattering mechanisms are at times available at VHF frequencies. Sporadic E layers may attain electron densities sufficiently large to provide a long lasting channel (lasting 30 minutes to hours). This channel is presumed to have properties similar to those of a HF channel. Thus they provide a long lasting communication possibility with linited bandwidth due to multipath distortion.

4. EXAMPLES OF METEOR SCATTER RETURNS FROM GREENLAND.

This section presents a selection of meteor scatter returns recorded in Greenland with the RADC High Latitude Meteor Scatter Test Bed. This instrument is a oblique meteor scatter sounding link between Sondrestrom AB and Thule AB, Greenland, a distance of 1210 km entirely within the Polar Cap. The transmitter which has an output power of 600 - 900 W is situated at Sondrestrom AB, and the receiver is situated at Thule AB. The antennas used in either end are six element, horizontally polarized Yagis. The transmitter provides periodic transmissions on the frequencies 45, 65 and 104 MHz. The frequencies are changed every 30 minutes. The receiver which is situated at Thule AB is galactically noise limited, and both the system noise and the instantaneous signal to noise ratio for the signals received from the transmitter are recorded for later analysis.

Conventionally the assessment of the communication capacity and bit error rate performance for meteor scatter system is obtained by deploying a full duplex message handling system. This method has the advantage of providing some precise communication information pertaining to that particular system, but it provides little information on

the specific propagation characteristics of the meteor scatter channel. The RADC High Latitude Meteor Scatter Test Bed provides data which can be used to predict, on the basis of the signal to noise ratio, the performance of meteor scatter communication systems using a variety of modulations and signaling schemes.

To characterize the returns, a classification system has been adopted which includes five categories: Those from underdense, overdense, and tiny meteor trails, and those from sporadic-E layers and other ionospheric scatter modes. In the data presentation that follow, the received powers are scaled in dBm in a 50 Ohm load. This means that the amplitude scales are compressed, and that an exponential change in amplitude is expressed as a straight line.

4.1 Returns from Underdense Meteor Trails.

The returns from underdense trails are characterized by a fast rising edge and a slower exponential decay. Underdense trails account for approximately 80 - 90 percent of the signals observed. Figure 4 shows a number of returns from underdense trails. The maximum amplitudes of the waveforms vary over a range of 25 dB, and the durations vary from less than 0.1 second to several seconds.

Some underdense returns exhibit fading during their exponential decay. This phenomenon usually is observed with long lasting trails, and may be attributed to wind moving portions of the trail to different positions and attitudes such that they fulfill the geometric conditions for scattering between the terminals of the link. These fades can be deep, occationally reaching down to the receiver noise level; i.e., a complete cancellation of the total received power by destructive interference between signals from the different scatter paths. The fades attributed to wind distortion of the trail differs from the amplitude oscillations seen on most of the waveforms near the point of maximum amplitude. These oscillations occur during the formation of the meteor trail, as the meteorite moves through the Fresnel zones of the scattering geometry (e.g., see Manning, 1959). Figure 5 presents examples of underdense returns with fading.

4.2 Returns from Overdense Meteor Trails.

The returns from overdense trails are also characterized by a fast rising leading edge; however, the amplitude continues to increase after the trail is fully formed, and reaches a maximum before decaying exponentially. Examples of returns from overdense trails, which constitute approximately 5 percent of the returns observed, are shown in Figure 6. The maximum amplitudes are generally larger than those from underdense trails. The returns from overdense trails can have very long durations. The longest lasting return so far observed with the Test Bed lasted in excess of 40 seconds. The majority of the returns which lasts longer than 1 second can be classified as overdense returns. As the overdense trails generally tend to last longer, they are especially prone to wind distortion. Some atypical returns from overdense trails are shown in Figure 7.

4.3 Returns from Tiny Meteor Trails.

This classification has been established to account for the multitude of very small returns from meteor trails which have a duration of less than 0.1 second and maximum signal to noise ratios of 10 to 12 dB (30 KHz receiver bandwidth). These returns do not exhibit the exponential decay associated with returns from underdense trails. Returns from tiny meteor trails account for 10 - 15 percent of the total population. Examples are shown in Figure 8. Two possible physical processes are suggested to explain the shape of these waveforms. One explanation is that the returns are caused by meteor trails diffusing so rapidly that only the part of the trail closest to the burning head of the meteorite scatters the radio waves. This is described by McKinley(1961) and Eshleman (1957) and termed "head echoes". According to this theory, the head echoes are predominantly observed at the higher frequencies. The shape of the return from a head echo is "bell like", which fits a large number of waveforms in this category. Many of the tiny returns consist of more than one bell like shape in succession. This leads to the second suggested explanation; that these waveforms originate from underdense trails whose maximum electron line densities are just large enough to produce a signal which exceeds the receivers noise floor and a signal is only received for a short while immediately after the passage of the meteorite through the principal Fresnel zone. Whichever explanation is valid cannot be determined with the present configuration of the Test Bed, as this would require simultaneous observation at different frequencies, and the measurement of the Doppler shift produced by the trail.

4.4 Returns from Sporadic E Layers.

This classification is used to account for the occurrence of very strong signals (up tp - 75 dBm), which can last from a few minutes to more than 25 minutes. Examples of such signals, which are observed occasionally at Thule AB at 45 MHz and 65 MHz are shown in Figure 9. These signals do not originate from meteor trails, nor can they originate from the ionospheres F-layer, as this does not reflect obliquely at VHF frequencies on a path as short as the Sondrestrom AB - Thule AB path. Sporadic E layers are known to have electron densities large enough to permit oblique reflections at frequencies in the lower VHF spectrum. The main characteristics of the signals, apart from their long duration, are the large amplitudes, and the slow, deep fades. Considering the signal to noise ratio alone, which for sporadic E layers can exceed 40 dB (in a 30 KHz bandwidth), a high signaling speed could be allowed as supported by the high signal - to - noise ratio in the required bandwidth. In reality, however, the usable signaling speed is most likely

limited by other propagation factors such as time decorrellation. This is presently being investigated under the AFGL program.

4.5 Unidentified Scatter Returns.

Occasionally, relatively weak but long lasting signals are received that are characterized by rapid fading as illustrated in Figure 10. These returns fit, superficially, the description of scatter from field aligned irregularities as described by Dyce (1955). However such scattering as a mode of propagation is not plausible for the Sondrestrom AB - Thule AB path which is entirely within the Auroral oval. Since the signals often precede sporadic E events, they may be reclassified as returns from weak sporadic E layers after more detailled analysis. Scattering from Field aligned irregularities is however, a well known mode of propagation for East - West paths south of the Auroral oval. It is known that the coherence bandwidth is small (few hundred Hz), and the doppler spread is large (up to 1000 Hz), but few measurements exist of these quantities. These will be needed, as the field aligned scatter mode of propagation which is incompatible with the most commonly used digital modulations can account for a large percentage of the connectivity for such paths.

5.NOISE AND INTERFERENCE IN METEOR SCATTER SYSTEMS.

5.1 Galactic Noise in Meteor Scatter Systems.

The noise in meteor scatter communication systems is primarily galactic and man-made, as receivers with low noise figure front ends (1.5 dB or less) are readily available. The receiver noise can be much less than the galactic noise at frequencies up tpo 120 MHz. The galactic noise can be calculated from existing data sources. Noise maps of the galactic sphere exist for 136 MHz and 400 MHz (Taylor 1973), and it is known that the noise power at other frequencies in the VHF spectrum can be calculated from these, as the noise power is a function of the frequency ratio to a -2.3 power.

The galactic noise power received by an antenna is proportional to the average brightness of the sky within the antenna aperture. For very large apertures covering a large part of the sky and foreground terrain, little diurnal variation will be seen in the system noise, but the diurnal variation will be several dB for small apertures. In addition, the galactic noise at different terminals within the system will not be the same at any instant of time, as their antennas point at different parts of the sky with unequal galactic noise temperatures. The impact of this is that the meteor scatter channel is not in a strict sense a reciprocal channel when the receiving systems are galactically noise limited. Such systems are operating optimally compared to systems which are not galactically noise limited, as any man made noise or interference will reduce the signal to noise ratio in the channel, which results in either decreased throughput or a higher bit error rate. The system noise for the meteor scatter receiver at Thule AB is presented in Figure 11 for the month of february 1985. The diurnal variation of the galactic noise which amounts to 4.5 dB is seen at 45 MHz and 65 MHz. 104 MHz was periodically disturbed by interference from a nearby FM transmitter.

5.2 Man-Made Noise and Interference in Meteor Scatter Systems.

Man made noise can often exceed the galactic noise by more than 10 - 20 dB, and it can be seen from Figures 10 - 14, that such levels of noise easily destructs the operation of a meteor scatter communication system, as many returns would be obscured by noise and interference. Several destinct sources of man made noise are important to consider. Interference from other users of the VHF spectrum can easily destroy the throughput of a meteor scatter system, and it follows from this that frequency agility and allocation are important issues to be addressed. Two noise sources within the meteor scatter communication system itself are of special importance. The transmitters in systems which operate in full duplex signaling modes must possess very low sideband noise in the spectrum used for reception. The sideband noise should ideally not exceed the galactic noise level. This demand is especially difficult to fulfill if the transmitter and the receiver share a common antenna through a duplex filter. The noise generated by the systems own data processor and other digital system components can exceed the galactic noise power by several tens of dB if no precautions have been taken to shield these components. Experience shows that many standard laboratory instruments and small computers which contain microprocessors are not sufficiently well shielded to be operated near a meteor scatter receiving terminal. Recently, more and more microprocessors are being incorporated in communications equipment with success, and the required shielding procedures are thus well known.

6.ANTENNAS AND THEIR POSITIONS.

As mentioned previously, certain regions of the sky, termed the hot spots, are more important than others in terms of their contributions to the overall performance of a meteor scatter communication system. For a 1000 - 1200 km link the more important regions are to the side of the great circle path between the terminals in the vicinity of the midpoint of the path. The importance of the individual hot spots has a diurnal variation as illustrated by Brown (1985), and the antennas of a meteor scatter communication system should be selected and positioned to illuminate these optimally. This calls for adaptive antennas. Such antennas are very large for frequencies in the lower VHF spectrum, they are presumably not usable for mobile terminals, and their use have not been reported yet. Most fixed meteor scatter communication systems use simple Yagi antennas with 2 to 10 elements, in some cases stacked in arrays of four antennas to further increase directivity.

6.1 Effects of the Foreground Terrain.

It is important to select the height of the antennas to illuminate the scattering volume uniformly. The effective radiation pattern of an elevated antenna is determined by the directive properties of thaeantenna it self, and by the influence of the foreground in front of it. If the foreground is perfectly flat, it can be shown that antennas should be placed 1.5 wavelenghts above the ground to obtain the best radiation pattern for a meteor scatter system with a pathlength of 1000 - 1200 km. This position will produce only one antenna pattern lobe in the elevation interval of interest, and further, if the foreground is perfectly reflecting, a desirable 6 dB enhancement of the EIRP will be obtained at an elevation angle of approximately 8 degrees. The reflections from even a very dry foreground will significantly modify the radiation pattern, although a full 6 dB enhancement of the EIRP may not be obtained. The importance of the influence of the foreground terrain can be illustrated by computing the antenna radiation pattern and the part of the path loss originating from gain of the antennas and the spreading of the propagating waves. Although these losses do not include those associated with the meteor scatter mechanism itself, it is instructive to investigate how they vary in accordance with the common volume that is illuminated by the transmitting and receiving antennas. The results of such calculations are shown in Figure 12. The computed radiation patterns for a horizontally polarized, 6 element, 104 MHz Yagi antenna mounted 10 m and 4.5 m above the ground at Thule AB are shown. The last position corresponds to a height of 1.5 wavelengths. The Figure does also present the actual radiation patterns of the two antennas as measured in situ with a helicopter. A good agreement is found between the computations and the measurement. The antenna patterns for the antennas elevated 10 m has an excessive number of nulls in the elevation range 1 - 20 degrees, the range that illuminates the parts of the sky which are most important to a meteor scatter system.

6.2 Illumination of the Scattering Volume.

Figure 13 shows the result of calculations of the path losses associated with the position of scatterers within the scattering volume for the Sondrestrom AB - Thule AB link for the antenna heights 10 m and 4.5 m. A small variation of the losses with respect to the position of the scatterers is noted when the antennas are elevated 4.5 m or 1.5 wavelength. Less than 8 dB variation is obtained over the most important portion of the scattering volume. When the antennas are 10 m or 3.3 wavelenghts above the ground, the variation in losses is considerably greater, about 20 dB, and includes a 16 db valley in the middle of the primary scattering area.

7. CHANNEL CAPACITY.

The high latitude meteor scatter test bed is the first instrument available that can deliver measurements in great detail of the time history of the meteor scatter channel. This information can be used to compute the performance of hypothetical meteor scatter communication systems using a variety of modulations and signaling schemes. The procedures used for such computations are based on the concept of the meteor scatter channel as a transfer function, and the usefulness of this approach depends upon the accuracy and the detail with which the function can be measured. Presently, only the instantaneous signal to noise ratio is measurted by the test bed, and the noise is presumed to be of galactic origin with a gaussian distribution. This does not constitute a fully valid channel transfer function, as the time and frequency dispersion expected has not been accounted for. Howevre, it has been reported by Grossi and Jahved (1977) and Weitzen et al (1984) that the meteor scatter channel does not exhibit significant channel distortion due to time and frequency dispersion. This leads to the conclusion that, to a first approximation, the properties of the meteor scatter channel can be analyzed on the basis of the signal to noise ratio. It is important that this model be validated or modified if necessary with the results of measurements of the doppler spreads and multipath properties of the channel. Only few such measurements exist, and more measurements will be needed to gain a thorough insight into these, for the meteor scatter channel, as well as for the other scatter modes of propagation observed on a meteor burst communication system.

The performance of a digital modulation system, in terms of bit-error-rate (BER) can, if the channel distortion is Gaussian distributed noise, be expressed as a function of the ratio between signal power and noise power in the channel. Various measures can be used, but the most common is the carrier-to-noise ratio (C/N), expressed as the ratio of the unmodulated carrier power to the noise power in a bandwidth equal to the bit rate. Alternatively, the baseband equivalent (E_b/E_0), defined as the ratio between the average signal energy per bit and the noise power spectral density, can be used.

The performance of a number of hypothetical meteor scatter communication systems has been computed based on data from the high latitude meteor scatter test bed. Using key numbers which link BER and signaling speed performance specifications to the equivalent C/N or E_b/N_0 requirements presented by Oetting (1979), every recorded meteor return for the month of february 1985 has been analyzed to determine the maximum signaling speed it could have supported. Distributions of this quantity are calculated to evaluate the range of possible signaling speeds which could be supported by the meteor scatter mechanism. These distributions are important in that the signaling speed is a prime factor in choosing a communications processor, especially if adaptive signaling schemes are contemplated. The capacity in bits for each individual return has been calculated from the instantaneous carrier to noise ratio, for the chosen modulations and signaling schemes. Aset of five commonly used modulations have been selected for evaluation: BPSK,

QPSK, DPSK and FSK. Coherent detection is assumed for all the chosen modulations except for FSK. Non coherent detection of FSK is a simple demodulation technique and is especially attractive for meteor scatter systems because of the relative ease of signal acquisition. However its performance relative to the C/N for the channel is not as good as the performance of other coherently detected phase modulation techniques.

The signaling schemes used in the analysis includes adaptive signaling, for which the instantaneous signaling speed is chosen to match the instantaneous C/N in the channel; and fixed signaling speeds of 5, 10 and 15 kBits/second. A BER performance of 10^{-4} is assumed in all cases. The use of a fixed signaling rate is not a very efficient mode of communication for a meteor scatter communication system, as the inherent capacity of a meteor return will not be fully utilized during the periods when the received C/N exceeds that minimally required for the specified modulation technique and signaling speed. Nevertheless a fixed signaling rate system might be preferred for many applications, based on requirements for simplicity and low cost, at the expense of optimal performance.

In the following section examples are given to illustrate some of the propagation and communication parameters that can be derived from returns received at Thule AB. The results are confined to february 1985, a period with no significant ionospheric disturbances.

7.2 Basic Return Properties.

Figure 14 presents cumulative distributions of results for the time block between 100 - 1200 UTC, as obtained from the 45, 65, and 104 MHz returns. Included are return durations, intervals between returns, maximum antenna voltages across 50 Ohms, maximum signaling speeds and the capacities in bits for the returns when using BPSK modulation with adaptive and fixed signaling rates. The figure indicates, that for this particular period, the return durations and the maximum antenna voltages decreased slightly with increasing frequency while the intervals between returns increased sharply as the frequency increased. It was predicted that the maximum signaling rate would decrease with frequency to the -0.7 power. This was based on the expectation that the maximum received signal power would decrease with frequency to the -3 power and the galactic noise power would decrease with frequency to the -2.3 power. From the measurements it was found that the galactic noise does decrease as expected, but the maximum received signal power decreased less than expexted for the February 1985 data, and the maximum signaling rate remained approximately constant with frequency, if the variation of the transmitter power with frequency is taken into consideration. This condition is only obtained with a galactically noise limited system. A system which is receiver noise limited would show a significant decrease in C/N with frequency.

It is known from earlier experiments (Weitzen, 1984; Grossi, 1977) that the coherence bandwidth for the meteor scatter channel generally exceeds 1 - 2 MHz. The February data indicate that, based simply on the signal to noise ratio, few returns will support signaling rates in excess of 1 Mbit/s. This observation supports the use of a channel model for the computation of communication capacities which only accounts for Gaussian distributed noise as the channel distortion. The computed communication capacities have for this reason been calculated with the constraint that the signaling rate cannot exceed 1 Mbit/s. The limit eshuures that no single very large return can distort a months distribution by contributing a large fraction of the total communication capacity. The communication capacity is seen to decrease with an increase in frequency. However, the decrease does not exceed a factor of two over the 45 to 104 MHz range. Based on the above observations it is concluded that the single most dominant cause of the decrease in the communication capacity of a meteor scatter link with an increase in frequency is the increase of the interval between returns.

7.3 Means of Basic Return Properties.

Figure 15 shows the mean diurnal variations for February 1985 for the following basic parameters: Return duration, maximum antenna voltage and intervals between returns. The diurnal variation in the interval between returns is seen to be approximately 1 : 2, which is less than that observed at lower latitudes (1 : 4-5 at northern mid-latitudes).

The diurnal variation in the means of the maximum signaling rate and the return capacity for BPSK modulation is presented in Figure 16 for the three operating frequencies of 45, 65 and 104 MHz. The return capacity is shown as a function of the four basic signaling schemes: Adaptive signaling, and fixed rate signaling at 5, 10 and 15 kbits/s.

7.4 Distributions of Waiting Times for Message Transfer.

Cumulative distributions of the waiting times for fixed length messages have been computed for BPSK modulation. The waiting times are computed for adaptive signaling, and for fixed rate signaling at 5, 10 and 15 kbits/s signaling rates, and based on messagec containing 50, 100, 200 and 400 8 bit characters. The computations of the waiting times includes message piecing, such that the waiting time for a message which can be contained in a single return is equal to the time elapsed since the last return was received, and the waiting time for messages which must be transferred by a number of returns is calculated as the time elapsed untill all fragments of the message have been transferred.

In message handling meteor scatter systems a certain time overhead is required for synchronization purposes at the beginning of each return. The overhead is a hardware dependent parameter, which is not included in the present computations. The results shown

are the theoretical minimum waiting times obtaineable. The overhead can be included in future computations if it is desired to evaluate the performance of a specific proposed system.

Figure 17 presents the distribution of waiting times at 45 MHz for the time block 1000 - 1200 UTC. The curves coincide for short messages, indicating that the four signaling schemes perform equally well. For the longer messages of 200 and 400 characters a difference in performance is noted. Adaptive signaling produces the shortest waiting time, while no significant variations is seen for the different fixed speeds. This trend is also observed at 65 and 104 MHz, and further the relative improvement obtained when adaptive signaling is used increases dramatically with frequency. The hardware needed for adaptive signaling is much more complex than the hardware needed for fixed rate signaling and the choice to be made between the two schemes in relation to communication system performance must be based both on the required throughput, the frequency of operation and the required waiting time. A system which is restricted to short messages, such as an environmental data collection system, will not gain any real advantage from the use of adaptive signaling unless the system must be frequency agile, or must use the higher frequencies exclusively. Systems which need to maximize the throughput or need to minimize the waiting time for long messages, should consider the use of adaptive signaling. The diurnal variations in the mean of the waiting times at 45, 65, and 104 MHz for BPSK modulation and adaptive and 5 kbits/s fixed signaling rates are given in Figure 18.

7.5 Overall Performance Data.

An overall statistical summary of some of the communication related parameters obtained from the data reduction of the data acquired with the high latitude test bed is shown in Table 1. A total of 13,789 returns were analyzed for the month of February. The table contain the total throughput in kbits for BPSK and FSK modulations for adaptive signaling and fixed rate signaling at 5, 10 and 15 kbits/s. BPSK and non coherently detected FSK have been chosen to illustrate the difference between the yield of the most efficient and the least efficient modulations analyzed. The portion of the throughput which is supported by the overdense returns and the tiny returns are presented separately to evaluate the importance of these two classes of returns. The mean capacities for the returns are also presented as are the mean throughput rates in bits/s. Finally, the relative performance of the modulation and signaling schemes have been computed. The FSK, 5 kbits/s modulation has been chosen as reference, as it has the lowest performance of the chosen modulations and signaling schemes.

The overdense returns account for a large percentage of the total throughput, ranging from 55% at 45 and 104 MHz to 68% at 65 MHz for adaptive signaling, although they only account for approximately 5% of the population of returns. The contribution of the overdense returns is less, 30% to 50%, but still very significant for the fixed signaling rate schemes. The tiny returns, however, contribute less than 1% of the throughput, indicating that this class of returns is probably insignificant in the operation of a link. This contribution will be even less in a real system, as the synchronization overhead must be subtracted from the computed throughput. The overdense returns are usually not included in present day computer models of meteor return availability and communication capacity. The large contribution of the overdense returns found in this analysis indicates that future statistical modeling of meteor scatter communication systems should include this class of returns.

The mean computed throughput exceeded 1000 bits/s at 45 MHz, decreasing to 333 bits/s at 65 MHz and to 90 bits/s at 104 MHz when ideal adaptive signaling and BPSK modulation is assumed. These results indicate that the total capacity for the month of February 1985, without correction for the variation in transmitted power vs. frequency (45 MHz: 475W, 65 MHz:780W, 104 MHz: 680W), varied as a function of frequency to the power -3. The corresponding mean throughput rates for FSK fixed signaling at 5 kbits/s range from 50 bits/s at 45 MHz to less than 4 bits/s at 104 MHz. For FSK modulation the mean throughput continues to increase as the signaling rate increases to 15 kbits/s, but an increase of the signaling rate by a factor 3 only produces a 40% increase in the mean throughput. For BPSK modulation the mean throughput is essentially independent of the fixed signaling rate. The optimum fixed signaling rate may exceed 15 kbits/s, but further analysis is needed to determine both the optimum fixed signaling rate and the corresponding waiting time distributions. The optimum fixed signaling rate for maximum throughput might not produce the shortest waiting times for short messages, as fewer returns support the high rate. This increased threshold constraint is not important for adaptive signaling, as this scheme will also utilize the weak returns by decreasing the signaling rate.

The use of ideal adaptive signaling can theoretically improve the throughput by as much as a factor of 25 relative to the mean throughput obtained when fixed signaling rates are assumed. This factor, however, is too optimistic, as the synchronization overhead for an actual system is not included. This overhead generally occurs near the leading edge of the returns, where the signal strenghts and thus the signaling speed is largest. The data from the RADC High Latitude Meteor Scatter Test Bed can be used to predict the performance of proposed meteor scatter communication systems once the hardware dependent parameters such as modulation and synchronization time is known.

8 COMPUTER MODELS FOR SYSTEM DESIGN

Although, many valuable results on the properties of the meteor scatter channel are obtained from previous and existing meteor scatter communication systems and test beds,

the planner of future meteor scatter links will need validated computer models of the meteor scatter channel to be able to predict the performance of any wanted meteor scatter link regardless of its geographical position or path length. However, measurements will still be needed to generate data which can be used to validate the computer models. Such models of various sophistication have recently been reported. A very involved model is presented by Brown (1985). The model can compute various properties of a meteor scatter link from the positions of meteor radiants and the statistics of variation of the number of meteorites swept up by the earth. This information, together with the position of the link terminals, the properties of the link hardware and the antennas, their position and the properties of the antenna foreground is used to compute the number of meteor returns and their duration, and the positions and variation of the hot spots. From this information, the optimum antenna patterns can be computed as a function of the time of day, The effects of terrain blockage, as well as the communication related parameters such as message waiting time and throughput. The model can be used to evaluate the transfer efficiency of different message packet schemes at fixed signaling rates, an issue of importance to system designers.

The model is limited to underdense returns, but is currently being refined to include both overdense returns and certain scatter modes of propagaton. An example of the computation of a hot spot diagram, and a comparison of the calculated and measured duty cycle for the STC meteor scatter link operated in Central Europe in the 1960's are shown in Figure 19.

9EFFECTS OF NATURALLY OCCURRING DISTURBANCES.

Two types of naturally occurring absorption is frequently encountered at high latitudes: Auroral absorption, which is confined mainly to paths penetrating the upper D- and lower E-regions of the ionosphere within the auroral oval; and Polar Cap Absorption (PCA or SPE - Solar Protron Event) which effects paths penetrating the D-region inside the polar cap. Of the two the SPE's are generally much more severe and longer lasting. They occur following certain solar flares which dump energetic particles (protrons) into the lower iinosphere within the polar cap. The effects of such absorptions are known to seriously degrade or cut off HF communications in the polar regions for many days.

Knowledge of the morphology of high latitude absorption events has increased greatly in resent years as the result of a wide variety of experiments conducted both on ground and in space (Kossey et al, 1983; Reagan et al, 1982). The possible mitigation of SPE absorption effects by increasing the frequency of operation of a meteor scatter communication system is illustrated in Figure 20. For simplicity the figure construction assumes that the wave frequency is much greater than the collision frequency in the absorbing region of the ionosphere. Theoretical results shown are for a 1000 km link, a variety of ionospheric conditions, rangimng from undisturbed (0 dB riommeter absorption) to severely distrubed conditions (greater than 10 dB riometer absorption) and the absorption is assumed to be inversely proportional to the square of the frequency. This assumption can be supported for elevation angles exceeding 15 deg. by means of raytracing with a computer program. The results clearly indicate how even a modest SPE can strongly affect a meteor scatter system operating in the lower VHF band, and how absorption effects can be reduced significantly by operating at higher frequencies.

The advantage of using higher frequencies to mitigate absorption is offset by a decrease in the capacity of the system, due to the decrease of the throughput of the system as frequency increases. Galactic noise also suffers absorption under a SPE event, but that loss is much less severe, as the noise traverses the absorbing region only once, whereas the the signal must go through the region twice to reach the receiver. The prediction of the overall capacity for a meteor scatter communication system as a function of frequency during a SPE event is a question of considerable debate, which can be resolved by further measurements of system performance during SPE's supported by modeling efforts.

9.1 The Height of the Scattering Region.

The scattering of radio waves by meteor trails occurs at an altitude of 85 - 120 km. The higher the frequency, the lower the predominant trail height (McKinley, 1961). This brings up the question whether it is possible to chose a frequency of operation, for which the radio waves do not have to pass through the entire absorbing region of the lower ionosphere, before it is scattered from a meteor trail. To investigate this question, an electron density profile from a 10 dB riometer absorption SPE event was used to calculate the one way absorption as a function of height with a raytracing program. The results for 30 MHz at 90^{o} elevation, and for 104 MHz at 0^{o}, 30^{o} and 90^{o} elevation are shown in Figure 21. It can be seen that more than 95% of the total absorption in dB occurs below 85 km. Thus, there is no advantage in selecting a higher frequency of operation to obtain a lower scattering height.

9.2 Polarization

The ionosphere is an anisotropic medium and is capable of depolarizing radio waves known as Faraday rotation. This depolarization is not of concern at VHF frequencies during undisturbed ionospheric conditions. It has been shown by Nes (1985), that the polarization of meteor scatter returns generally do not deviate more than +- 20^{o} from the transmitted polarization. However, the electron density in the lower ionosphere can increase orders of magnitude during SPE events. This will create a frequency dependent Faraday rotation which might result in severe polarization mismatches at the receiving antennas of a meteor scatter link, if no precautions have been taken to avoid this. The

approximate magnitude of the Faraday rotation and its effect has be computed by Cannon (1985) for a linearly polarized receiving antenna system. Polarization losses of 3 - 6 dB was found at 40 MHz during a 1.5 dB SPE event. The Faraday rotation may therefore be another factor of concern for the design og meteor scatter communication systems for high latitudes.

ACKNOWLEDGEMENTS

The readers will notice that this presentation leans heavily on the paper referenced as 96 in the bibliography. The author is deeply indepted to his coauthors of that paper mr. J.E. Rasmussen, Mr. M.J. Sowa, Mr. J.M. Quinn, Dr. P.A. Kossey for the opportunities given to him to participate in the work with the RADC High Latitude Meteor Scatter Test Bed. Also the author is indepted to Dr. E.A. Lewis for patient and outstanding teaching, to Dr. Per Enge, Mr. R. Bono, Dr. S. Parl, Dr. A. Malaga and Dr. J. Weitzen for valuable discussions.

BIBLIOGRAPHY.

INTRODUCTION TO METEOR SCATTER.

1. McKinley D.W.R: Meteor Science and Engineering, McGraw-Hill 1961.
2.Sites F.J.:Communication via meteor trails. AGARD 23rd. EWPP Symposium 3-7 oct. 1977
3.Oetting J.D.: An analysis of meteor burst communication for military applications. IEEE Trans. Com; Com-28 no.9. Sept. 1980.
4.Grossi M.D, A. Jahved: Time and frequency spread in meteor burst propagation paths. AGARD,23rd. EWPP Symposium 3-7 oct. 1977.
5. Proc. IRE. Dec. 1957. This volume contain 12 papers on all aspects of meteor scatter communication. All papers are referenced below.

METEOR SCATTER IN GENERAL.

6.Forsyth P.A., E.L. Vogan, D.R. Hansen, and C.O. Hines: The principles of JANET, a meteor burst communication system. Proc. IRE Dec. 1957.
7. Campbell L.L., C.O. Hines: Bandwidth considerations in a JANET system. Proc IRE. Dec. 1957.
8. L.L. Campbell: Storage capacity in burst type communication systems. Proc. IRE. Dec 1957.
9. Davis G.W.L., S.J. Gladys, G.R.Lang, L.M.Luke, M.K.Taylor: The Canadian JANET system. Proc. IRE. Dec 1957.
10. Montgomery G.F.: Intermittent communication with a fluctuating signal. Proc. IRE. Dec. 1957.
11. Montgomery G.F., G.R. Sugar: The ability of meteor bursts for intermittent radio communication. Proc. IRE. Dec. 1957.
12. Vincent W.R., R.T. Wolfram, B.M. Sifford, W.E.Jayne, A.M. Peterson: A meteor burst system for extended range VHF communication. Proc. IRE. Dec. 1957.
13. Vincent W.R., R.T. Wolfram, B.M. Sifford, W.E. Jayne, A.M. Peterson: Analysis of oblique path meteor burst propagation data from the communications viewpoint. Proc. IRE. Dec.1957.
14. Rach R.A.: An investigation of storage capacity required for a meteor burst communication system. Proc. IRE. Dec. 1957.
15. Eshleman V.R.: On the wavelength dependence of the information capacity of meteor burst propagation. Proc. IRE. Dec. 1957.
16. Eshleman V.R., R.F. Mlodnosky: Directional characteristics of metepr propagation derived from radar measurements. Proc. IRE. Dec. 1957.
17. Meeks J.C., M.L. James: On the influence of meteor radiant distribution in meteor scatter communication. Proc. IRE. Dec. 1957.
18. Forsyth P.A., E.L. Vogan: Forward scattering of radiowaves by meteor trails. Can. J. Phys. V33. 1955.
19. Hines C.O.: Diurnal variation in the number of shover meteors detected by the forward scattering of radio waves. I. Theory. Can. J. Phys. V33. 1955.
20. Forsyth P.A., C.O. Hines, E.L.Vogan: Diurnal variation in the number of shower meteors detected by the forward scattering of radio waves. II: Experiment. Can. J. Phys. V33. 1955.
21. Forsyth P.A., Vogan: The duration of forward scattered signals from meteor trails. Can. J. Phys. V34. 1956.
22. Pugh R.E.: The number density of meteor trails observable by the forward scattering of radio waves. Can. J. Phys. V34. 1956.
23. C.O. Hines, R.E. Pugh: The spatial distribution of signal sources in meteoric forward scattering. Can. J. Phys. V34. 1956.
24. C.O. Hines: Diurnal variation in the number of shower meteors detected by the forward scattering of radio waves. III. Ellipsoidal theory. Can. J. Phys. V36. 1958.
25. Forsyth P.A.: The forward scattered signal from an overdense meteor trail. Can. J. Phys. V36. 1958.
26. Hines C.O.: Diurnal variation in forward scattered meteor signals. JATP. V9. N4. 1956.
27. Hines C.O.: The efficiency of meteor scattering over short and long ranges. Can. DRTE, Rep. 12-2-15, Aug. 1955.
28. Campbell L.L.: Information loss and delays in two proposed techniques for meteor communication. Can DRTE, rep. 12-3-4, nov. 1953.
29. Vogan E.L., P.A. Forsyth: Privacy in system JANET. Can DRTE, Rep.12-3-4 , Nov. 1953.
30. Hines C.O.: Information transfer in JANET systems of fixed and of variable bandwidths. Can DRTE, Rep. 12-3-14. July 1955.

31. Hansen D.R., L.L. Campbell: A proposed variable information rate JANET system. Can. DRTE, July 1955.
32. Shockley T.D., W.B. Jones: Effects of transmission path parameters on meteor trail propagation. US. Navy, NONR-991(02), Tech. rep. No.15.
33. Spezio A.E.: Meteor burst communication systems: Analysis and synthesis. NRL. Washington DC. Dec. 1978.
34. James J.C., M.L. Meeks: On the relative contribution of various sky regions to meteor trail communication. ONR. NONR-991-02, june 1956, tech. rep. No.1.
35. Nupen: Bibliography on meteoric radiowave propagation. NBS. Boulder, tech. note 94, may 1961.
36. Meeks M.L., J.C. James: Meteor radiant distributions and the radio echos observed by forward scatter. JATP. V16. 1959.
37. Meteor Comm. Consultants: Analysis of meteor burst communications for Navy strategic applications. NOSC, N66001-79-c-0460. Feb 1980.
38. Eshleman V.R.: Meteors and radio propagation. Part A. Meteor ionozation trails: Their formation and radio echoing properties. RPL. Stanford Univ. Feb. 1955.
39. Eshleman V.R.: Meteor scatter. RPL Stanford Univ. Sci. Rep. 4, Aug. 1958.
40. Heritage J.L., J.E. Bickel, C.P.Kugel: Meteor communication in Minimum Essential Emergency Communication Network (MEECN). DCA, NOSC, Tech. Rep. 138, Aug 1977.
41. Brown D.W., H.P.Williams: The performance of meteor burst communications at different frequencies. AGARD CP-244, Oct. 1977.
42. Bartholome' P.J., I.M. Vogt: Comet- a new meteor burst system incorporating ARQ and diversity reception. IEEE. Trans. Com-16, no.2, April 1968.
43. Sugar G.R.: Radio propagation by reflection from meteor trails. Proc. IEEE, Feb 1964.
44. Manning L.A., V.R. Eshleman: Meteors in the ionosphere. Proc. IRE, Feb. 1959.
45. Akram F, N.M.Sheikh, A. Jahved, M.D. Grossi: Impulse response of a meteor burst communication channel determined by raytracing. IEEE Trans. Com. April 1977.
46. Santeford H.S.: Meteor burst communication system. Alaska winther field test program - Test report, Boeing Aerospace Company Document No. D182-10423-1.
47. Ince A.N.: Spatial properties of meteor burst propagation. IEEE Trans. Com, Com-28, no.6, June 1980.

IONOSPHERIC ABSORPTION.

48. Lawrence R.S., C.G. Little, H.J.A. Chivers: A survey of ionospheric effects upon earth-space radio propagation. Proc. IEEE, Jan. 1964.
49. Hunsucker R.D.: Survey of polar and auroral region effects on HF propagation. Radio Sci. Vol.4, No.4, April 1969.
50. James J.C.: The influence of meteor radiant distributions on radio echo rates. ONR. NONR. 991(02). Tech. Rep. 3, May 1958.
51. Hunsucker R.D.: simultaneous riometer and incoherent scatter radar observations of the auroral D-region. Radio Sci. Vol.9, No.2, Feb 1974.
52. Crysdale J.H.: Analysis of the performance of the Edmonton-Yellowknife JANET circuit. IRE Trans. Com. March 1960.
53. Maynard L.A.: Propagation of meteor burst signals during the polar disturbance of november 12-16 1960. Can. J. Phys. V39, p628, 1961.
54. Cormier R.J.: Thule riometer absorptions of polar cap absorption events (1962-1972). AFCRL TR-73-70060. Jan.1973.
55. Cormier R.J.: Polar riometer observations. AFCRL TR-70-0690. Dec. 1970.
56. Cormier R.J.: Riometry as an aid to ionospheric forecasting. AFCRL 70-0689. Dec. 1970.
57. Kossey P.A., J.P. Turttle, R.P. Pagliarulo, W.I.Klemetti, J.E. Rasmussen: VLF reflection properties of the normal and disturbed polar ionosphere in northern Greenland. Radio Sci., Vol 18, No.6, p 907-916, Nov-Dec 1983.

AURORA AND SCATTER.

58. Greely A.W.: Account of auroral displays accompanying the great magnetic storm of nov. 15-19 1882, noted at Fort Conger, Grinnell Land. Three years of arctic service. 1884.
59. Barry G.H.: HF-VHF communications experiment using field aligned ionospheric scatterers. Radio Sci. Vol.9, No.11, Nov. 1974.
60. Stathacopulos A.D., G.H. Berry: Geometric considerations in the design of communications circuits using field aligned scatter. Radio Sci. Vol.9 No.11, Nov. 1974.
61. Heritage J.L., S.Weisbrod, W.J.Fay: Experimental studies of meteor echoes at 200 MHz. IRE Trans. A.P. Jan 1960
62. Heritage J.L., S. Weisbrod, W.J. Fay: Evidence for a 200 megacycles per second ionospheric forward scatter mode associated with the earths magnetic field. JGR Vol.64 No.9, Sept. 1959.
63. Philipp N.D., E.Y. Gliebman: Power of HE-scatter signals. Geom, Aeron.,1976
64. Hargreaves J.K.: Auroral absorption of HF radiowaves in the ionosphere: A review of the first decade of riometry. Proc. IEEE Vol.57 No.8, Aug 1969.
65. Dyce R.: VHF auroral and sporadic E propagation form Cedar Rapids, Iowa to Ithaca, New York. IRE Trans. PGAP. April 1955.
66. Booker H.G.: A theory of scattering by non-isotropic irregularities with application to radar reflections from the aurora. JATP Vol.8 p204, 1956.
67. Akasofu S.I.: Physics of magnetospheric substorms. Astrophysics and space science library, Vol.47. D. Reidl. Publ.
68. Ince A.N., I.M. Vogt, H.P. Williams: A review of scatter communications. AGARD. CP-244.

69. Sinno K.: On the time delay of the appearance of sporadic E following meteor activity. JATP. VOL.48 p35, 1980.
70. Nupen: Bibloigraphy on auroral radiowave propagation. NBS. Boulder. Tech. Note 128, 1962.
71. Unwin R.S., F.B. Knox: The morphology of the VHF radio aurora at sunspot maximum IV. JATP, Vol.30 p25, 1968.
72. Elkins T.J.: A model for high frequency radar auroral clutter. RADC-TR-80-122, march 1980.
73. Avery S.K., A.C.Riddle, B.B. Balsey: The Poker Flat MST radar as a meteor radar. Radio Sci. Vol.18, No.6, Nov-Dec 1983.
74. Lewis E.W.: Numerical search for Sondrestrom-Thule VHF propagation paths via specularly reflecting field-aligned ionization columns. Megapulse Inc. RADC-TR Interim Technical Report, May 1984.

SYSTEM NOISE.

75. Taylor R.E.: 136 MHz/ 400 MHz radio-sky maps. Proc.IEEE, Apr.1973.
76. Cottony H.V., J.R. Johler: Cosmic radio noise intensities in the VHF band. Proc. IRE. Sept. 1952.
77. Brindle C., D.K. French, J.L. Osborne: An interpretation of the galactic continuum radiation II. A three dimensional model for the radioemissivity. Mon. Not. R. Astr. Soc. p184, 283 (1978).
78. Cane H.V.: A 30 MHz map of the whole sky. Aust. J. Phys. 1978,31,561.
79. Lewis E.A.: CRC Handbook of atmospherics, p276.
80. Hsieh H.C.: Characteristics of ionospheric thermal radiation. JATP Vol.28. p738, 1966.
81. Hsieh H.C.: A theory of ionospheric thermal radiation. JATP Vol.28, p769, 1966.
82. Hall M.R.M.: Effects of the troposphere on radio communication. chap.4, 1979.

COMMUNICATION THEORY AND TECHNOLOGY.

83. Oetting J.D.: A comparison of modulation techniques for digital radio. IEEE Trans. Com. Com-27. No12, Dec 1979.
84. Amoroso F.: The bandwidth of digital signals. IEEE communications Magazine, Nov. 1980.
85. Horstock N,: The flexible digital receiver. ElektronikCentralen, Denmark. ECR-120, June 1982.
86. Thompson R., D.R. Clouting: Digital angle modulation. Wireless World. Dec 1976, Feb 1977.
87. Monsen, P.: Fading channel communications. IEEE Comm. Soc. Mag. Vol.18 No.1 Jan 1980.
88. Ostergaard, J.C.: Meteor Burst Receiver manual. Darcom Electronics, Denmark. Feb 1984.
89. Weitzen, J.A., W.P. Birkemaier, M.D. Grossi: A high speed digital modem for the meteor scatter channel, Proceedings of Seventeenth Annual Conference on Information Science and Systems, Johns Hopkins University, March 23-25, 1983.
90. Albright, Latorre: Some applications of the meteor burst communications mode. The Boeing Company, Seattle, Washington.
91. Nes H.L.: Propagation characteristics of meteor burst communication systems. SHAPE Technical Centre, Tech, Memo STC-TM-710. July 1983.
92. Nes H.L., F.A.Eenhorn, W.K. Sekreve: Test equipment for meteor trail propagation studies. SHAPE Technical Centre. STC-IN-534. May 1983.
93. Brown. D.W. A physical meteor-burst propagation model and some significant results for communication system design, IEEE Journal on selected areas in communications, Vol SAC-3, No. 5, Sept. 1985.
94. Weitzen J.A., W.P. Birkemeier, M.D. Grossi: An estimate of the capacity of the meteor burst channel. IEEE Trans. Com. Com-32, No. 8, Aug. 1984.
95. Weitzen J.A., M.D. Grossi, W.P. Birkemaier: High resolution multipath measurements of the meteor scatter channel. Radio Sci., Vol 19, No.1, p375, Jan-Feb 1984.
96. Ostergaard J.C., J.E. Rasmussen, M.J. Sowa, J.M. Quinn, P.A. Kossey: Characteristics of high latitude meteor scatter propagation parameters over the 45 - 104 MHz band. AGARD CP-382, 1985.
97 Nes H.: Meteor burst polarization trials; Electronics Letters Vol 21, No. 24, Nov 1985
98 Cannon P.S., A.H. Dickson, M.H. Armstrong: Meteor Scatter Communication at High Latitudes. AGARD, CP-382, 1985.

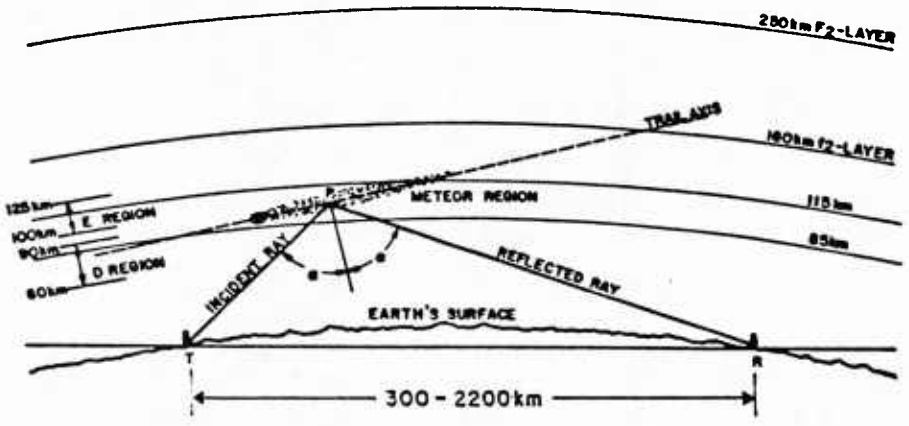


Fig. 1. Geometry of Meteor Scatter Propagation.

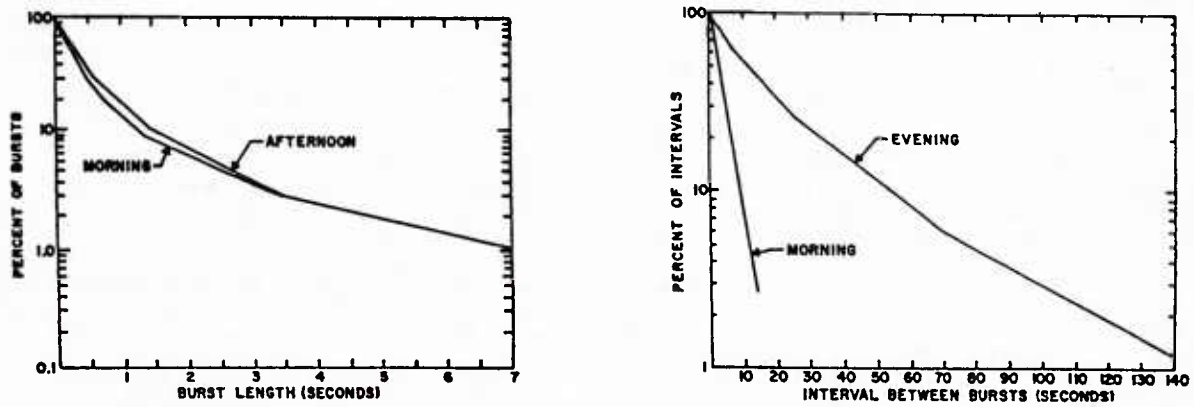


Fig.2. Mid latitude statistics for 40 MHz meteor trail returns. (a) Probability distribution of return length. (b) Probability distribution of return intervals. (Brown and Williams, 1977)

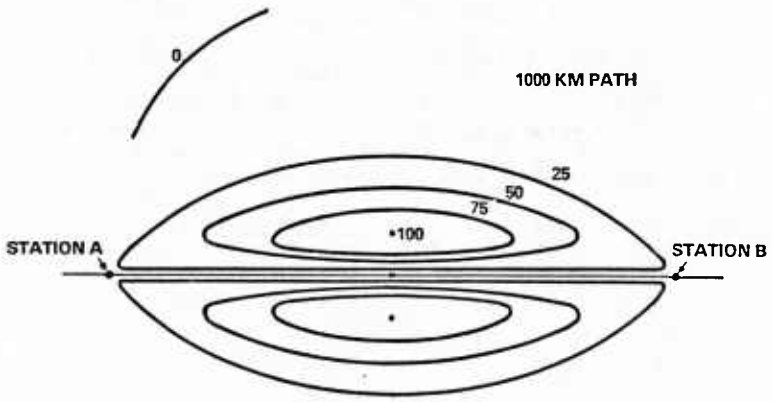


Fig. 3. Relative signal contributions from various parts of the meteor region computed for a 1000 km. path (Ince, 1980)

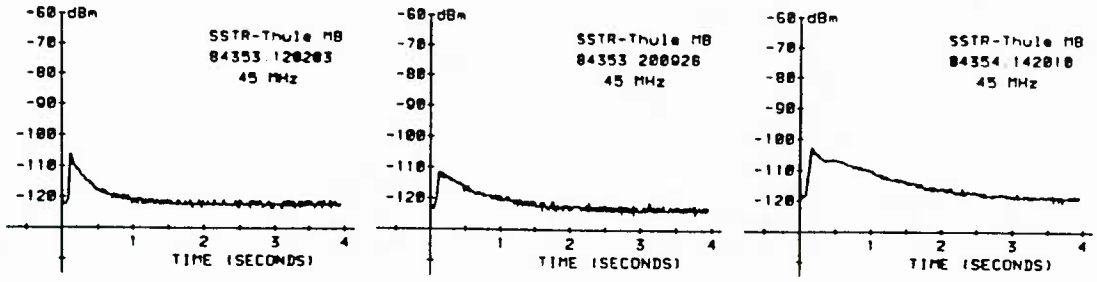


Fig. 4. Examples of returns from underdense trails.

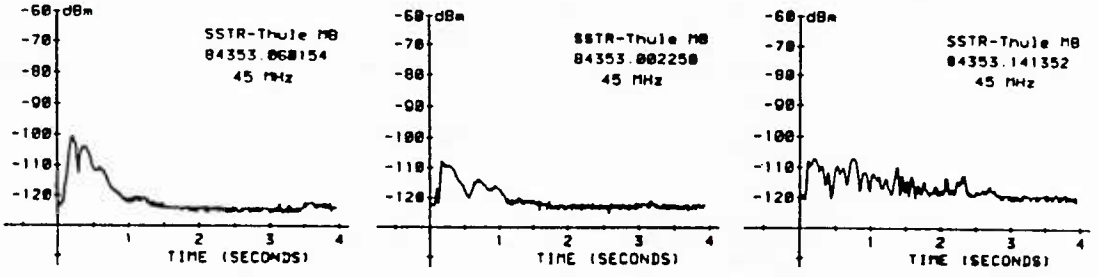


Fig. 5. Examples of returns from underdense trails with severe "wind distortion".

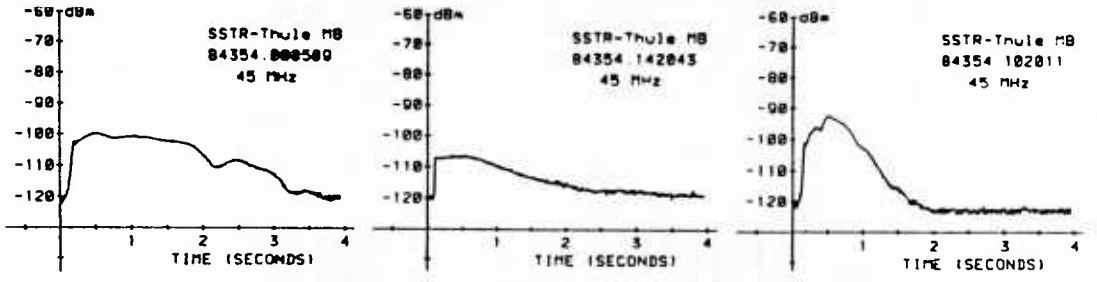


Fig.6. Examples of returns from overdense meteor trails.

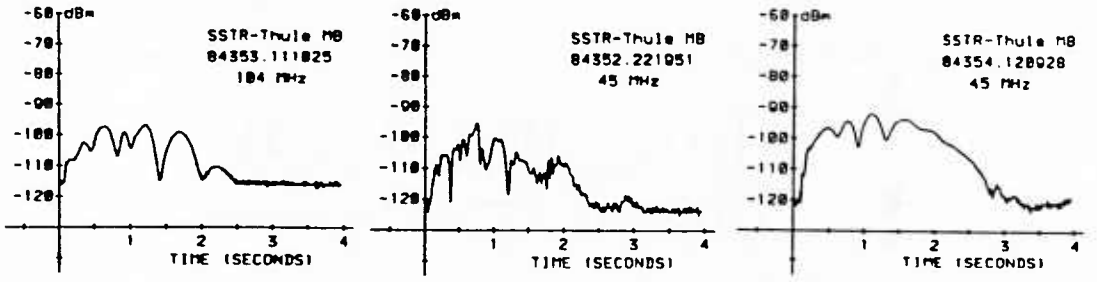


Fig. 7. Examples of returns from atypical overdense meteor trails.

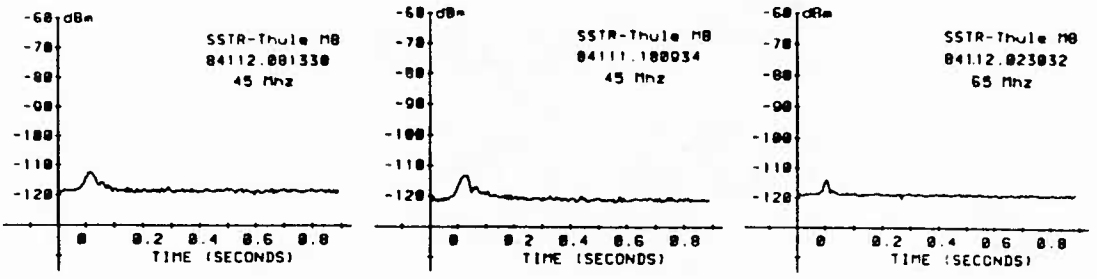


Fig. 8. Examples of returns from "Tiny" meteor trails.

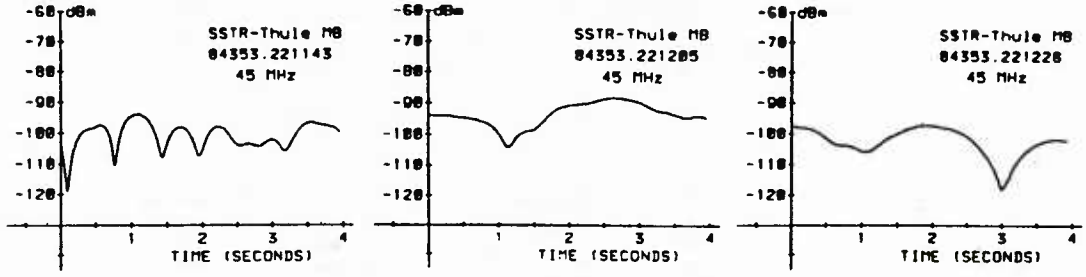


Fig.9. Examples of a series of signals reflected from sporadic E layers.

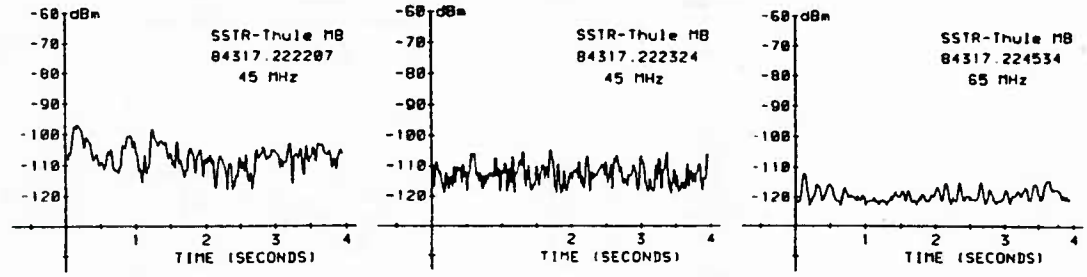


Fig. 10. Examples of a series of waveforms classified as unidentified scatter.

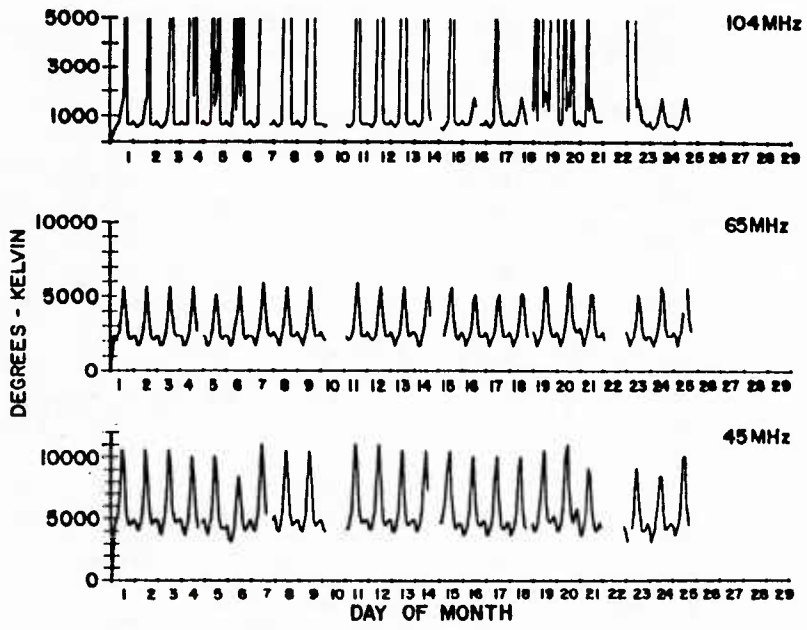


Fig. 11. System noise for the Thule AB meteor scatter receiver for February 1985. The diurnal variation of the galactic noise is seen at 45 and 65 MHz. 104 MHz was periodically disturbed by man made interference.

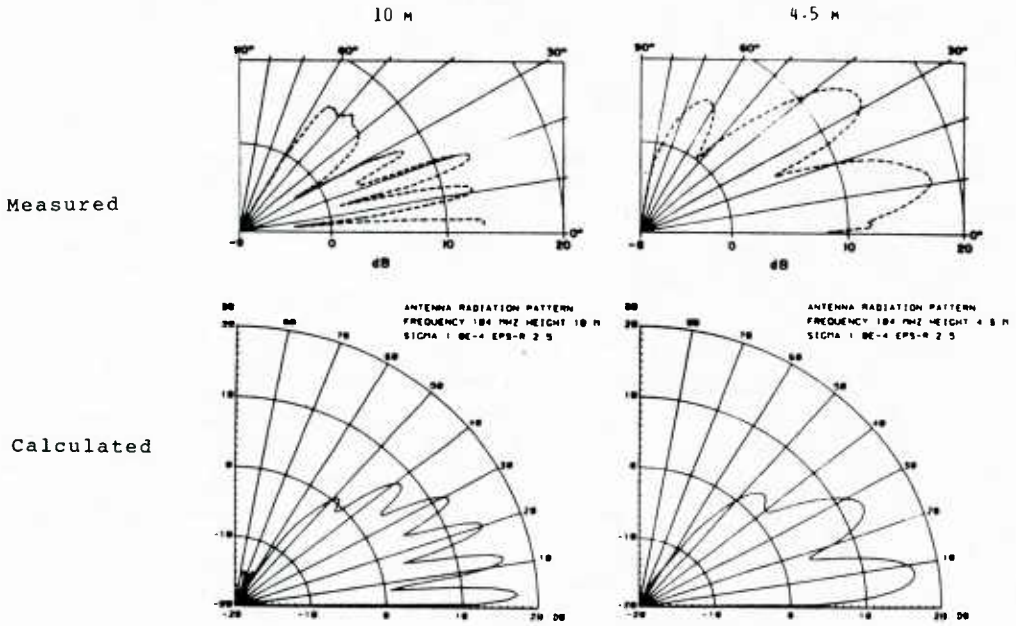


Fig.12 Measured and calculated antenna patterns for receiving antennas at 104 MHz. The antennas was mounted 4.5 m and 10 m above the ground.

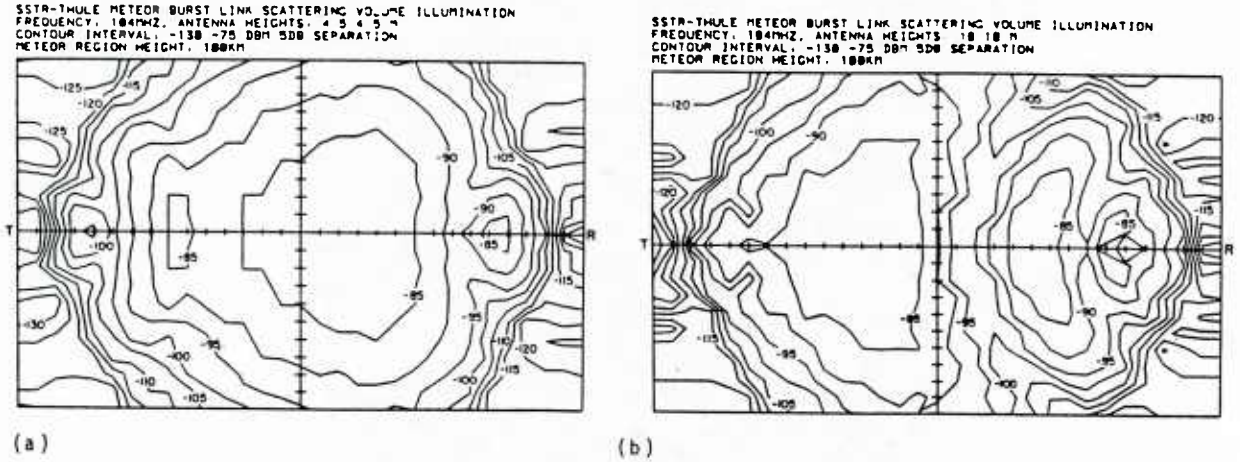


Fig. 13. Calculated illumination of the common scattering volume at 100 km for 104 MHz on the Sondrestom AB - Thule AB path. (a) The antennas are mounted 1.5 λ or 4.5 m above the ground. (b) The antennas are mounted 3.3 λ or 10 m above the ground.

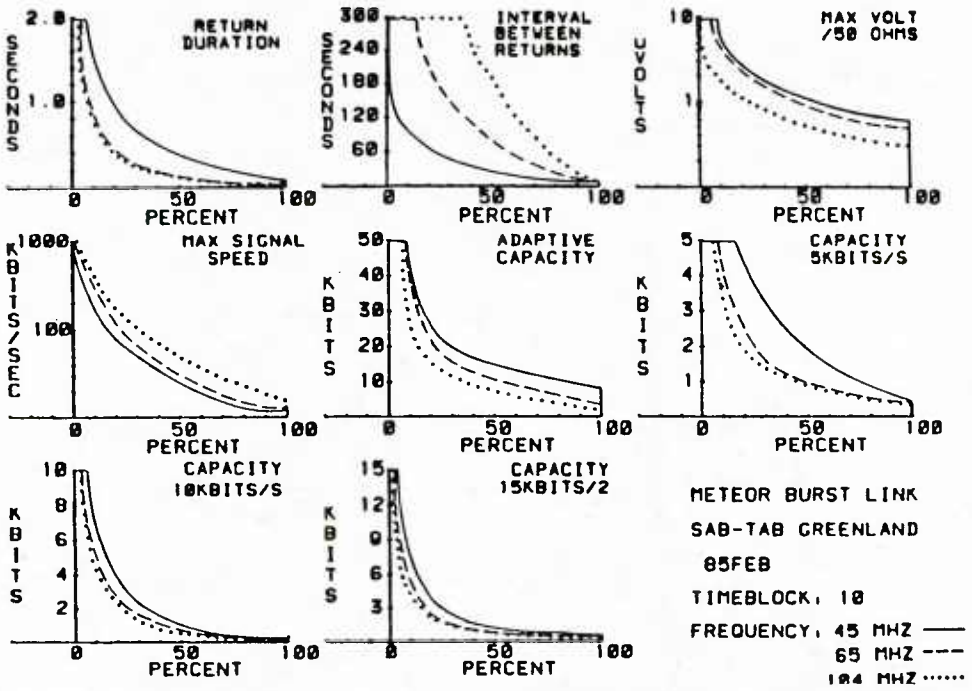


Fig. 14. Cumulative distributions of return duration, interval between returns, maximum antenna voltage across 50 Ohms, maximum signaling speed and the capacity in bits for BPSK modulation and adaptive and fixed signaling rates.

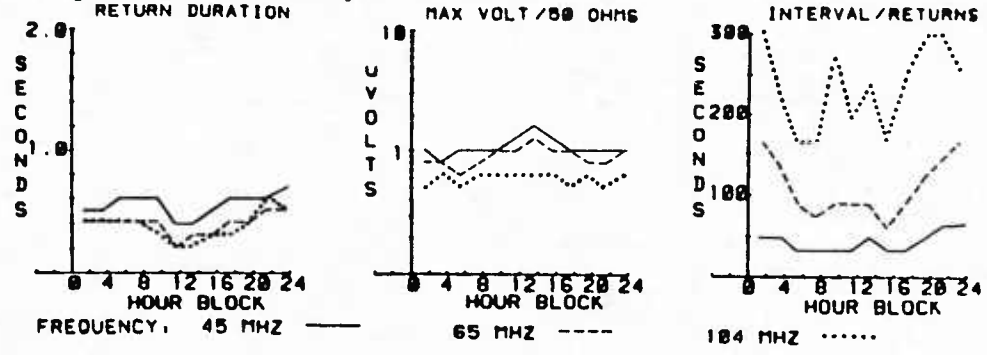


Fig. 15 Mean diurnal variation of return duration, maximum antenna voltage and interval between returns for February 1985.

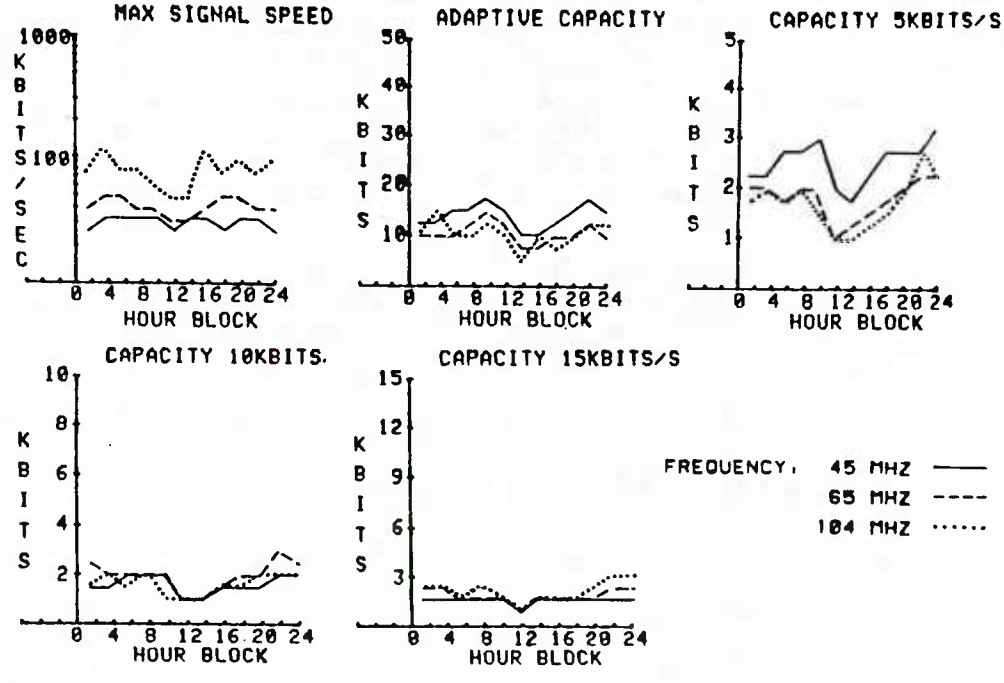


Fig. 16. Diurnal variation in the mean of the maximum signaling rates and return capacities at 45, 65 and 104 MHz during February 1985. BPSK modulation is assumed. The capacities are calculated for adaptive and fixed rate signaling schemes.

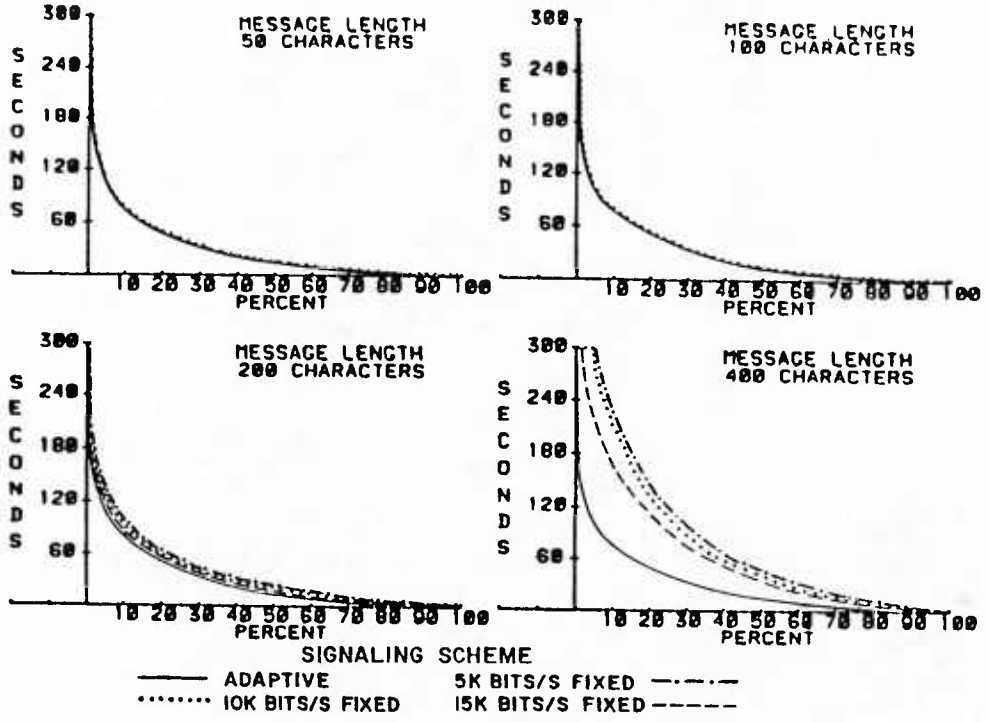


Fig. 17. Cumulative distributions of waiting times at 45 MHz for the 1000-1200 UTC time block in February 1985. The calculations are for messages lengths of 50, 100, 200 and 400 8 bit characters. The distributions are calculated for BPSK modulation and adaptive and fixed signaling rate schemes.

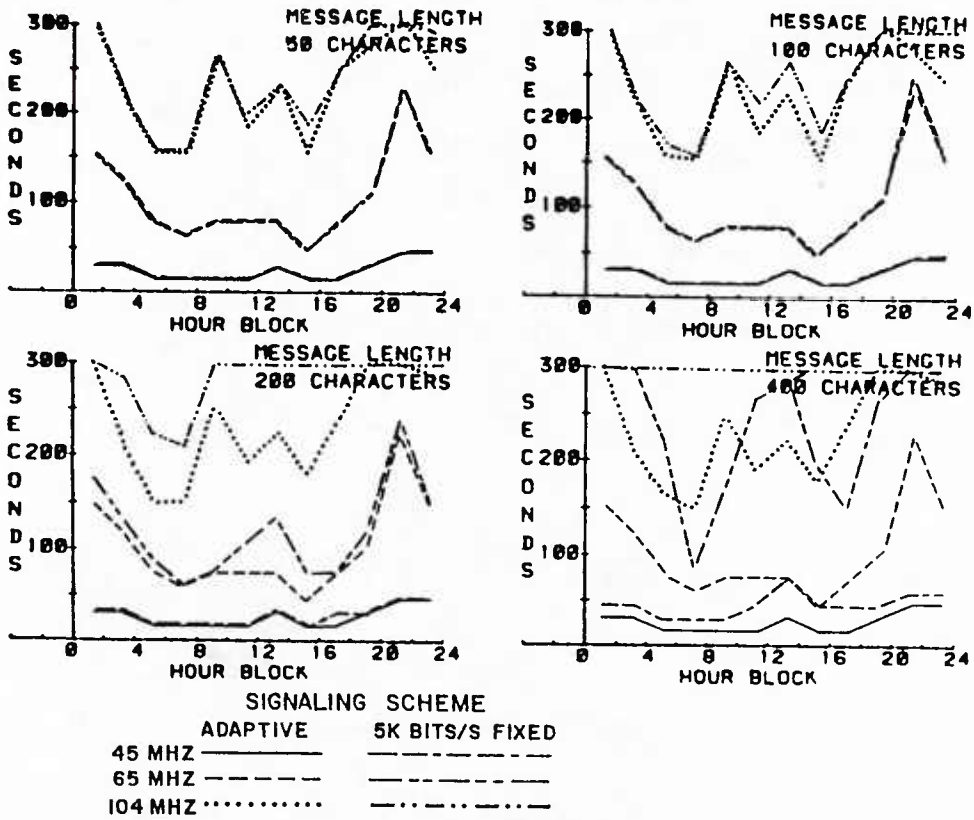


Fig. 18. Diurnal variation in the mean of the waiting times at 45, 65 and 104 MHz for February 1985. The values for adaptive and fixed rate signaling schemes with BPSK modulation are presented.

	ADAPTIVE SIGNALING		FIXED SIGNALING RATES (kbits/s)					
			BPSK			FSK		
	BPSK	FSK	5	10	15	5	10	15
CAPACITY (kbits)								
ALL RETURNS	2540000	1620000	225000	214000	234000	122000	142000	163000
OVERDENSE RETURNS	1400000	949000	70200	89700	107000	48700	66100	81700
%-OVERDENSE RETURNS	55.1	58.4	31.1	41.8	45.5	40.0	46.6	50.0
TINY RETURNS	31500	18100	2760	1090	998	715	605	680
%-TINY RETURNS	1.2	1.1	1.2	0.5	0.4	0.5	0.4	0.4
MEAN CAPACITY/RETURN (kbits)	44	28	3.9	3.7	4.1	2.1	2.5	2.8
MEAN THROUGHPUT (bits/s)	1050	670	93	89	97	50	59	67
REL. PERFORMANCE	20.9	13.3	1.8	1.7	1.9	1.0	1.1	1.3

a. Frequency = 45 MHz, Transmitted power = 475 w.

	ADAPTIVE SIGNALING		FIXED SIGNALING RATES (kbits/s)					
			BPSK			FSK		
	BPSK	FSK	5	10	15	5	10	15
CAPACITY (kbits)								
ALL RETURNS	807000	537000	52200	61700	65900	34200	39900	46600
OVERDENSE RETURNS	549000	389000	18500	26700	32800	14200	20400	25400
%-OVERDENSE RETURNS	68.1	72.3	35.3	43.3	49.7	41.4	51.1	54.6
TINY RETURNS	3710	2140	351	265	134	164	66	60
%-TINY RETURNS	0.4	0.3	0.6	0.4	0.2	0.4	0.1	0.1
MEAN CAPACITY/RETURN (kbits)	45	30	2.9	3.4	3.7	1.9	2.2	2.6
MEAN THROUGHPUT (bits/s)	333	222	22	26	27	14	17	19
REL. PERFORMANCE	23.5	15.7	1.5	1.8	1.9	1.0	1.1	1.3

b. Frequency = 65 MHz, Transmitted power = 780 w.

	ADAPTIVE SIGNALING		FIXED SIGNALING RATES (kbits/s)					
			BPSK			FSK		
	BPSK	FSK	5	10	15	5	10	15
CAPACITY (kbits)								
ALL RETURNS	217000	145000	13400	15200	17800	8310	10700	12200
OVERDENSE RETURNS	117000	82400	3510	5220	6830	2750	4310	5630
%-OVERDENSE RETURNS	54.0	56.8	26.2	34.3	38.3	33.1	40.3	46.1
TINY RETURNS	1000	577	165	129	83	72	40	26
%-TINY RETURNS	0.4	0.3	1.2	0.8	0.4	0.8	0.3	0.2
MEAN CAPACITY/RETURN (kbits)	35	23	2.1	2.4	2.9	1.3	1.7	2.0
MEAN THROUGHPUT (bits/s)	90	60	5.5	6.3	7.4	3.4	4.4	5.0
REL. PERFORMANCE	26.0	17.4	1.6	1.8	2.1	1.0	1.2	1.4

c. Frequency = 104 MHz, Transmitted power = 630 w.

Table 1. Summary of statistics during February 1985 assuming BPSK and FSK modulation and adaptive and fixed signaling schemes.

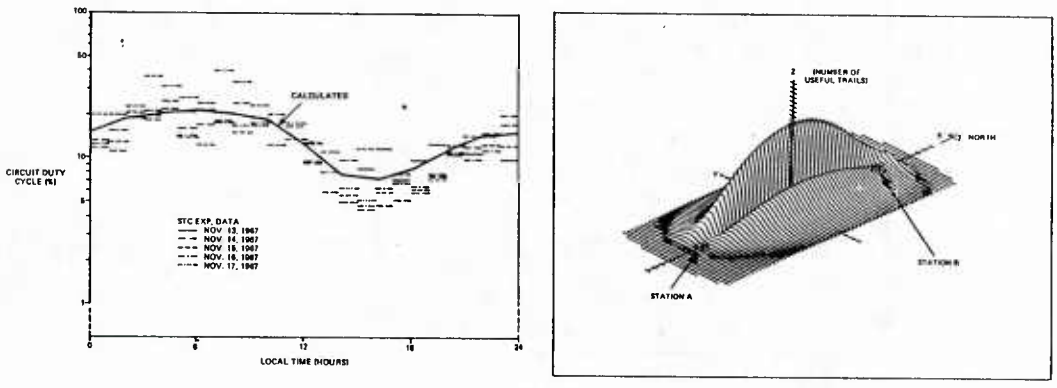


Fig. 19. Comparison of the computed and measured duty cycle for the STC meteor scatter link in Central Europe, and a presentation of a computed hot spot diagram. The results have been obtained with the meteor scatter model presented by Brown, 1985.

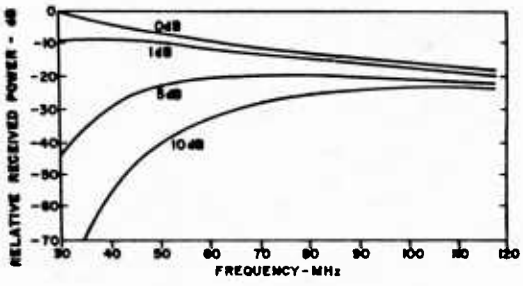


Fig. 20. Predicted relative peak power versus frequency for a 1000 km meteor scatter link in the presence of absorption. Four conditions of the ionosphere are illustrated in terms of 30 MHz vertical incidence riometer absorption. (Crysdale, 1960)

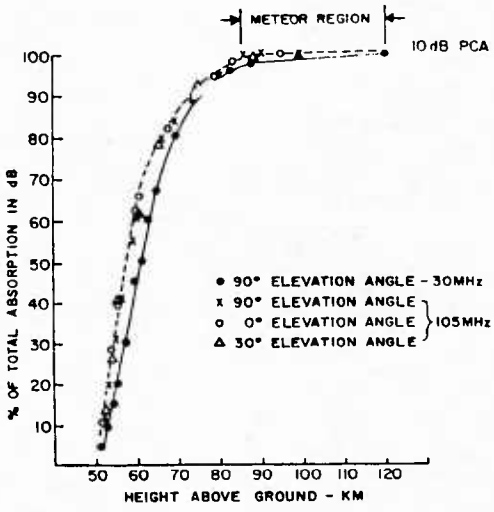


Fig. 21. Calculated D-layer absorption for a 10 dB SPE event as a function of height and angle of incidence at 30 and 104 MHz. at Thule, Greenland.

VHF/UHF/MICROWAVE LOS TERRESTRIAL PROPAGATION AND SYSTEM DESIGN

BY

T.K. FITZSIMONS

ALLIED RADIO FREQUENCY AGENCY,
IMS, NATO HQ, BRUSSELS

I. INTRODUCTION

By definition the frequency range covered by the title to this lecture is from 30 MHz up to tens of GHz. For military communications it means the use of systems such as Combat Net Radio, Single-Channel Radio Access, Tactical and Fixed Radio Relay, Air-Ground-Air Systems, Common User Information Systems and Tactical Satellite. For civil communications the systems are not dissimilar - mobile radio, fixed radio relay systems, Air Traffic Control, etc. Of course there is a considerable amount of broadcasting and there are military and civil radar systems in these frequency bands. The spectrum range from 30 MHz to about 5 GHz is for the military in particular, prime spectrum; wide bandwidths (hence digital transmission) can be achieved, most propagation mechanisms are relatively stable and most of this part of the spectrum can be used by tactical systems either operating on the move or moving frequently to new sites. In addition to considerations of wanted performance and necessary characteristics such as ECM performance, the military systems must be designed to meet requirements for rapid deployment in an environment which is difficult and unpredictable from the electromagnetic environment, topographic, and meteorological aspects.

2. NORMAL MODES OF PROPAGATION

Over the range 30 MHz to 15 GHz, there are several terrestrial propagation mechanisms. Excluding the tropospheric scatter mechanism which is the subject of a separate lecture, the main mechanisms are diffraction and line of sight. Both of these mechanisms can be modified by "normal" refraction and by reflection, though arguably where reflection plays a part, true line of sight conditions have not been established.

Diffraction. This is the main propagation mechanism exploited by systems operating over distances of a few tens of kilometres, requiring bandwidths which correspond to at least a single digital voice channel, and which operate on the move or from unplanned sites. Diffraction occurs when a sharp obstacle lies in the path between the transmitter and receiver; "sharp" being a relative term for the top of a natural or man-made feature which depends on the wavelength of the r.f. wave. At frequencies up to about 600 MHz, conditions exist in most of Europe to establish diffraction paths. Below this frequency, diffraction can occur due to rolling hills which are found in most regions. The geometry for classic diffraction is shown in Figure 1. Diffraction can occur with the tip of the obstacle either above or below the line joining the two antennas. The Fresnel-Kirchoff diffraction formula defines a dimensionless parameter:

$$\nu = \pm \sqrt{(2\Delta d/\lambda)}$$

where

$$\Delta d = r_1 + r_2 - d$$

the positive sign applying when the knife-edge is above the line of centres and the negative sign when the knife-edge is below. The diffraction loss, which is the ratio of free space field and the diffracted field is a function of this parameter which in turn is a function of d_1, d_2 and h_0. Diffraction can occur over a series of obstructions and there are several methods for calculating received field strengths (references 1-3). The models each have their advantages and disadvantages when applied to different situations and there are various degrees of difficulty in applying them.

The fact that diffraction is not only possible but can provide a reliable, steady signal makes the VHF and lower UHF range very attractive for use especially by mobile systems. The same characteristics which make the band attractive for this usage also makes prediction of signal level difficult and for military tactical systems not always practical for individual links.

The other normal mode of operation for terrestrial propagation in the VHF/UHF/SHF range is line of sight. The normal mode is of course subject to refraction due to the atmosphere's refractivity gradient and within the frequency range under consideration radio line of sight implies the use of a "straightening" factor of 4/3 applied to the earth's radius.

The normal modes of diffraction and LOS propagation through a standard atmosphere may of course include the occurrence of reflections producing extra signal components at the receiver. The reflections can occur at the earth's surface or it can be caused by the presence of hills off the line of the direct path. Analysing the effect of reflections on the signal level can therefore be very complex. A simplified approach can be taken for the case where there is a LOS mode and a single reflection from the earth's surface (Figure 2). For the conditions:

$$r \gg h_1', h_2'$$

and $$d \gg (h_1' + h_2')$$

the path length difference between the two rays, Δd, is approximately given by

$$\Delta d = 2h_1' h_2' / d \qquad (1)$$

If the magnitude of the direct wave field strength at the receiver is unity and that of the reflected wave is R, then the vector addition will be

$$\{1 + R^2 - 2R\cos(2\pi \Delta d/\lambda)\}^{1/2} \qquad (2)$$

for cases of small grazing angles where a phase change of 180 degrees occurs on reflection for both vertical and horizontal polarisations. The attenuation, A, relative to the direct ray is:

$$A = 10\log_{10}\{1 + R^2 - 2R\cos(4\pi h_1' h_2')/\lambda d\} \text{ dB} \qquad (3)$$

For the case of R = 1 and substituting for the Cosine term, the attentuation is

$$A = 10\log_{10}\{(-4\sin^2(2\pi h_1' h_2'/\lambda d))\} \text{ dB} \qquad (4)$$

$$A = -6 \text{ dB} - 20\log_{10}\sin(2\pi h_1' h_2'/\lambda d) \text{ dB} \qquad (5)$$

At short ranges, there is therefore a variation of signal level with change in antenna height (Figure 3). At VHF and lower UHF frequencies, height/gain tests are carried out when setting up a link to obtain the maximum signal.

Few paths occur, such that a single coherent reflection can be assumed. It is more likely that there will be contributions from several reflections. The effect on the steady state signal level may not be great but in the SHF range the situation can lead to multipath fading. The terrain in the reflection region is likely to be rough and for this condition an equivalent reflection surface has been proposed (Figure 4). The amount of terrain detail necessary for this type of approximation is not normally available when engineering tactical links. In addition, the effective antenna height is not easily determined in rough terrain and a reflection may in any case occur from an obstacle or terrain feature which is off-line and does not appear on the path profile chart. For these reasons, even though considerable amounts of digitised terrain data are now stored and available at short notice, it is doubtful whether it can contribute much to identifying and calculating the effects of multiple reflections.

Before looking at anomalous propagation, it is useful to review the structure of the atmosphere for normal propagation. Propagation of the radio wave is influenced by the atmosphere's refractive index, n. This is the ratio of the velocity of propagation in vacuum to that in air. It is not a uniform value but near the earth's surface the refractive index, n, is about 1.0003. A more useful indicator of the atmosphere's effect on e.m. waves is given by the refractivity, N, which by definition is:

$$N = (n - 1).10^6 \qquad (6)$$

This gives a value near the surface of about 300. It is related to the temperature, pressure and humidity of the air as follows:

$$N = 77.6\, P/T + 3.73.10^5\, e/T^2 \qquad (7)$$

where T = temperature (°k)
P = atmospheric pressure (millibars)
e = water vapour pressure (millibars).

This equation applies up to the EHF band.

In the standard atmosphere, the decrease of temperature and pressure are such as to produce a decrease in N of about 39 units/km. In such an atmosphere the bending of an e.m. wave towards the surface is such that the ray path can be depicted as a straight line when the earth's radius is increased by 4/3.

3. ANOMALOUS PROPAGATION

Deviations from the normal range of the refractivity gradient produce anomalous propagation. Normal refractivity, i.e. the radio wave bent towards the earth but with a curvature greater than that of the earth, occurs for refractivity gradients of zero to -79 units/km. Positive gradients produce upward bending (sub-refraction) and gradients of -79 to -157 N units/km produce downard bending greater than that of the standard atmosphere. Still more negative gradients produce curvature greater than the earth's and can cause ray trapping or ducting where the ray meets a reflecting lower boundary. These effects are summarised below and are shown in Figure 5.

TABLE 1

Refractivity Gradients

N Gradient (N/km)	Condition
< -157	Trapping
-157 to -79	Super-refractive
-79 to 0	Standard
> 0	Sub-refractive

When the atmosphere is not well mixed it is possible for super-refractive and trapping modes to become well established and produce exceptionally high signal levels over long distances. This can happen anywhere in the VHF/UHF/SHF range. Because it is anomalous it is not something that can be easily exploited and is likely to be the cause of interference.

Of more concern at the higher microwave frequencies (say above 2 GHz) is the phenomenon of multipath fading. Fading on microwave links has been observed for many years and occurs where the atmosphere is not well mixed. For the early (1950's) analogue FDM Links the distortion caused to them was not significant and early studies reported on the variations in mean signal levels. It was recognised though that the deepest and most rapid fades were associated with multipaths. Diversity techniques were used to overcome the problems. With the introduction of wideband digital systems the problem of inter-symbol interference caused by multipath fading became evident. As digital bandwidths have increased, spectrally efficient modulation methods have been introduced but these can make the system more susceptible to propagation effects, such as multipath fading. Multilevel modulation systems in particular can suffer a high level of intersymbol interference. In order to overcome the problem it is necessary to know the statistical properties of the propagation parameters or those of the signal itself. Hubbard (reference 4) has observed that since multipath is a multiplicative effect it cannot be overcome by additive processes such as increasing transmitter power or antenna gain. Adaptive techniques are required, able to follow the dynamics of the multipath disturbance.

4. OTHER SPECIAL PROPAGATION EFFECTS

Above about 2 GHz, attenuation due to precipitation can be observed on radio links. The extent though is not appreciable and for most of the world's climates the attentuation only becomes appreciable above about 8 GHz. Precipitation of any kind can attenuate or scatter the signal but the most likely to affect radio link performance for a significant amount of time and to a significant degree, is rain. The extent to which the signal is attenuated is a function of radio frequency and rainfall rate. The levels predicted by Ryde (reference 5) are shown in Figure 6. The two curves are for rainfall rates (12.5 and 50 mm/hour) which correspond roughly to "moderate" and "heavy rain". In designing to meet a required system performance it is obviously necessary to know how frequently it rains at the different rates. Figure 7 gives two rainfall rate distribution curves to illustrate two quite different climates - Singapore and north-west Europe. Low rainfall rates of less than about 5 mm/hour occur about as frequently in both climates; however there is more than an order difference for the incidence of heavy rainfall rates. The theoretical estimates of attenuation due to rain and the statistics of rainfall rates can be combined to produce predictions of attenuation due to rain. Figure 8 shows an example applicable to operation at 15 GHz. For a high reliability link, one that will have a specified received signal level higher than a stated value for say 99.9% of the time, provision would obviously be made at no great cost in N.W. Europe, while in Singapore it would obviously be necessary to restrict the length of the link to a few km or to provide a large signal margin probably in the form of additional transmitter power and antenna gain.

Absorption. Above about 20 GHz, atmospheric absorption becomes a significant factor. Figure 9 shows the familiar curves of attenuation against frequency for absorption by oxygen and water vapour molecules. At present it is mainly a problem for the design of fixed systems since few tactical systems are even potentially suited to operating above 20 GHz. However, the value of high signal attenuation in avoiding detection is recognised and it is probable that this will be a factor in exploitation of the windows or even the peaks in the absorption spectrum for special application.

5. EFFECTS OF TERRAIN AND MOBILITY OF TERMINALS

It is clear from the foregoing section on propagation effects that the terrain will play an important part in the design and deployment of communication systems operating in the VHF, UHF and SHF bands. Military users do have some fixed systems operating in this part of the spectrum but for the most part they are mobile (i.e. operating while moving) or transportable (high mobility but operating while halted). For a permanent, fixed system it is possible, for both civil and military systems to take a fairly complete account of the terrain and its effects. There is usually some choice available within a given locality to site antennas so as to clear local obstructions, avoid local reflections and to obtain first Fresnel zone clearance. There is also the possibility, at VHF and lower UHF frequencies to take calculated account of natural obstructions where they can provide a good diffracted signal. In these circumstances it should be possible to predict to a reasonably close degree, the effects of the

terrain on the signal level. Where this approach is taken with fixed systems, there may still be uncertainty as to the received signal level and its behaviour because of the presence of heavily wooded sections of the radio path. The effects of trees, both the trunks and leaves, on the received signal level have been studied for some time, particularly at VHF and UHF but gaps still exist in the knowledge concerning, for example, the permitivity of green wood, the effects of changing foliage, movement of foliage in the wind and so on (reference 6). Since the leaves can act as scatterers, there is obviously the possibility of the short-term nature of the signal as well as the mean signal level being affected.

Terrain problems become more difficult to deal with, and therefore potentially more restrictive for system design, when considering mobile or transportable systems. The military systems will be operated mainly in rural areas and have to contend with a wide variety of terrain and siting conditions. Civil mobile systems will be operated predominantly in urban areas with a high level of man-made noise and with propagation paths dominated by man-made structures. These terrain and environmental conditions play a significant part in determining the frequency band in which to operate the system and other important aspects of system design. Figures 10 and 11 show data collected from field observations of military tactical radio relay during exercises. When a tactical radio relay terminal is to be established it will have moved by man or vehicle to a previously agreed position, given by map location perhaps together with some description of paths and gateways to be used. Typical requirements for set-up time, after arrival on site, are of the order of 1 hour (tear down is of the order of 1/2 hour) for the larger systems such as radio relay. The mast carrying the antenna may be free-standing or it may be fixed to the side of a vehicle. In either case the exact location will not be chosen for the best propagation performance but will be determined by the need for concealment, the connection of power supplies, the stability of the terrain, the availability of real-estate on what could be a crowded co-site and many other factors.

In the absence of information about the exact situation of the remote end of a radio link it is normal practice to set up antennas and masts to give a horizontal line of shoot. If the two ends of the link are sited on terrain at different altitudes there will be a difference (the elevation angle) between the bore-sight angle of the antenna and the ray path joining the two antennas. The elevation angle can be positive or negative with respect to the horizontal and is a function of the difference in antenna altitudes, the length of the link and the earth's radius. Figure 10 shows the elevation angle as a function of link length based on measurements made on 208 LOS links established under exercise conditions. The curve is for values of elevation angle averaged over 5 km intervals. For link lengths above 20 km the average elevation is small ($\pm 0.5°$) but for links of 10 km and below, the range which might be used for SHF Links, the elevation angle becomes significant when compared with the beamwidths of the antennas one might wish to use. The average elevation angle for 10 km links was 1.2°. This will clearly limit the narrowness of antenna beamwidth and hence gain that can be employed on a tactical link.

Another way in which terrain features can affect the performance of tactical systems and influence their design is the number and distribution of obstacles in the immediate vicinity of the antenna. Figure 11 is based on data again obtained from the field during exercises. It shows the distribution of the heights of obstructions in the first 300 metres from the antenna position in the line towards the distant station. The data was obtained from 555 sites which were typical of those used for tactical radio relay. Different distributions will of course pertain to different geographic regions, however it can be seen that for the case chosen it is probably necessary to establish an antenna height of at least 10m above the local ground level if frequencies about about 300 MHz are to be used and good reliable links to be obtained. Other terrain factors affecting design and performance are mentioned in the concluding section of this paper which deals with some of the parameters which determine the bounds for a tactical microwave system.

6. PARAMETER BOUNDS FOR A TACTICAL MICROWAVE DIGITAL SYSTEM

In designing a new tactical communications system, the designer is frequently forced to consider the use of "new" frequency bands especially when the system is to carry digital transmissions. Single channel digital systems may require 25 or 50 kHz bandwidths compared with about 3 or 6 kHz used by analogue systems still used by the military. A 32 channel multiplexed link carrying 512 kb/s may require 0.5-1.0 MHz of bandwidth; up to eight times that required for an analogue system. The extra bandwidth requirements together with increased demand from all quarters for spectrum space has caused examination of the use of microwave frequencies in the tactical environment. Possible applications are for conventional tactical area network radio relay, local distribution and access and so-called down-the-hill, a tail from the area network to be an access point. Many of the considerations discussed here will apply though to other digital systems where there is a need to operate on the move or from unplanned sites where it is not possible to "engineer" the link. The MOD UK studied the problem of determining the upper limits of frequency band in which to operate a military tactical radio relay equipment and to assess the optimum frequency band to use in the SHF (LOS) range (reference 7). Factors which played an important part in the consideration were those of propagation and antenna performance.

Propagation. The required lengths of the links, in the range 10-30 km militates against the use of frequencies much above 10 GHz unless it were accepted that the system would only be deployed in a temperate (low rainfall) region. Predictions for rainfall attenuation at 15 GHz for example (Figure 8) are for approximately 1 to 3 dB/km attenuation for 0.1% of the time depending on the climatic region. Admittedly this type of attenuation can be overcome by increasing the power budget, either antenna gain or r.f. power, but there are limitations for a

tactical system on primary power generation and as will be discussed there are limitations on the antenna directivity that can be utilised. Microwave fading, especially frequency selective fading presents even more of a problem since an increase in power budget may not be the answer. Diversity techniques such as space or frequency, which are very suitable for fixed systems are not attractive solutions in a military tactical system for reasons of complexity, mobility and cost. Since there is no guarantee of avoiding fading above about 2 GHz one must reduce the probability of inter-symbol interference by limiting the baseband bit-rate and by choice of the modulation method.

Antenna and Feeder. The size of the antenna will be limited in the lower microwave frequencies by size, wieght or visibility considerations. As one goes higher in frequency the size of the antenna becomes limited because of problems associated with very narrow beamdwidths: the stability of the mast, the coupling between antenna beams as a consequence of height differences, mast altitudes, and accuracy of pointing. For a given size of antenna, there is a benefit in the path loss capability of a system by going up in frequency. In free space condition the path loss capability goes up as the gain of one of the antennas; however a point is reached where, for a tactical system the beamwidth becomes unmanageably small. It has already been shown that the antenna beam misalignment due to terrain factors can be in excess of 1° at each end of the link. A series of field trials (reference 7) showed that with the use of practical alignment aids it is possible to achieve a standard distribution of about $\pm 2°$ on the azimuth pointing angle. After a signal has been acquired it is then an easy matter to adjust the antenna aligment to maximise the received signal. However during the signal acquisition phase probably only 10-20 dB antenna coupling loss can be tolerated. Within rms errors of the order of 2°, the beamwidth of the antenna can not be less than about 6° without there being a substantial probability of failing to achieve signal acquisition at the first attempt and obviously, in a tactical situation, failure to acquire a signal in the first instance can lead to uncoordinated actions resulting in long delays or total failure.

In order to overcome losses caused by obstructions in the vicinity of the radio terminal it will normally be necessary to employ masts of about 10-12 metres. Feeders of either coaxial cable or flexible waveguides can be used if the r.f. amplifier is to be on the ground. In this case a tolerable loss of up to about 5 dB will obtain up to about 8 GHz. Above this frequency it becomes difficult to obtain feeders with an acceptable combination of mechanical and electrical characteristics and at about 15 GHz it is probably necessary to mount all or part of the radio equipment at the head of the mast.

7. THE CHOICE OF FREQUENCY BAND

In addition to the two important factors already discussed, things such as the r.f. power r generation efficiency, size of prime power source, overall size of the equipment and its match to the vehicle, the other systems which may be cosited, must all be considered. Many of them are inter-related and can not be considered in isolation. Of course not all of the radio frequency spectrum is available but for the military systems there are various bands which can be considered (reference 8). In the assessment study referred to (reference 7) it was concluded that the optimum frequency bands for an SHF radio relay equipment lie in the range 1.5 to 5 GHz. The problems which were identified, do not exclude working in the 15 GHz band but there is no doubt that they restrict the transmitted bit rate, the link length, the regions in which the system can be deployed and the ease of setting up a tactical system. Therefore even if there are many opportunities to establish LOS conditions, equipments operating in the 15 GHz band can not be considered as interchangeable with those in the UHF band for making links in a tactical area system.

8. ANALYSIS AND DESIGN TOOLS

Numerous computer models exist to aid the design and analysis of equipments and systems. Digital computer models based on the Fast Fourier Transform have existed since the 1970's to analyse analogue and digital communications equipment performance. They produce very accurate results when modelling equipments operating under normal propagation conditions. To predict performance in the presence of say intersymbol interference caused by multipaths or to compute the effect of different modulation methods or filter design in multipath conditions is probably best suited to an analogue or hybrid computer. Unfortunately these are very expensive to instal and require considerable efforts in setting up and maintenance. There are system computer models which complement the equipment models. Specific ones are available for studies of reliability and availability. In order to assess say the effects that anomalous propagation might have on a particular aspect of network performance, the results from one model need to be used as the input to another though the output and input formats may be dissimilar and require some form of translation.

9. CONCLUDING REMARKS

For military users, the application of digital radio transmission is in some cases still in the initial stages. In the tactical environment particularly there is still the need to experience the effects of propoagation and antenna characteristics on link and system performance. During the period that this experience is gained and the presently perceived problems of working at SHF frequencies are overcome, it is likely that the demand for information transfer will go on increasing. This could mean higher transmission rates for both fixed and mobile systems and designers will be presented with a new set of problems caused by the large bandwidths and the need to exploit other lightly used parts of the spectrum.

References

1. Bullington, K. "Radio Propgation at Frequencies above 30 Megacycles" Proc. I.R.E, Oct 1947.

2. Epstein, K., Peterson, D.W., "An experimental study of wave propagation at 850 Mc/s" Trans. I.R.E. Antennas and Propagation, April 1955.

3. Deygout, J. "Multiple Knife Edge Diffraction of Microwaves". IEEE Trans. Vol. AP-14, July 1966.

4. Hubbard, R.W. "A Review of atmospherice multipath measurements and digital system performance". AGARD Conference Proceedings N° 363. June 1984.

5. Ryle, J.W., Ryle, D., "Attenuation of centimetric waves by Rain, Hail and Clouds". Report 8516, GEC Research Labs., Wembley, England. May 1945.

6. AGARD Conference Proceedings, CP-363. June 1984 - Discussion of Session IIA.

7. Mitchell, A.M.J., FitzSimons, T.K., "Some basic design considerations for an SHF military radio relay equipment". SRDE (MOD, PE) Report 71012. Nov 1971.

8. FitzSimons, T.K., Problems of Spectrum Management and System Compatibility. AGARD Special Course: Interaction of Propagation and Digital Transmission Techniques. Oct 1986.

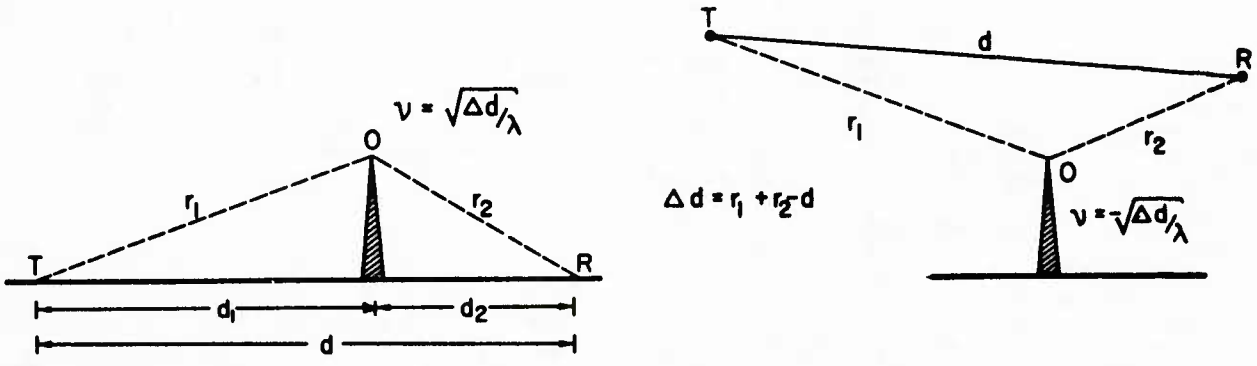


Figure 1 KNIFE-EDGE DIFFRACTION

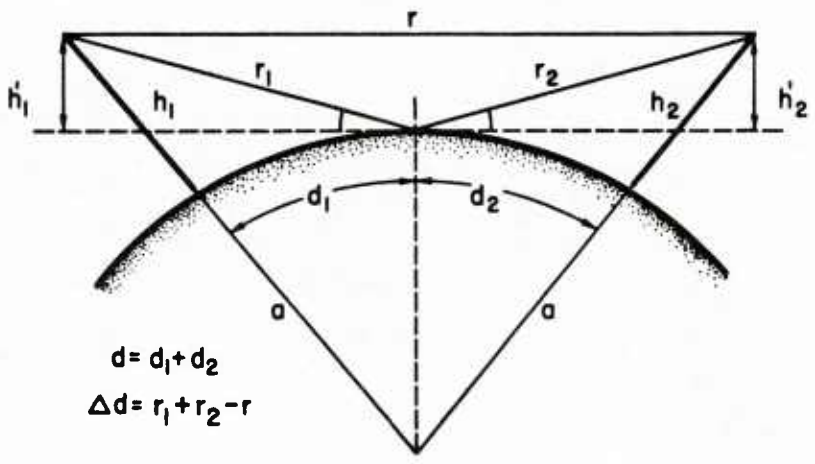


Figure 2 REFLECTING PATH OVER SMOOTH SPHERICAL EARTH

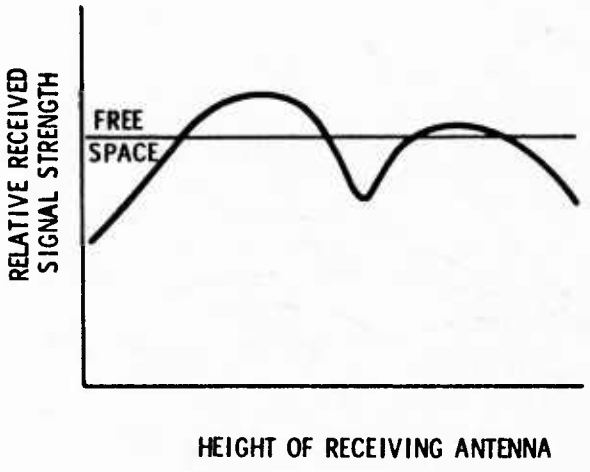


Figure 3 VARIATION OF SIGNAL STRENGTH WITH HEIGHT IN THE REFLECTING REGION

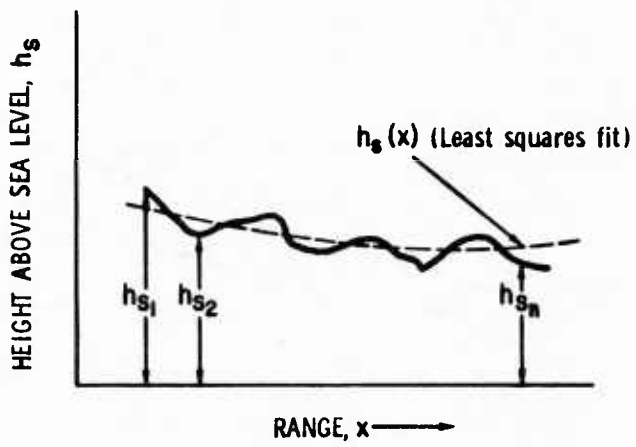


Figure 4 CURVE FIT TO ROUGH TERRAIN TO OBTAIN EQUIVALENT REFLECTING SURFACE

a. Trapping
b. Super-refractive
c. Standard
d. Sub-refractive

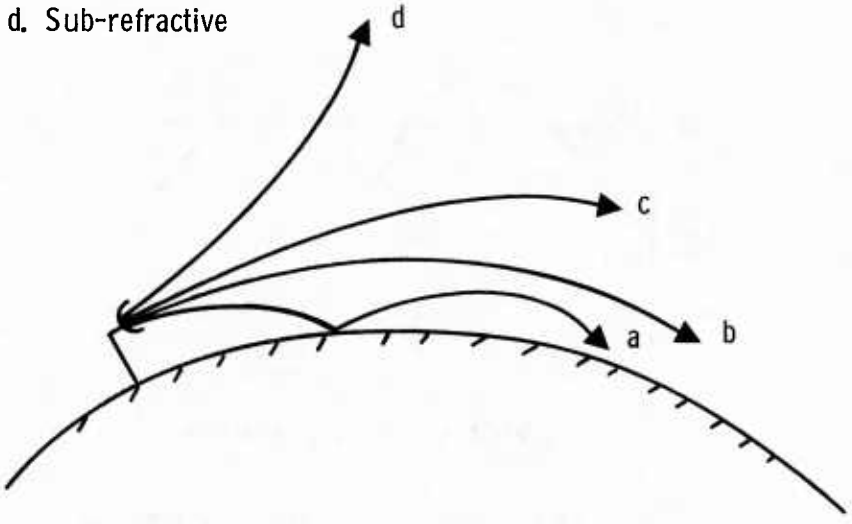


Figure 5 Ray bending above earth's surface

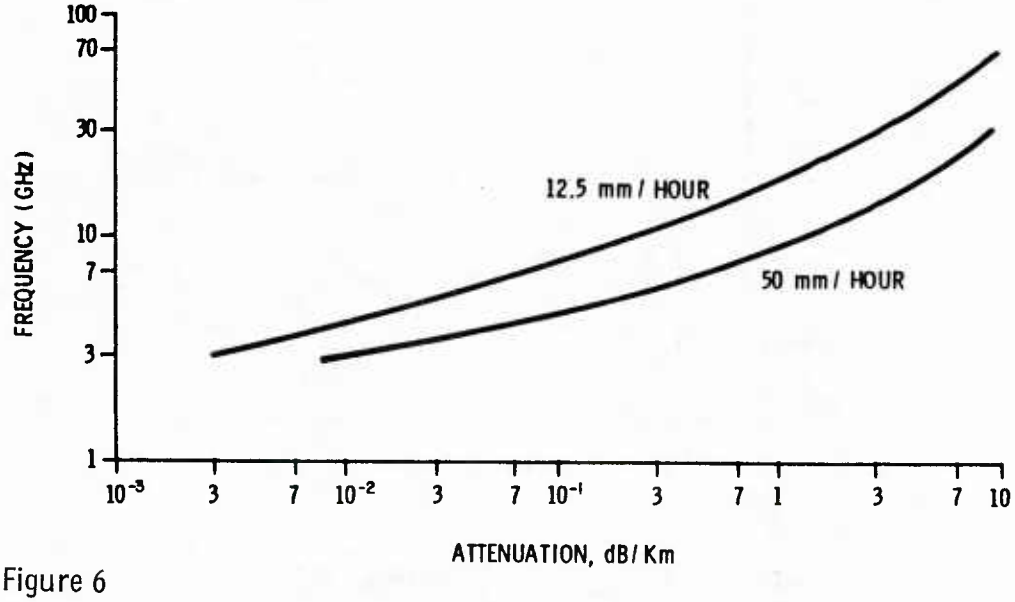


Figure 6

RAIN ATTENUATION AS A FUNCTION OF FREQUENCY. (RYDE THEORY)

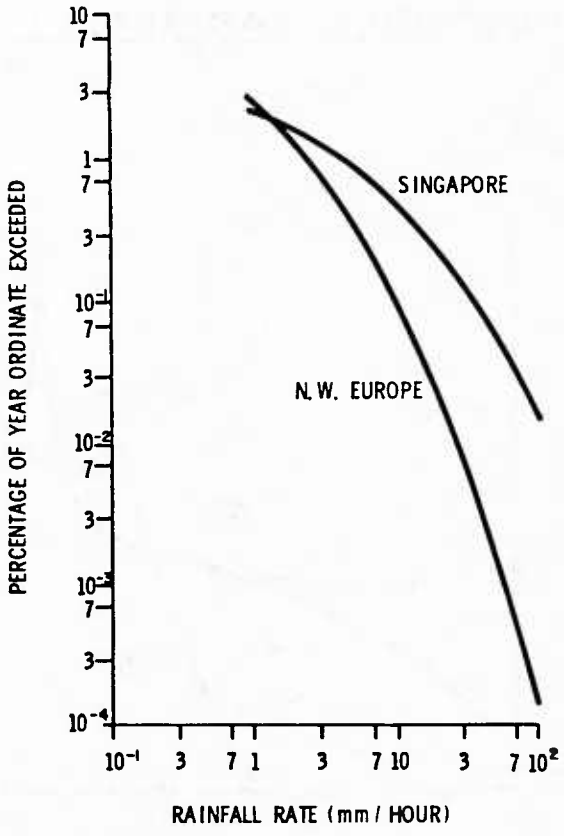


Figure 7

RAINFALL RATE DISTRIBUTIONS

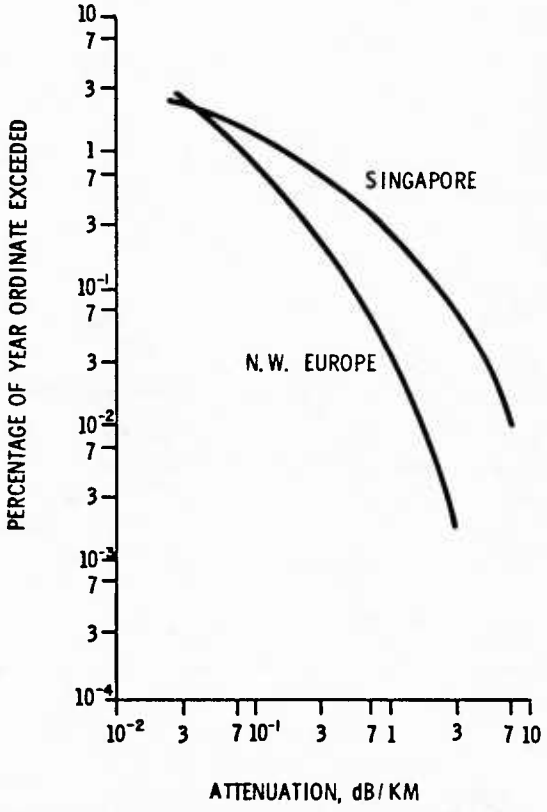


Figure 8

PREDICTED RAIN ATTENUATION
DISTRIBUTION AT 15 GHz

ATMOSPHERIC ABSORBTION

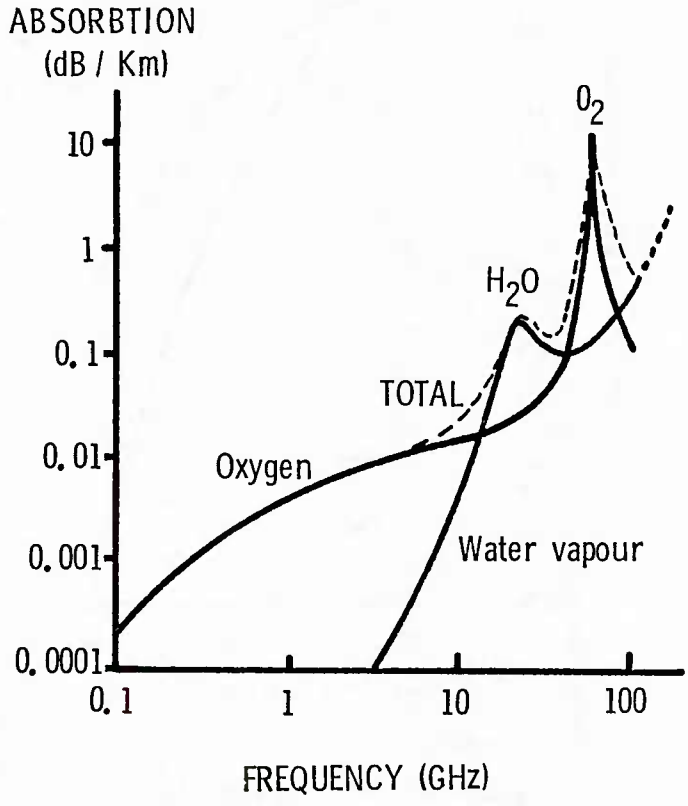


FIGURE 9

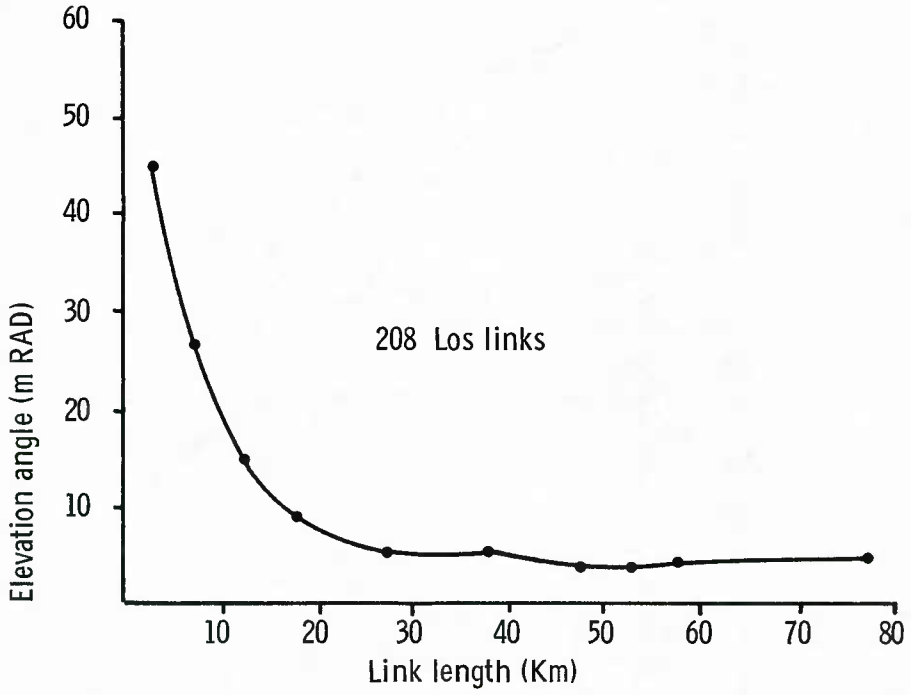


Figure 10 Variation of elevation angle with link length

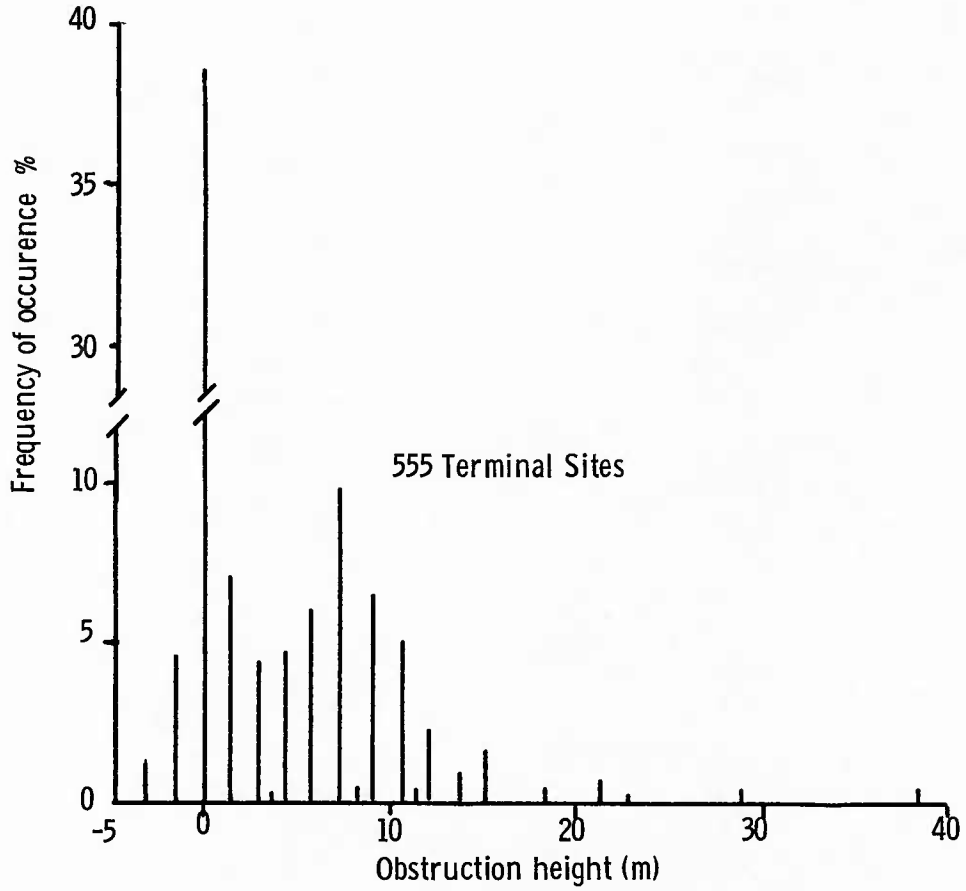


Figure 11 Distribution of local obstruction heights in first 300 m.

UHF/SHF SATCOM PROPAGATION AND SYSTEM DESIGN

by

Allen L.Johnson
US Air Force Wright Aeronautical Laboratories
AFWAL/AAAI
Wright Patterson AFB, Ohio 45433-6543
USA

UHF/SHF SATCOM PROPAGATION AND SYSTEM DESIGN

I. Propagation Aspects of Satellite Communications

One year after the launch of the Russian Sputnik, the Americans orbited the world's first active communications satellite, SCORE. The communications planners hoped that the line-of-sight communications through satellites would overcome the various propagation problems plaquing long-distance High Frequency (HF) communications. The communications satellite did overcome many of the traditional HF problems but due to the law of conservation of difficulty, a new set of propagation problems emerged. This section deals with some of those problems.

A satellite orbiting at approximately 19,200 nautical mile altitude rotates at the same rate as the earth and, if in an equatorial plane, appears to remain fixed over one point on the earth. The coverage of an equatorial synchronous satellite is shown in Figure 1.

A. Normal Propagation Mechanisms

The earth-to-satellite link experiences a variety of propagation effects such as free space loss, attenuation, polarization changes, refraction, and delay as a result of the physics of the link and the normal earth's atmosphere.

1. Free Space Loss

The largest loss factor in a earth-to-satellite link is the so-called "free space loss". In fact, there is no loss of energy associated with free space, just a space spreading of the energy. As the energy is emitted from a point source, it radiates out spherically and the farther it travels, the less energy per square unit of measure exists.

The equation for the power received at a distant receiver can be stated as:

$$P_r = P_t \cdot A_s \cdot (G_t/4\pi R^2)$$

where P_r = received power
P_t = transmitted power
A_s = receive antenna capture area
G_t = transmitter antenna gain
R = distance between transmitter and receiver

and $G_t/4\pi R^2$ = isotropic radiating factor.

The isotropic radiation factor is the relationship between the spherical angle of the main beam of the high gain antenna compared with an omni-directional antenna.

Note that the amount of power received at the distant receiver is independent of frequency in this equation. This is inherently obvious if you consider, for example, a low frequency transmitter radiating one watt of energy omni directionally and an optical transmitter, such as a light bulb, radiating one watt of optical energy omni directionally in free space. If spheres were constructed with approximately a 282 meter radius from each of these transmitters, the spheres would intercept the entire one watt of energy. Each sphere would contain approximately 1 million square meters of surface area. The energy density intercepted by each sphere would be 1 microwatt per square meter. If one square meter antennas were located on each of these spheres, both antennas would intercept the same amount of radiated energy, one microwatt, Figure 2. There would be some practical differences in the antenna efficiency at different frequencies but these differences are not related to free space loss. In the typical link calculation equations free space loss appears to be frequency sensitive. This is because the typical equation inserts receiver gain, G_r, in place of antenna intercept area.

$$G = 4\pi A/\lambda^2$$

where λ = transmitted wavelength.

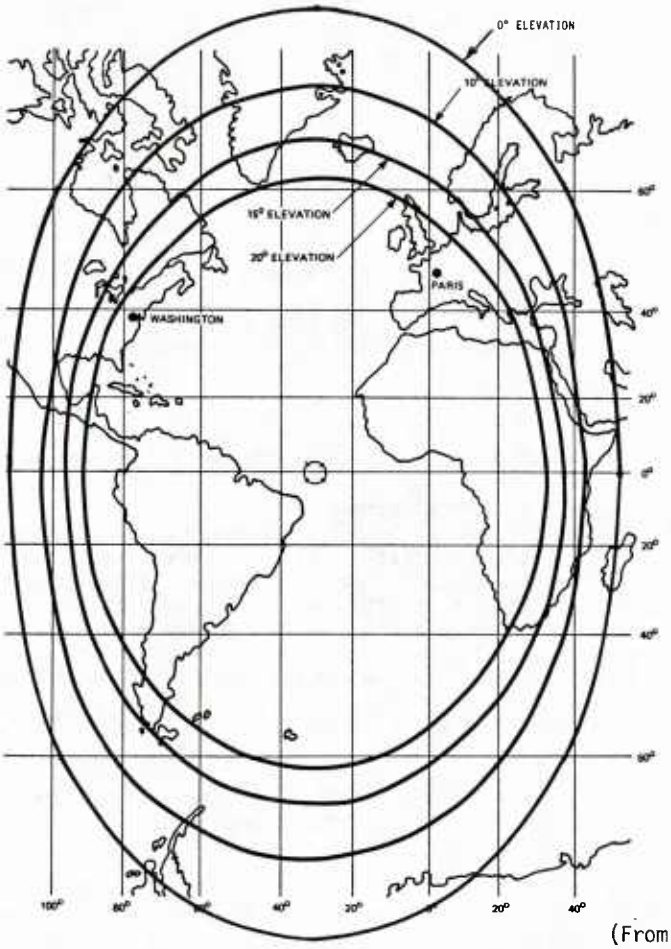
Therefore, if the cross-sectional area of the receive antenna is replaced with gain, the power received equation becomes:

$$P_r = P_t \cdot G_t \cdot G_r \cdot (\lambda/4\pi R)^2.$$

The quantity in parentheses is normally called the "free space loss," α.

$$\alpha = (\lambda/4\pi R)^2.$$

A plot of free space loss, α, versus frequency for an earth-to-synchronous satellite is shown in Figure 3.



(From DCA, 1972)

FIGURE 1. COVERAGE CONTOURS FOR SYNCHRONOUS SATELLITE

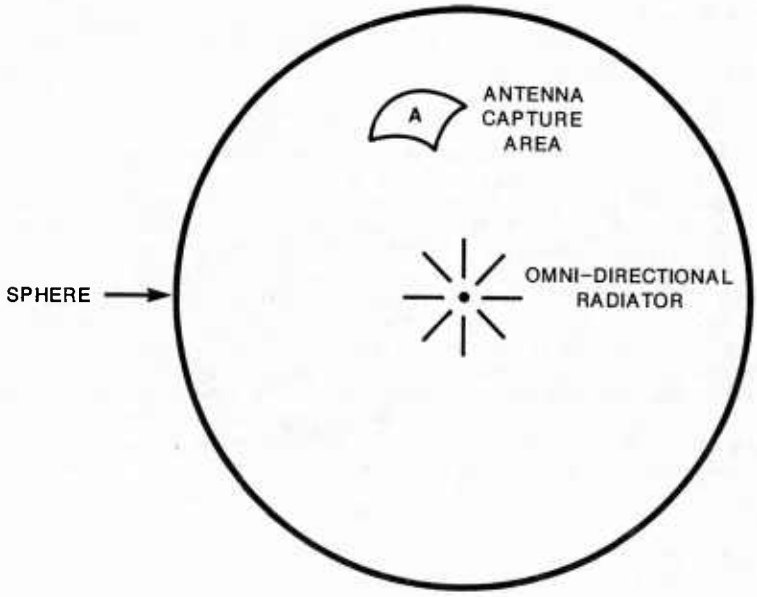


FIGURE 2. SPHERE INTERCEPTING ENERGY RADIATED BY OMNI-DIRECTIONAL SOURCE

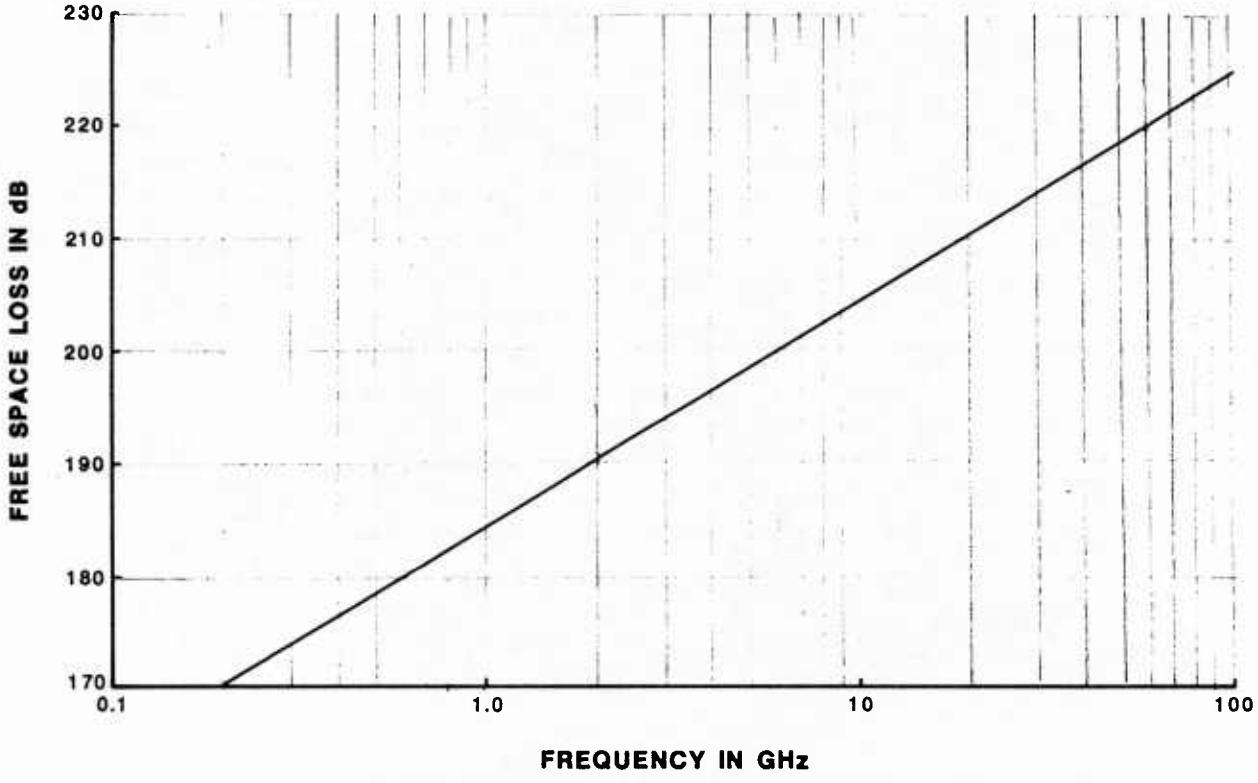
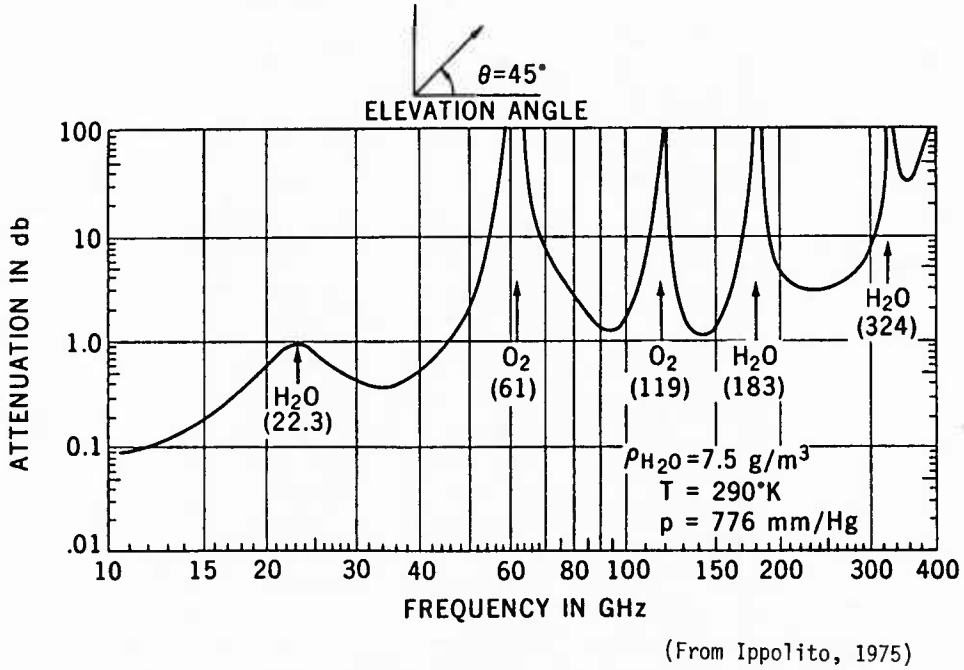
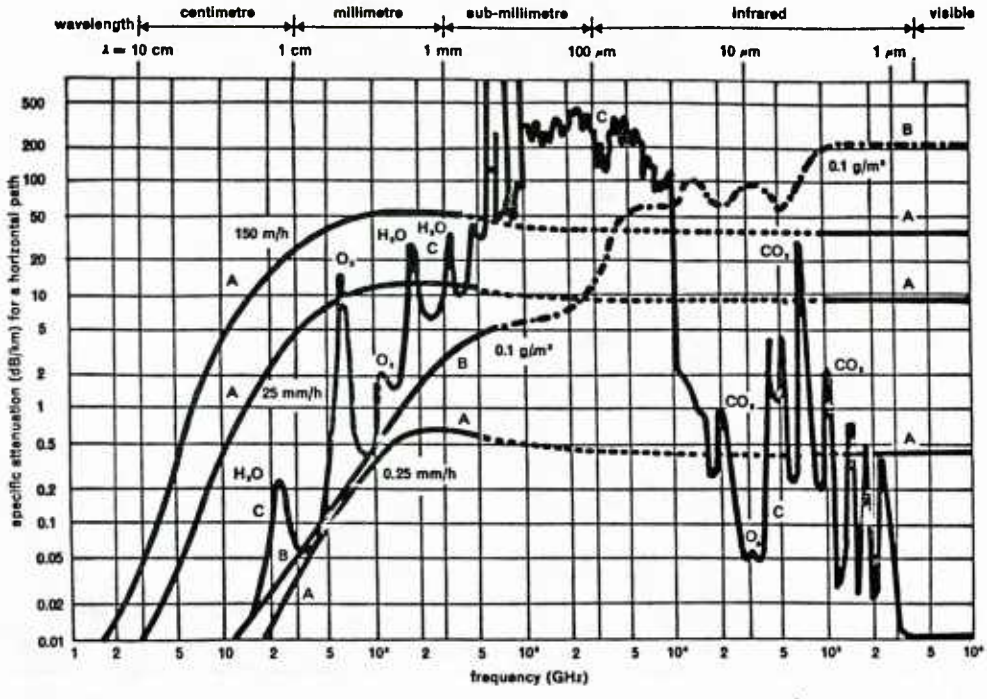


FIGURE 3. FREE SPACE LOSS FROM EARTH TERMINAL TO SYNCHRONOUS ALTITUDE SATELLITE



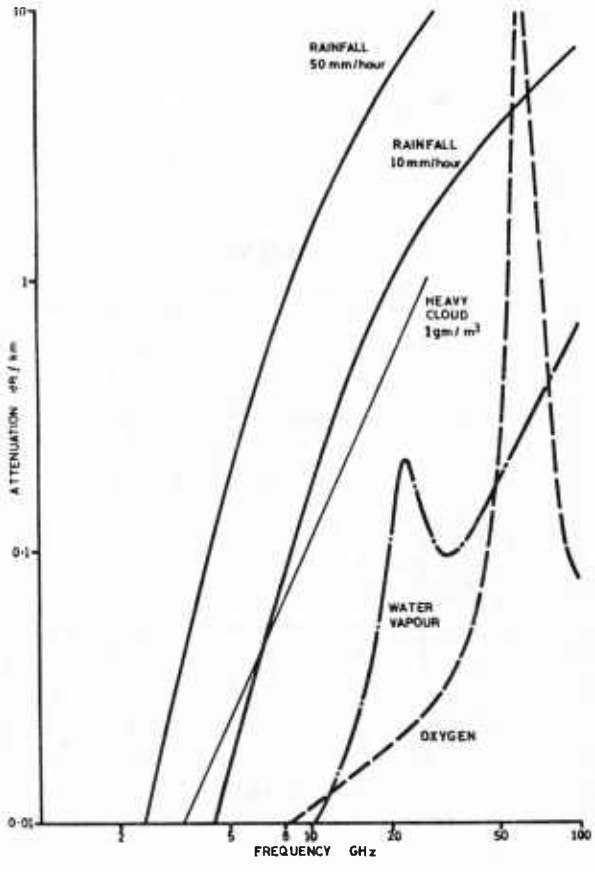
(From Ippolito, 1975)

FIGURE 4. TOTAL ATTENUATION OF THE ATMOSPHERE DUE TO RESONANCE ABSORPTION



(From Bem, 1979)

FIGURE 5. SPECIFIC ATTENUATION FOR RF, IR, AND VISIBLE LIGHT



(From Crane, 1971)

FIGURE 6. CAUSES OF ATMOSPHERIC ATTENUATION

2. Attenuation

The attenuation on the earth/space path primarily occurs in the first ten kilometers above the earth (extracted from Goodman, 1980). The upper troposphere, ionosphere, and beyond have negligible attenuation. The radio frequency attenuation of the atmosphere is due primarily to gas absorption, fog or clouds, and precipitation.

Radio frequency absorption in gases is primarily limited to H_2O and O_2 absorption lines. This absorption starts to become significant at resonant frequencies around 24 GHz and 60 GHz. Other than those resonant points, the gas absorption is very low as can be seen in Figure 4. Likewise the absorption in fog or clouds is less than one hundredth of a dB per kilometer below 10 GHz, Figure 5.

Attenuation due to rainfall can vary from 0 to 2 dB at the UHF frequencies to 2 to 10 dB for the SHF frequencies, Figure 6. At EHF the rainfall attenuation can get as high 25 or 30 dB for a heavy thunderstorm.

Another attenuation factor that must be considered, especially in the wet tropical areas, is the foliage loss. Field measurements from a ground transmitter to a satellite through foliage indicated that in the UHF band the attenuation may vary from 1 to 15 dB and for SHF from 2 to 25 dB. By choosing a new site for the antenna it is often possible to significantly reduce the foliage attenuation.

3. Polarization Effects

The polarization of the signal between earth and space is affected by hydrometeors and by electrons or ions in the path which cause Faraday rotation (extracted from Goodman, 1980). The major effect is the Faraday effect where the rotation is related to the total electron content of the ionosphere and the wavelength squared. As the frequency gets larger, or conversely the wave length gets smaller, the amount of Faraday rotation diminishes rapidly. For example, in the SHF band, 7GHz, the polarization change due to Faraday may be a few degrees for a typical total electron content. At a 250 MHz UHF frequency the same total electric content would give 1000° of Faraday rotation and would, therefore, make the use of linear polarization questionable. This can be overcome by using circular polarization, especially for the lower frequency transmissions. The other polarization problem has to do with the depolarization caused by fog, rain, or ice particles. In general, on the earth/space path the effect of depolarization can be considered negligible.

4. Refraction

Refraction occurs when a radio wave propagates from one medium to a medium that has a different refractive index (extracted from Goodman, 1980). In effect, the propagation on the earth space paths is passing through a large number of micro medias, each with a small refractive index change, Figure 7. This causes a bending of the radio wave and results in the ability to communicate below the geometric horizon. The correction made for a standard earth's atmosphere is to substitute a slightly larger (four-thirds) earth radius, in which case the bending radio waves can now be drawn as a straight line from the earth to the satellite.

In summary, the amount of refraction is related to the wave length squared and therefore decreases rapidly with frequency. The range of the refraction angles varies from a few degrees at UHF to fractions of a degree at SHF and are, therefore, not particularly significant.

5. Propagation Delays

The troposphere and ionosphere delay the propagation of the radio signal on the earth/space paths by a small amount must be added to the speed-of-light delay to determine the total delay (extracted from Goodman, 1980). The ionospheric delay is proportional to the wavelength squared, but the tropospheric delay is basically frequency independent. These delays could add a few microseconds at 250 MHz or a few nano seconds at 8 GHz. Unless the satellite signal is being used to measure exact range, for say navigation purposes, these delays are negligible.

B. Ionospheric Scintillation Fading

The major geographic effect of concern in satellite communications, other than rain or foliage attenuation, is ionospheric scintillation. There are six major factors which affect the intensity of ionospheric scintillation on the earth-satellite path. These are the transmission frequency, geographic location, time of day, season, magnetic activity, and solar flux.

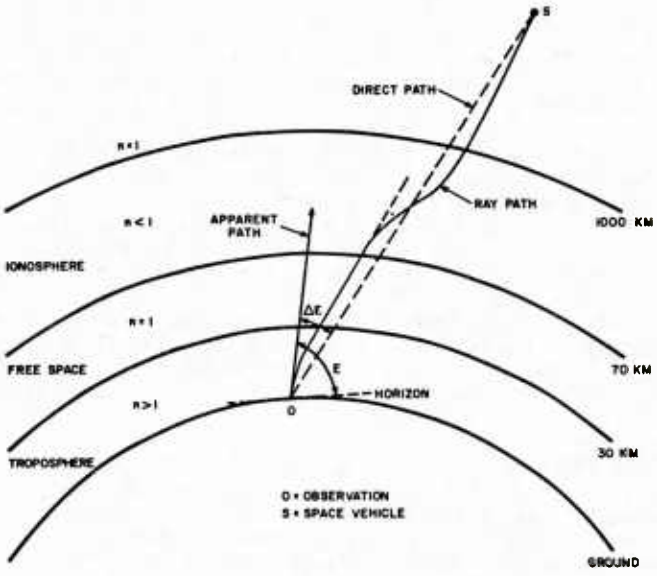
Ionospheric scintillation fading occurs as a result of sharp ion or electron gradients that occur in the ionosphere, Figure 8. These sharp gradients are caused by ionospheric irregularities that tend to refract or focus the radio waves as they pass from the earth's surface to the satellite, Figure 9.

There are three major scintillation regions, the equatorial zone ±20° from the magnetic equator, the mid-latitude zone, and the high-latitude zone approximately 60° to 90° from the magnetic equator, Figure 10.

1. Equatorial Ionospheric Scintillation

The past decade has seen an extensive multi-nation campaign to quantify and model the equatorial ionospheric scintillation phenomena (Aarons, 1977; Ossakow, 1979; Koster, 1971; Johnson, 1980). A comprehensive picture of the ionospheric irregularity has been developed by theoretical studies, earth terminal-to-satellite observations, and in situ measurements along the magnetic equator. The current knowledge indicates that the ionospheric irregularities form in the F region (200 to 400 kilometers altitude) one to two hours after local sunset. The irregularities develop in patches along the equator, developing westward at a velocity approximating the sun terminator velocity. The patches develop with an east-west size to 50 to 200 kilometers (Aarons et al, 1980-1). Following the initial irregularity development, an upward velocity occurs causing the patch to extend from its 200 kilometer development altitude to as high as 1,000 kilometers. During this vertical development there is also a north-south propagation of the irregularity along the magnetic field lines. Recent measurements indicated the north-south propagation to occur at speeds of approximately 400 meters per second (Johnson et al, 1980).

The equatorial ionospheric irregularities result from a depletion of electrons compared with the surrounding F region. This depletion results in sharp gradients with abrupt changes in electron content of one to three orders of magnitude. Once formed, the irregularities drift in an easterly direction with velocities of 100 to 200 meters per second. The initial velocities, immediately after formation, appear to be higher than the later velocities. Individual irregularity patches have been tracked for periods of up to three hours using an all-sky photometer at 6300 Å to measure the air glow (Weber et al, 1978). Irregularity patches with a north-south extent of two thousand kilometers have been observed. Irregularities



(After Millman, 1967)

FIGURE 7. RADIO WAVE TRAJECTORY THROUGH THE TROPOSPHERE AND IONOSPHERE

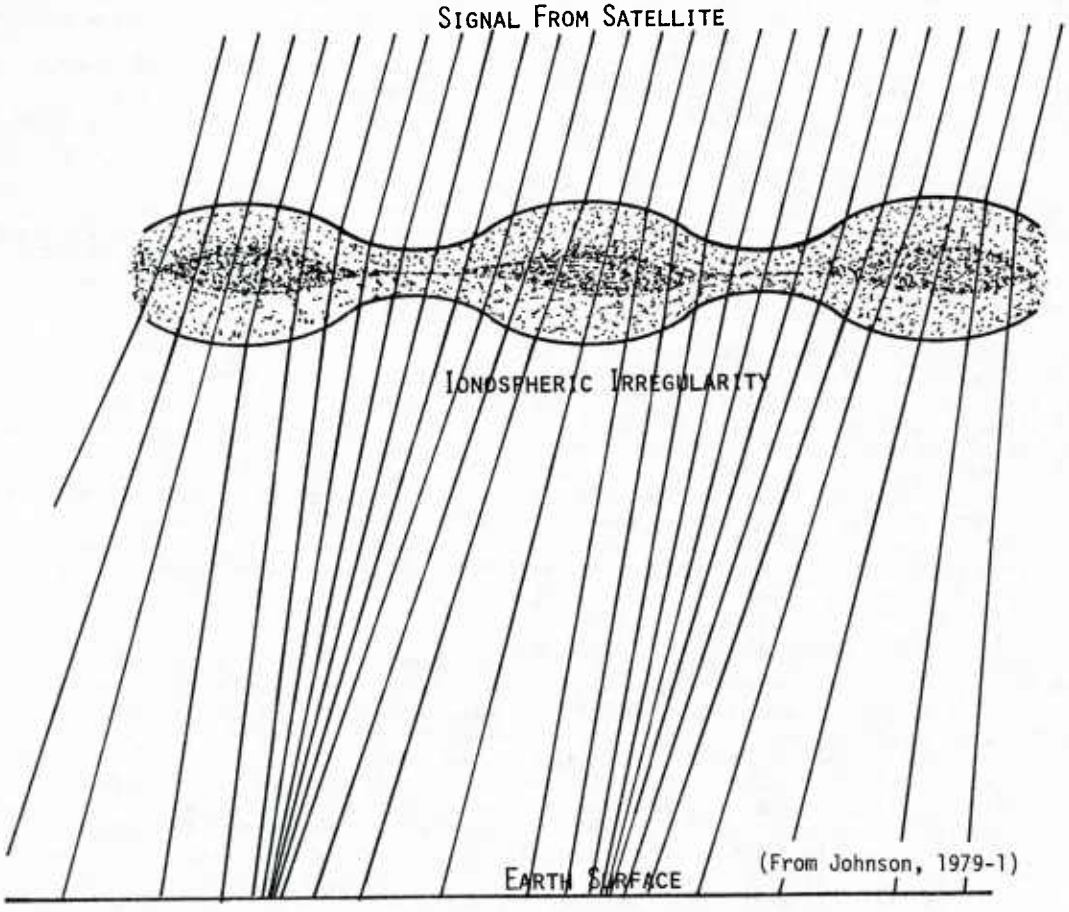
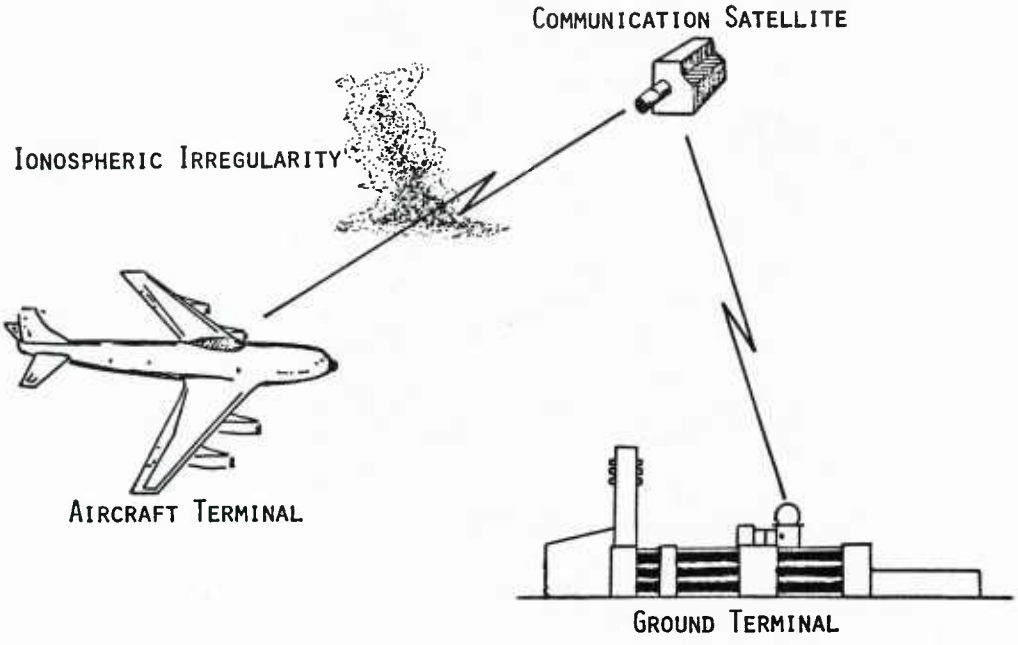
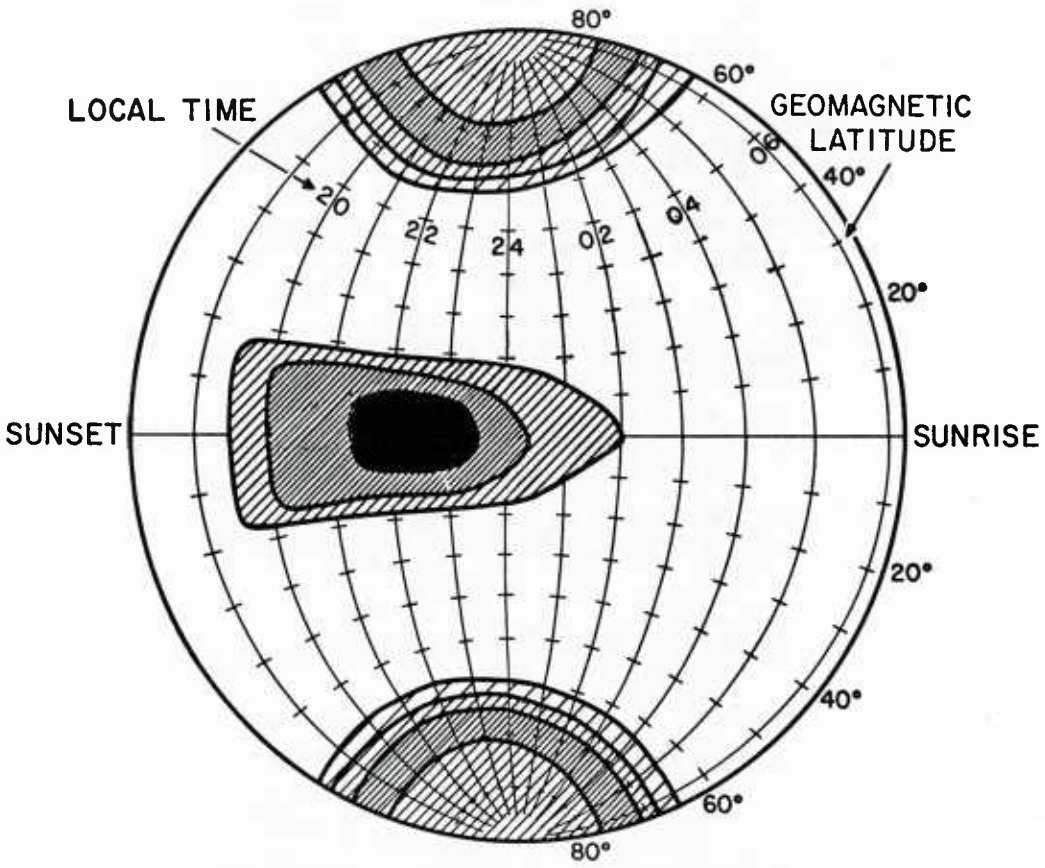


FIGURE 8. SIMPLIFIED REPRESENTATION OF IONOSPHERIC FOCUSING EFFECT



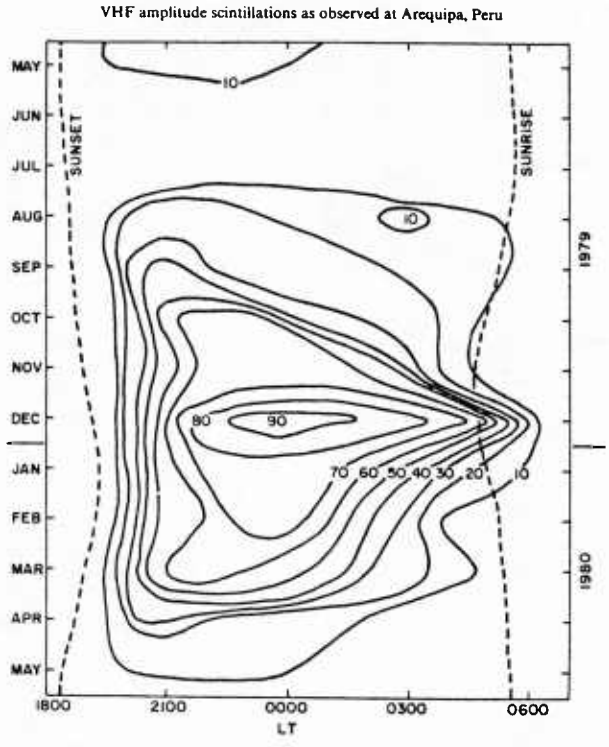
(From Johnson, 1979-1)

FIGURE 9. COMMUNICATIONS PATH THROUGH IONOSPHERIC IRREGULARITY



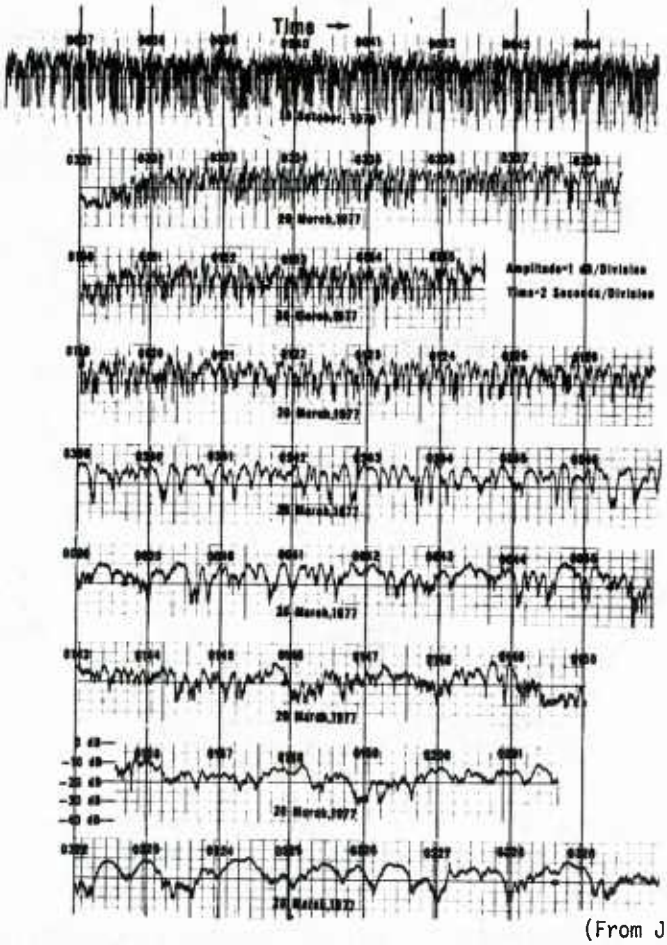
(From Aarons, 1982)

FIGURE 10. GEOGRAPHIC DISTRIBUTION OF IONOSPHERIC SCINTILLATION FADING



(From DasGupta et al, 1983)

FIGURE 11. MONTHLY CONTOUR PLOT SHOWING PERCENTAGE OCCURRENCE OF SCINTILLATION



(From Johnson, 1979-1)

FIGURE 12. EQUATORIAL UHF IONOSPHERIC SCINTILLATION FADING VARIATION

often occur in series with alternating disturbed and undisturbed ionospheric patches. Ground measurements and all-sky photometer measurements have identified groups of four or five patches more or less uniformly spaced (Weber et al, 1978). Along the magnetic equator the irregularity patches tend to have a thin fabric of irregularities connecting the patches. This thin irregularity structure between patches causes only minor ionospheric scintillation. As one travels away from the magnetic equator, the fabric connecting the irregularities disappears leaving distinct field aligned irregularities separated by undisturbed ionosphere.

The occurrence of equatorial ionospheric disturbances appears to increase during the sun spot maximum. There also appears to be a negative correlation between irregularity formation and magnetic index (Koster 1971, Singleton 1978). During period of strong magnetic activity scintillation occurrence is minimized. The occurrence of equatorial ionospheric irregularities follows a seasonal pattern which varies, depending upon the observer's location along the magnetic equator. In general, the occurrence peaks up during the equinox periods and remains relatively strong during the local summer, Figure 11. A minimum occurrence is noted during the local winter.

The trigger mechanism which cause the irregularity formation has not been identified. During the scintillation season an observer along the magnetic equator may experience several nights of strong irregularity formation interspersed with nights of absolutely no ionospheric disturbances. Correlation with solar flux density, or magnetic intensity offers little explanation for the erratic occurrence of the irregularity patches.

The irregularities primarily affect the VHF and UHF bands. Earth terminal-to-satellite links of 100 to 500 MHz are severely affected by the equatorial ionospheric scintillation. When the irregularity patches occur, fading of 20 to 35 dB peak-to-peak is regularly encountered, Figure 12. Frequencies from 500 MHz up to 2 GHz experience moderate scintillation fading with 2 GHz fades of 10 to 25 dB under severe conditions, Figure 13. Frequencies from 7 to 8 GHz experience only minor scintillation effects of 1 or 2 dB fading during severe irregularities, Figure 14.

The onset of scintillation is very abrupt, especially for an airborne terminal. The fading often changes from less than 1 dB to 20 or 30 dB in sixty seconds. With aircraft flying in a westerly direction, against the eastwardly drift of the irregularity, abrupt onsets of scintillation have been observed in 10 to 20 second periods. The fading rate tends to be greater in the premidnight period when the velocity of the irregularity is normally greatest. Airborne measurements show that the fading rate is sensitive to aircraft direction. An aircraft flying westwardly, against the irregularity drift, may record fading rates of 1 Hz or greater. An aircraft flying in an eastward direction can effectively synchronizes its motion with that of the irregularity patch and may encounter one fade in 60 to 120 seconds. The severe amplitude fading encountered in the equatorial region tends to be Rayleigh distributed. The equatorial ionospheric irregularities do not present an attenuation factor to the satellite-to-earth signal but merely redistribute the energy into enhance areas and null areas. Enhancements of 6 to 10 dB are encountered during severe fading along with fades of 20 to 30 dB.

The strong amplitude scintillation is accompanied by phase scintillation. Measurements made at 250 MHz in the equatorial region showed phase scintillations with abrupt phase changes at a rate of 2 pi to 16 pi radians per second, Figure 15 (Prettie et al, 1977).

In an effort to obtain additional information about the characteristics of the field aligned equatorial irregularities, an extensive measurement campaign was conducted on Ascension Island, near the equator in the South Atlantic, in 1981, (Johnson, 1986). Spaced antenna/receivers were set out in several configurations to observe UHF signals from the MARISAT satellite and evaluate signal amplitude fading correlation over north-south distances of up to 1.6 kilometers.

The cross-correlation of the signals from two spaced receivers provided a measure of their similarity. Cross-correlations for the different north-south antenna spacings were calculated and analyzed for dependence on receiver separation and fade rate. No account was taken of other factors, such as time of occurrence, which could cause variations in the dependence relationship. For each distance, the cross-correlation was plotted versus fade rate. The data points displayed some scattering but they show unmistakable trends, and a linear curve fit was made for each set of data. The family of north-south distance curves showing cross-correlation versus fade rate are depicted in Figure 16, (Hocutt, 1983).

In general, the measured dependence of cross-correlation on fade rate increased as distance increased. The 560 meters and the 775 meters curves, calculated from data collected on different days and times, did not follow this trend. Also, the two 1 km curves had different slopes. The curves have not been corrected for time of occurrence or other factors which could enter the fade rate - cross-correlation relationship.

Very slow fading (0-20 fades/min) provided almost perfect cross-correlations for all observed north-south distances, including the 1.6 km range, Figure 17. As the fade rate increases to 40 fades/min, the correlation is still better than 0.9. More rapid fading causes the correlation distance to decrease until by 80 fades/min the distance which provides 0.90 correlation has dropped to 260 meters or less.

In addition to the relationship between cross-correlation, fade-rate, and receiver separation noted by Hocutt, 1983, east-west skewing of the north-south irregularity alignment was observed on several occasions (Johnson, 1986). The north-south antenna alignment was initially adjusted to assure similar fades arrived at both ends of the 1.6 km spaces receivers to within less than .03 seconds. For the average west-to-east velocity encountered during this campaign (125 m/sec) this corresponded to an angular measurement accuracy of approximately 0.1°.

The alignment of the magnetic field varies in a complex way from the earth's surface out to the magnetosphere. On the earth's surface 99% of the main magnetic field originates from sources within the earth (Vestine et al, 1947). At increased altitude above the earth's surface the contribution of local magnetic anomalies decrease rapidly and the magnetic field becomes more dipolar. Contributions of the equatorial electrojet, solar particles, and other currents make the magnetic field at altitude very dynamic. In addition to the long term magnetic field, diurnal variations, and magnetic storm disturbances, there are also short term variations with periods of seconds, (Johnson, 1961).

To investigate the magnetic alignment variations, data from 1 February was analyzed. The majority of the data collected on that night showed a cross-correlation between the 1.6 km receivers with zero time offset. However, on several occasions, distinct time offsets of 0.1 or 0.2 seconds were observed, Figure 18. An offset of 0.2 seconds corresponds to an angular change of the north-south alignment of approximately 0.9°. Since these occurred late in the evening when the irregularity turbulence tends to be small, the offset is not likely due to turbulence. One hypothesis is that the irregularity height was different for the non-zero time-lag compared to the zero time-lag irregularities. If this were the case then the mapping of the irregularity alignment on the earths surface would provide information on the irregularity height.

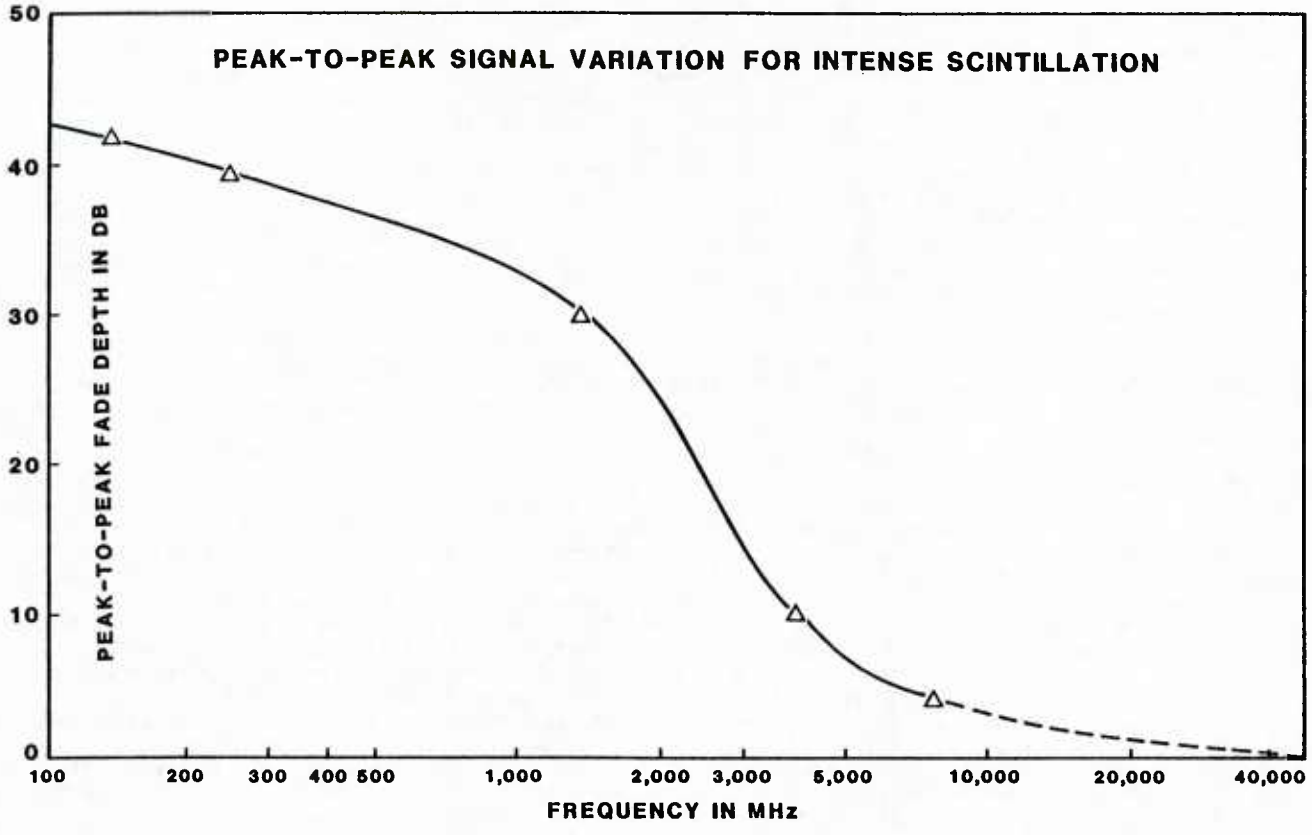


FIGURE 13. FREQUENCY DEPENDENCE OF IONOSPHERIC SCINTILLATION FADING

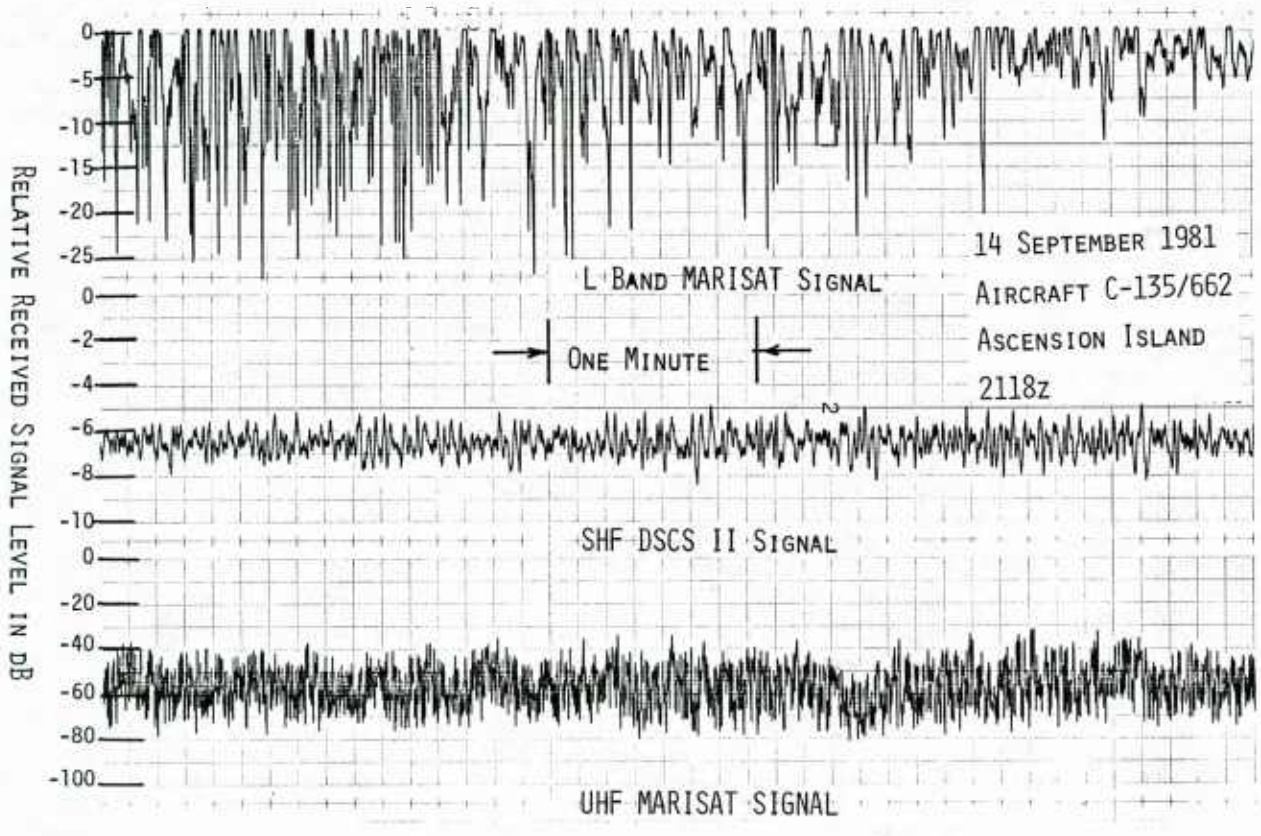
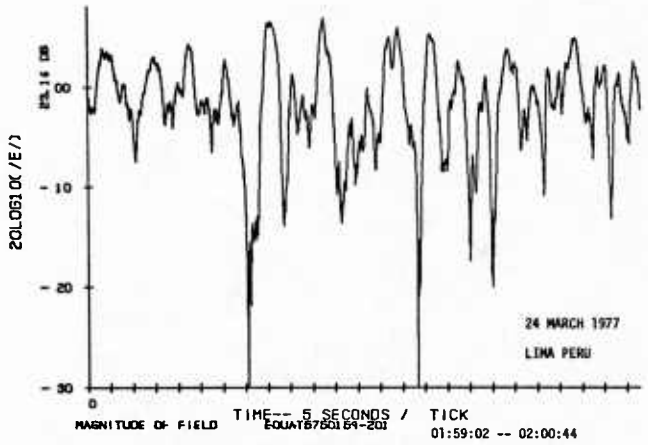
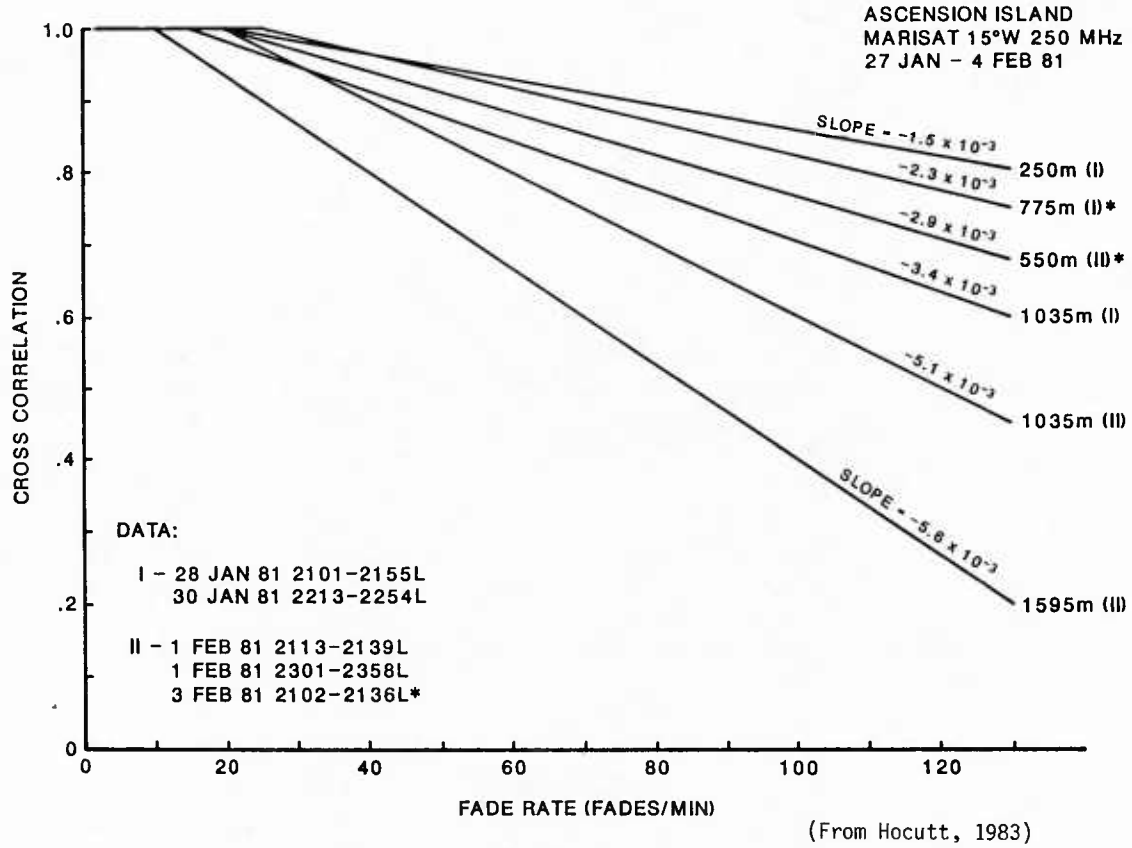


FIGURE 14. GIGAHERTZ IONOSPHERIC SCINTILLATION FADING



(From Johnson, 1979-1)

FIGURE 15. PHASE AND AMPLITUDE OF IONOSPHERIC SCINTILLATION FADING



(From Hocutt, 1983)

FIGURE 16. SLOPE OF CROSS-CORRELATION VERSUS FADE RATE

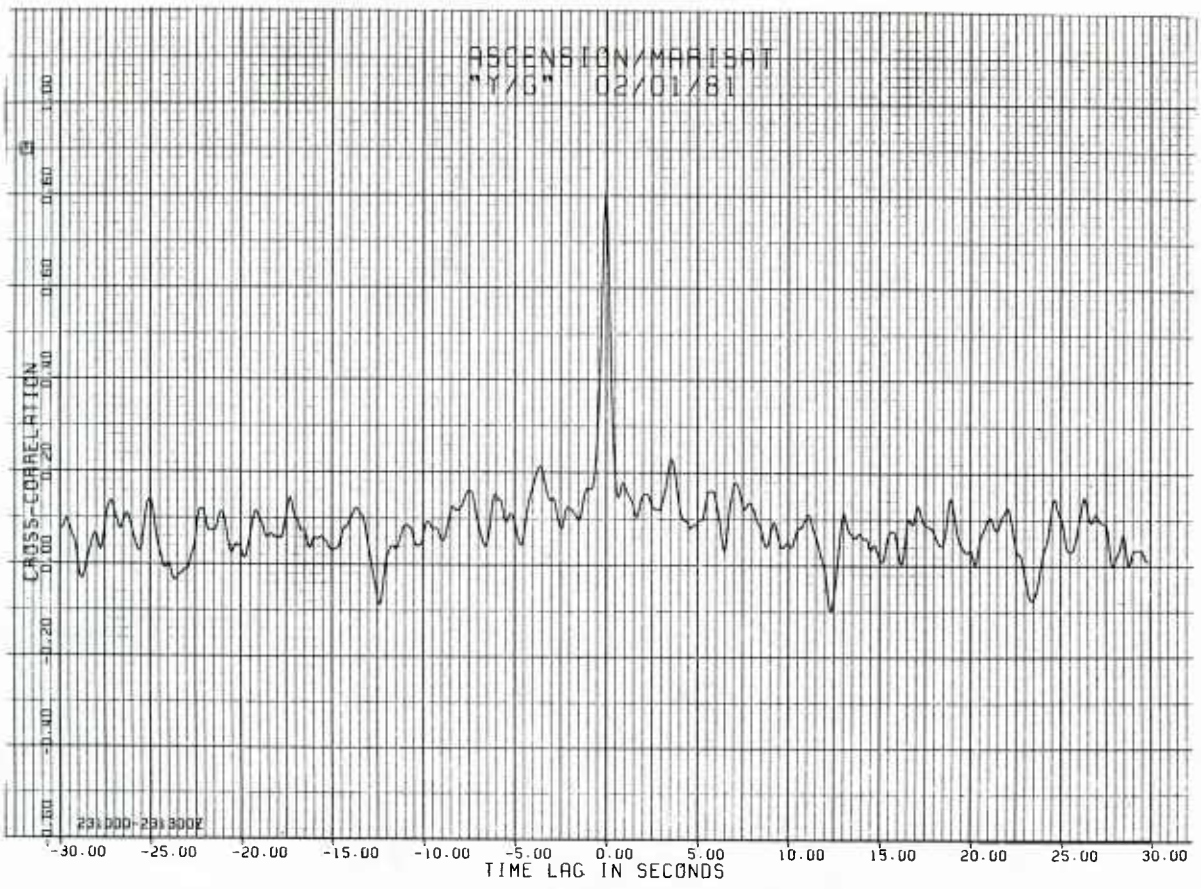
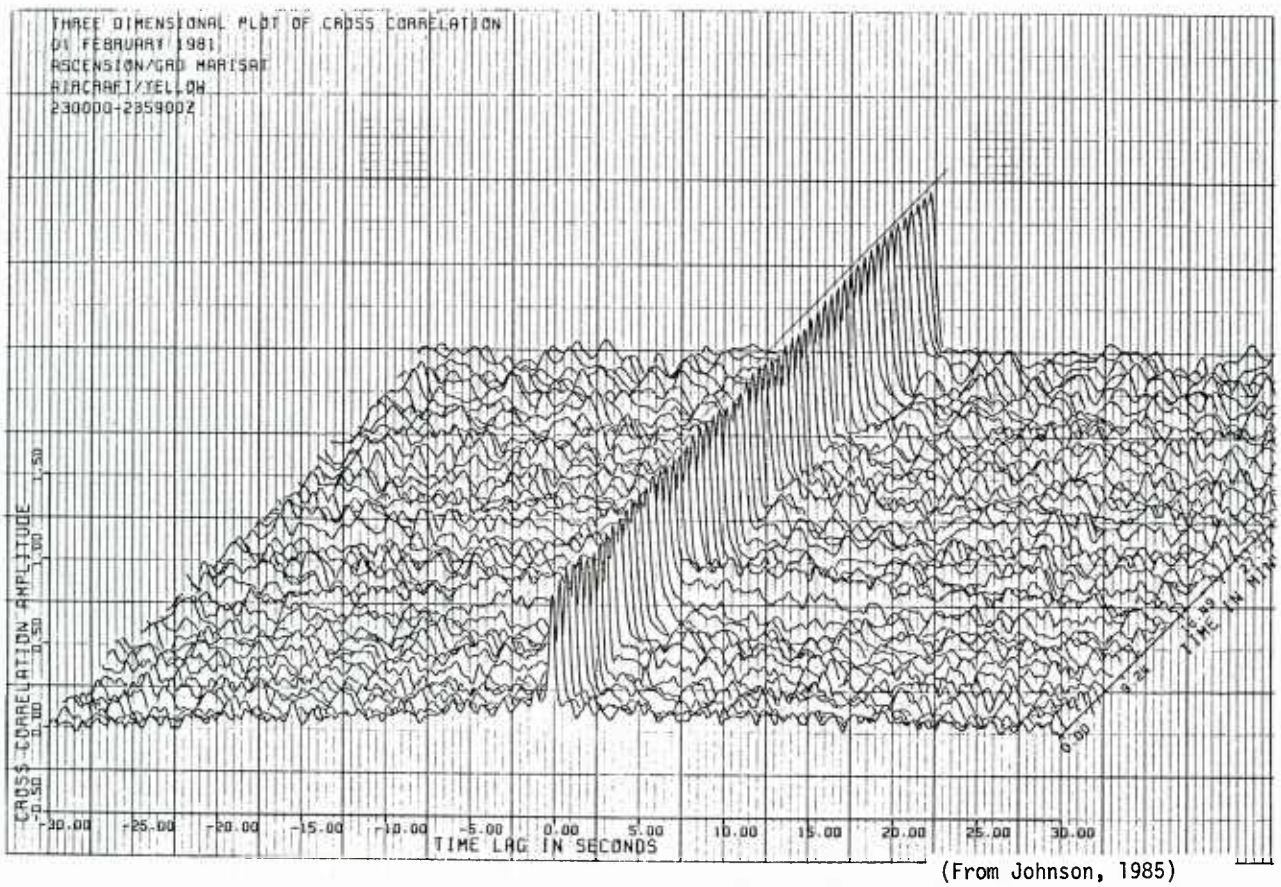


FIGURE 17. CROSS-CORRELATION TIME LAG FOR 1.6 KM RECEIVER SEPARATION



(From Johnson, 1985)

FIGURE 18. CROSS-CORRELATION PLOTS WITH VARIOUS TIME LAGS

In summary, the north-south correlation of the ionospheric scintillation fading over distances of up to 1.6 km was very high for a considerable portion of the test period at Ascension Island. The time that the fades occurred at the north-south spaced receivers varied from a zero time lag to ±0.2 seconds. This time lag corresponds to a tilt of the north-south irregularity alignment of approximately ±1°. It is postulated that the tilt is generated by irregularities at different altitudes aligning with the altitude-dependent magnetic field lines. Other possible explanations are short term changes in the magnetic field alignment due to electrojet currents or magnetic storms.

2. Equatorial Ionospheric Modification Experiments

a) STRESS

The overall objective of the STRESS (Satellite Transmission Effects Simulation Study) experiment was to evaluate the performance of a UHF satellite communications system in an artificially disturbed ionosphere in an effort to better predict the expected performance of such communications systems in a nuclear environment (extracted from Prettie et al, 1977). One objective of the experiment was to exercise the techniques used to verify assumptions made in predicting the performance of communications systems operating through striated plasmas. These techniques involve gradient drift plasma instability phenomenology for the determination of the striated environment, multiple thin phase screen propagation theory, and computer simulations of system performance that utilize propagation inputs. The other objective was to obtain data on late-time striation dissipation mechanisms. At the time of the experiment, no theories exist that describe how long striations from barium or nuclear detonations are expected to persist.

The basic concept of the STRESS Experiment involves at least two communications terminals, a striated plasma in the ionosphere, and a UHF satellite. In the experiment the two terminals attempt to communicate via the satellite with UHF signals between one terminal and the satellite transversing the striated plasma. The properties of the striated plasma perturb the UHF signals and, thereby stress the communications link.

The use of chemicals to modify or artificially disturb the ionosphere is a technique that has received extensive development over the past years. An artificial barium ion cloud was used to produce propagation path disturbances during the ARPA SECEDE Program, which involved radar propagation through the disturbed ionosphere.

The barium clouds used in the STRESS test were generated with the launch of 48 kilograms of barium chips to an altitude of approximately 185 kilometers. The barium was vaporized by a small thermite explosion. Action of the sun's ultra violet rays on the barium generated barium ions and free electrons. The barium that did not ionize formed spherical clouds (neutral clouds) which drifted according to the high altitude winds (30 to 100 meters per second generally away from the sun). The ionized barium also formed in spherical clouds initially but soon changed into elliptical clouds tilted along the magnetic field lines. The ionized plasma was confined, and its diffusion spread occurred only in the direction of the magnetic field lines. Figure 19, (McDaniel, 1972) illustrates the subsequent ion cloud evolution from two different views. The bottom row or sketches represents the more typical view of an ion cloud in the process of striating as it would appear from sites with arbitrary magnetic field line aspect angles. The top row of sketches show the corresponding appearances of the ion cloud when viewed up the field lines.

The typical cloud evolution from the elliptical form with the circular cross-section (labelled "AMBIPOLAR DIFFUSION") to a striated cloud is driven by the neutral wind attempting to drag the denser regions of the barium cloud with it (and thus with the neutral cloud) in conflict with the magnetic field confinement forces (E x B forces) on the entire ion cloud. (If the neutral cloud were shown in this figure, it would be seen moving the the left.) Initially the denser portion of the cloud is dragged to one side of the cloud forming the "hard edge," or "BACKSIDE STEEPENING." Further wind drag pulls "fingers," or "sheets," of dense plasma from the "hard edge" which eventually pinch off to form isolated "striations." When viewed with a typical magnetic aspect as in the bottom row, the appearance of isolated striations embedded in a background plasma cannot be distinguished from the appearance of the overlap of several sheets of varying thicknesses.

Both the sheets and striations cause UHF signal amplitude scintillations while the effect of the unstriated, or "smooth," ion cloud (farthest from the neutral cloud and to the right in the figure) is a slight phase shift due to the elevated integrated electron content through the medium. While the initial barium release occurs at approximately 185 kilometers altitude, the free electrons tend to drift up and down the magnetic field lines between altitudes of approximately 140 kilometers and 210 kilometers, Figure 20. The development of the striated line might appear as in Figure 21 when viewed across the field lines.

Barium clouds resemble weather clouds in that all of the significant observation light which comes from them is reflected sunlight; their glow due to molecular recombination is insignificant. Barium clouds launched at sunset are best observed after the time when the sun is 6° below the horizon and before sunset at the 185 kilometer altitude. The spherical neutral cloud reflects sunlight of a bluish and greenish tint. The ionized portion of the cloud reflects sunlight of a pinkish or reddish tint, Figure 22.

For Project STRESS the barium releases occurred at various times relative to the 6° sun depression angle. The barium clouds which were released early and passed through their early stage of development obscured by the sky glow became visible well into their development and remained visible further into their development than those launched later. The variety of launch times allowed optical observation of the late-time cloud development and will provide data on structure dissipation mechanisms.

The concept of using an aircraft to fly under the barium cloud projections from the satellite was intrinsic to the test concept. Using satellite ephemeris data and nominal cloud drift assumptions two test windows for operation of the aircraft with the satellite were generated before the test. These windows give a qualitative feel for the geometry involved. The flight path of the aircraft in the shadow of the cloud was designed primarily to cut across the striations and to measure the signal fading caused by the diffraction pattern of the striations. Some passes, "parallel runs or end runs," were made along the striations to measure their extent and to investigate propagation phenomena.

For the STRESS tests, several satellite test configurations were utilized. The first test configuration, Figure 23, provided UHF forward downlink data to the aircraft and a CW UHF uplink probe from the aircraft through the barium cloud. The uplink probe was sampled at the satellite and sent down on the EHF downlink to the Rooftop where it was recorded. The EHF forward uplink was provided either by the Rooftop or by the aircraft.

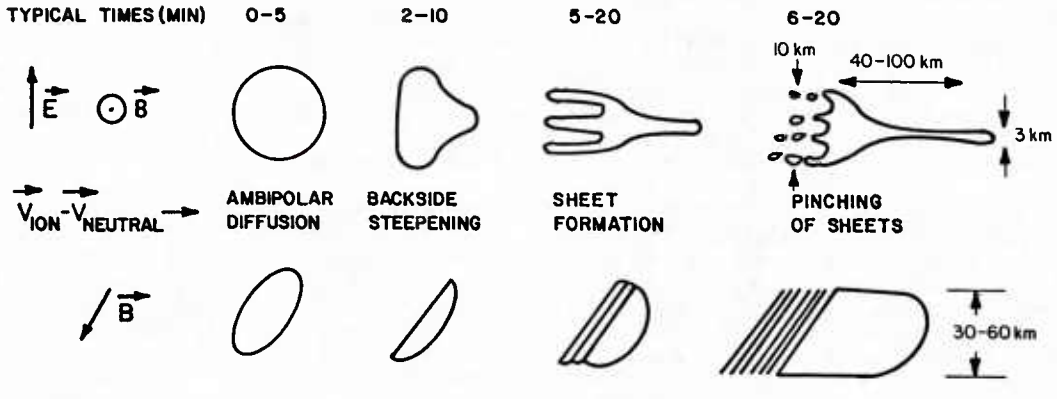
The fading from the Barium Cloud was characterized according to early-time like, Rayleigh-like, or Rician-like, and parallel pass fading. As the cloud developed into a series of individual irregularities, rapid and deep fading was produced often with a ringing type multipath caused by edge diffraction effects at the beginning or end of the pass. The downlink received signal level during some passes showed a broad decrease at the initial part of the pass with rapid Rayleigh-like fading toward the end. Excellent

examples of Rayleigh-like fading are seen in the uplink ESTHER, Pass 8, data, Figure 24. The diffraction edge in this pass is obvious on the left at the start of the pass. The phase shown in Figure 25 indicates vestiges of the smooth Gaussian early-time behavior. It is, however, corrugated both by changes in path integration electron content and by diffraction effects that can produce phase jumps associated with deep fades.

During the five STRESS tests a wide variety of ionospheric scintillation fading was encountered. Fades during the early development of the cloud usually consisted of a single, long defocus, roughly 10 dB deep with a single enhancement (focus) prior to and following the defocus. The mid-time fading was Rayleigh-like, usually with rapid fades from 15 to 30 dB deep with enhancement of 5 to 10 dB. During the late-time the fading was often patchy and Rician-like, with slower fading.

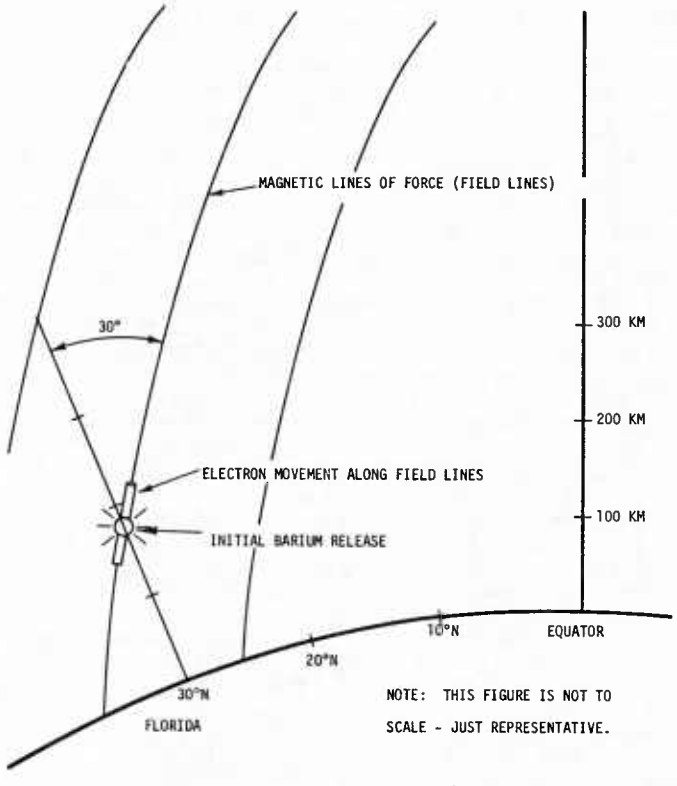
b) Coloured Bubbles

The Coloured Bubbles experiment was designed by Haerendel of the Max Planck Institute For Physics and Astro Physics to investigate the trigger mechanism for equatorial ionospheric irregularities (extracted from Johnson & Hocutt, 1984). If two barium packages were detonated in the ionosphere where the ion density gradient is steepest, Haerendel & Valenzuela (1982) postulated that the following would occur: As the barium ionizes, it enhances the plasma content of the ionospheric flux tubes. These flux tubes are forced down with respect to the ambient undisturbed plasma. The incompressible nature of this convection



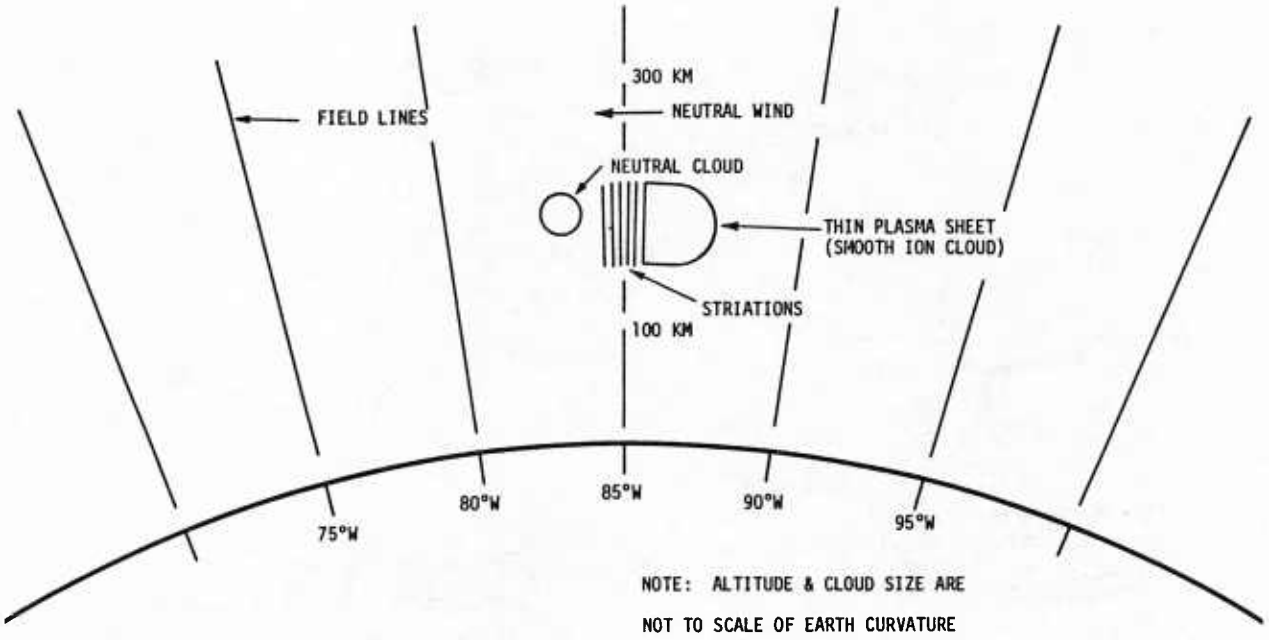
(From Prettie et al, 1977)

FIGURE 19. SCHEMATIC DIAGRAM OF BARIUM ION CLOUD MORPHOLOGICAL DEVELOPMENT



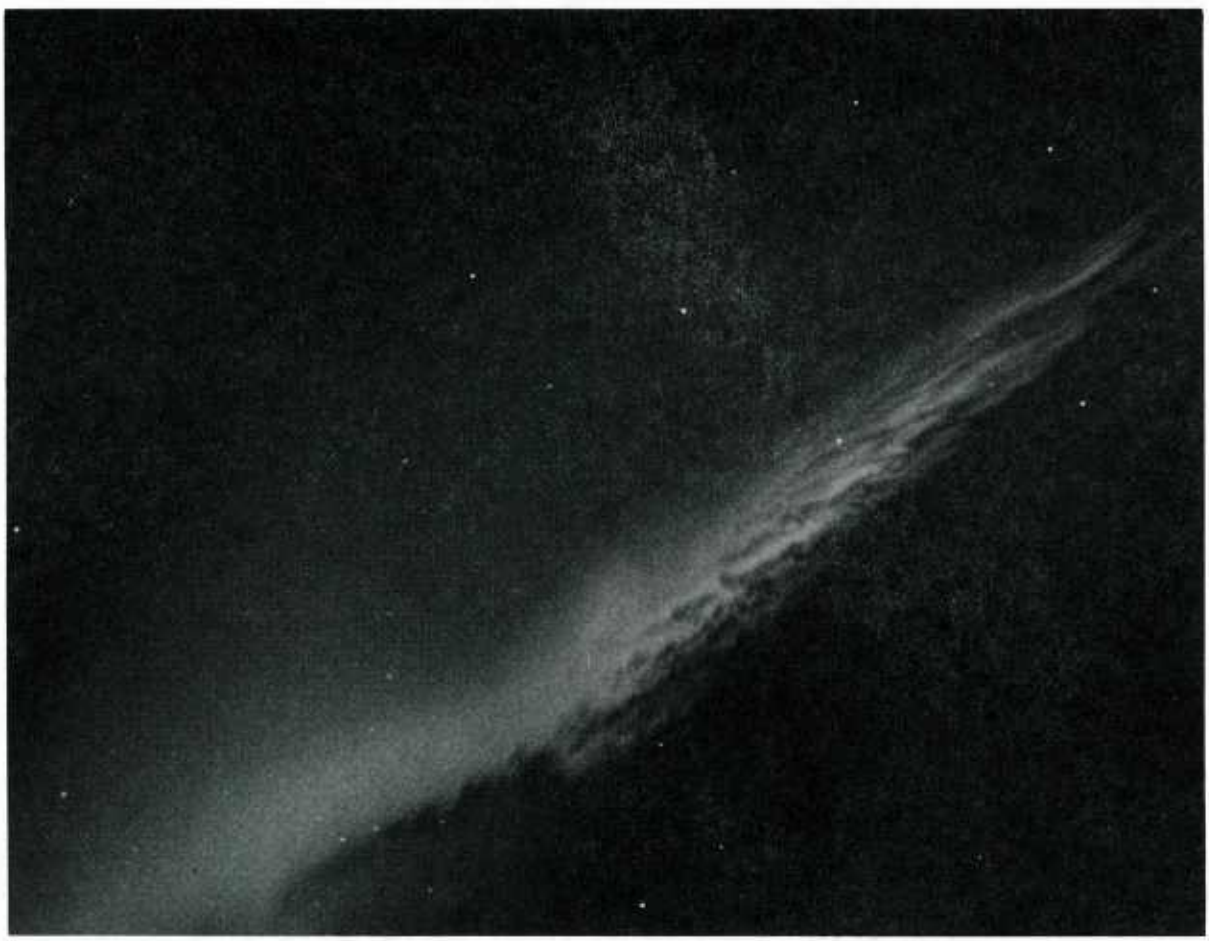
(From Prettie et al, 1977)

FIGURE 20. PROPAGATION OF FREE ELECTRONS ALONG FIELD LINES



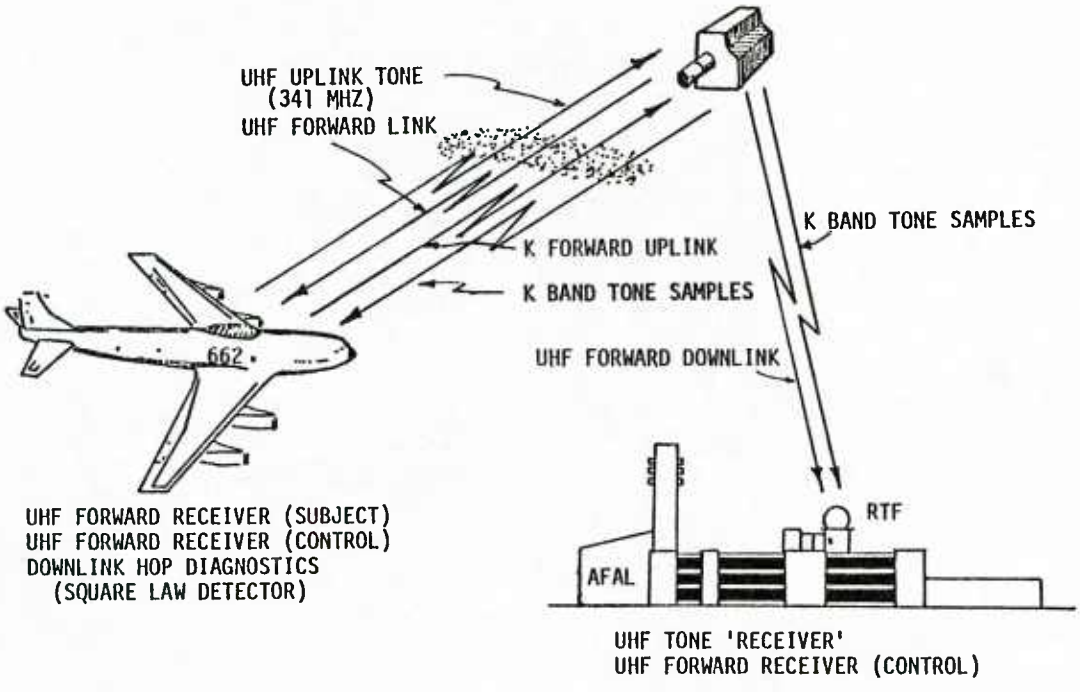
(From Prettie et al, 1977)

FIGURE 21. MOVEMENT OF BARIUM ION CLOUD



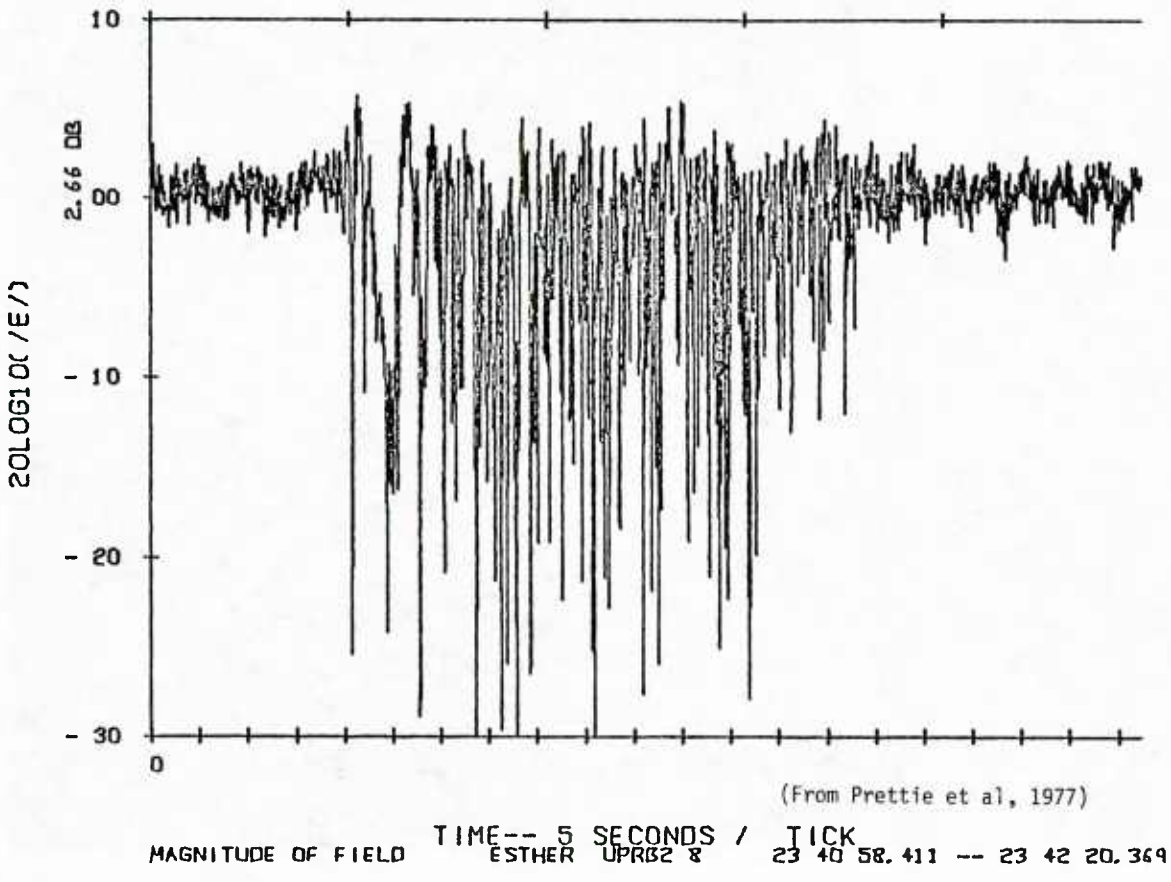
(From Linson, 1980)

FIGURE 22. PHOTO OF BARIUM ION CLOUD



(From Prettie et al, 1977)

FIGURE 23. STRESS TEST CONFIGURATION



(From Prettie et al, 1977)

FIGURE 24. RAYLEIGH LIKE UPLINK FADING FROM BARIUM CLOUD

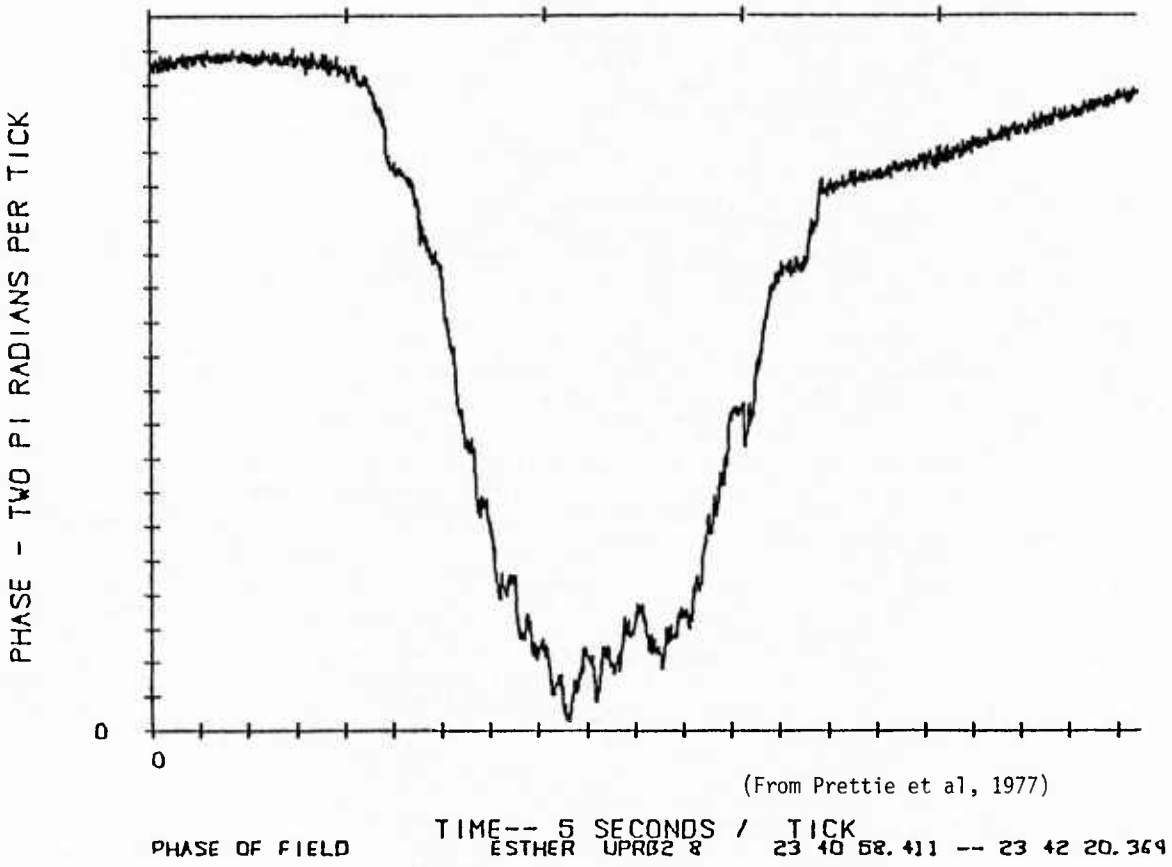


FIGURE 25. UPLINK PHASE EFFECTS FROM BARIUM CLOUD

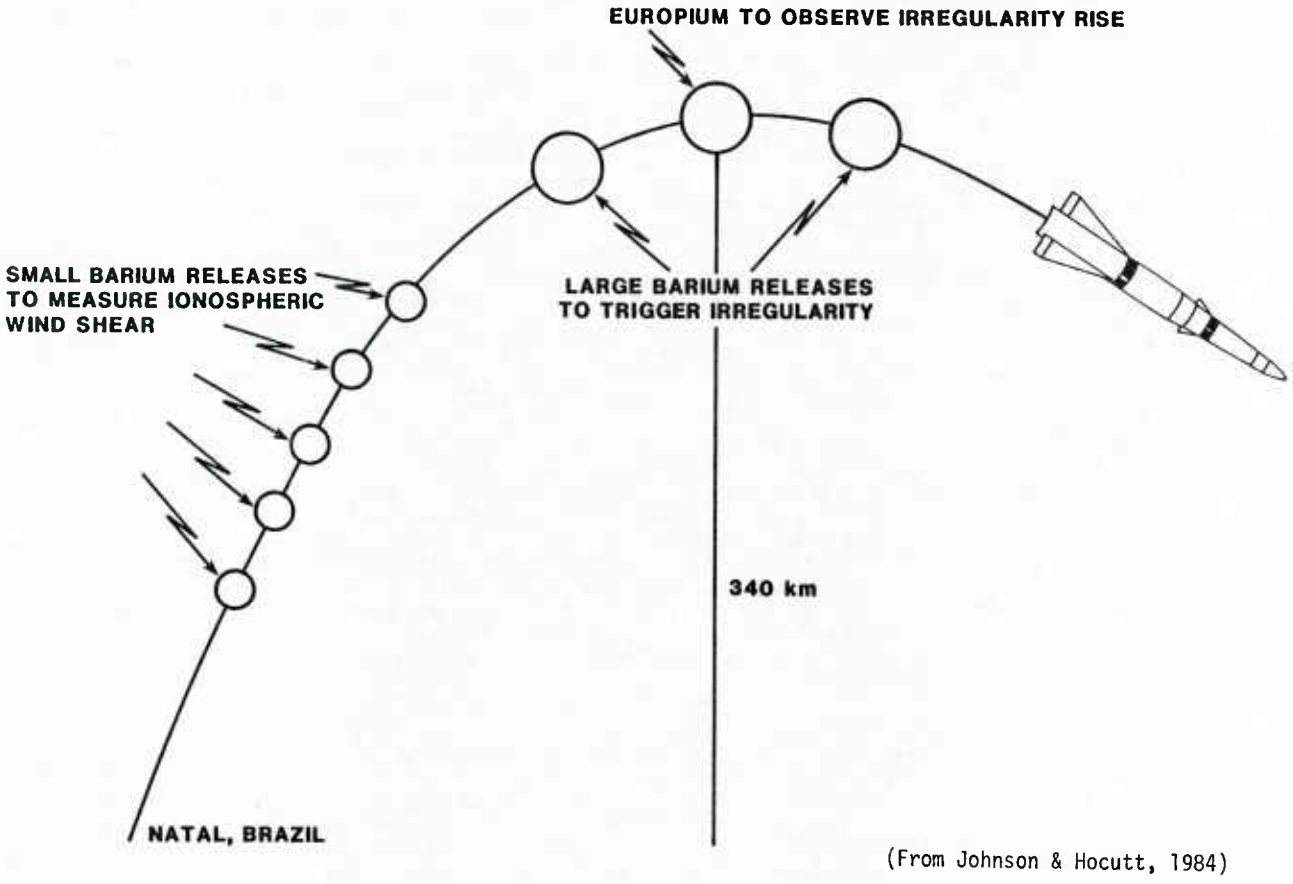


FIGURE 26. COLOURED BUBBLES EXPERIMENT GEOMETRY

motion, along with the electric field established between the two barium clouds, forces the plasma region between the barium upward. The upward motion of the low density plasma will grow exponentially if the barium is injected into the steep slope of the post-sunset F region density, near an altitude of 350 km. Strong plasma density gradients will be set up by this rising low density region. These regions or bubbles will have secondary instabilities which form small scale (1 meter to 10 kilometer) irregularities. The structural development of these small scale irregularities over time is the property of interest.

The Coloured Bubbles experiment consisted of eight chemical packages launched on a Sonda III rocket from Natal, Brazil, Figure 26. The first five chemical packages were small barium charges released to investigate wind shear in the 220 to 300 km region of the ionosphere. The next three packages were part of the ionospheric modification experiment. Two 40 kg barium packages were detonated 2 minutes apart at an altitude of 320 km where the ion density gradient was the steepest. Between them, a europium package was exploded. Observers used the europium to monitor the rise of the bubble of irregularities. The sequence of releases is shown in Figure 27, along with the ionization and elongation of the two barium clouds along the magnetic lines.

The Avionics Laboratory (AFWAL) equipped C-135/372 aircraft was flown in the shadow of the barium clouds, receiving a UHF signal from the FLTSATCOM satellite at 23° West longitude. Aircraft instrumentation recorded the phase and amplitude of the UHF CW signal while the aircraft line-of-sight to the satellite swept in and out of the area where the effects of the barium-induced irregularities were expected.

On 17 September 1982, the first Coloured Bubbles rocket was launched. The aircraft made five passes below the area of the barium releases and the irregularities that subsequently formed. It was apparent there was a single region of irregularities forming between the two barium clouds and the barium clouds themselves were not producing irregularity signatures.

By measuring the position and time when the aircraft encountered the shadows of the irregularity region's edge, the velocities of the eastern and western edges of the region and its width were calculated. The width of the region was 5 km when first encountered 10 minutes after release. It grew in a nearly linear fashion to 30 km at 30 minutes after release and 40 km approximately 40 minutes after release. Since definitive data on the vertical motion of the ambient ionosphere in the region of the irregularities was not available (Buchau, 1983), a zero vertical velocity was used in calculating the width and horizontal velocity of the irregularities. Any vertical velocity or vertical growth of the irregularities would affect the apparent horizontal velocity and width. The effect on the calculated values is expected to be small since the aircraft was viewing the irregularities from almost directly below.

On 18 September 1982, the second Coloured Bubbles rocket was launched. The aircraft flew a track similar to Coloured Bubbles #1. However, since the results of the Coloured Bubbles #1 had clearly shown the irregularities were located between the two major barium releases, the aircraft flight track for Coloured Bubbles #2 was altered to investigate only the region below the irregularities. The aircraft turned to reenter the shadow of the irregularities as soon as the aircraft had exited it. This gave more samples of the bubble during its early growth.

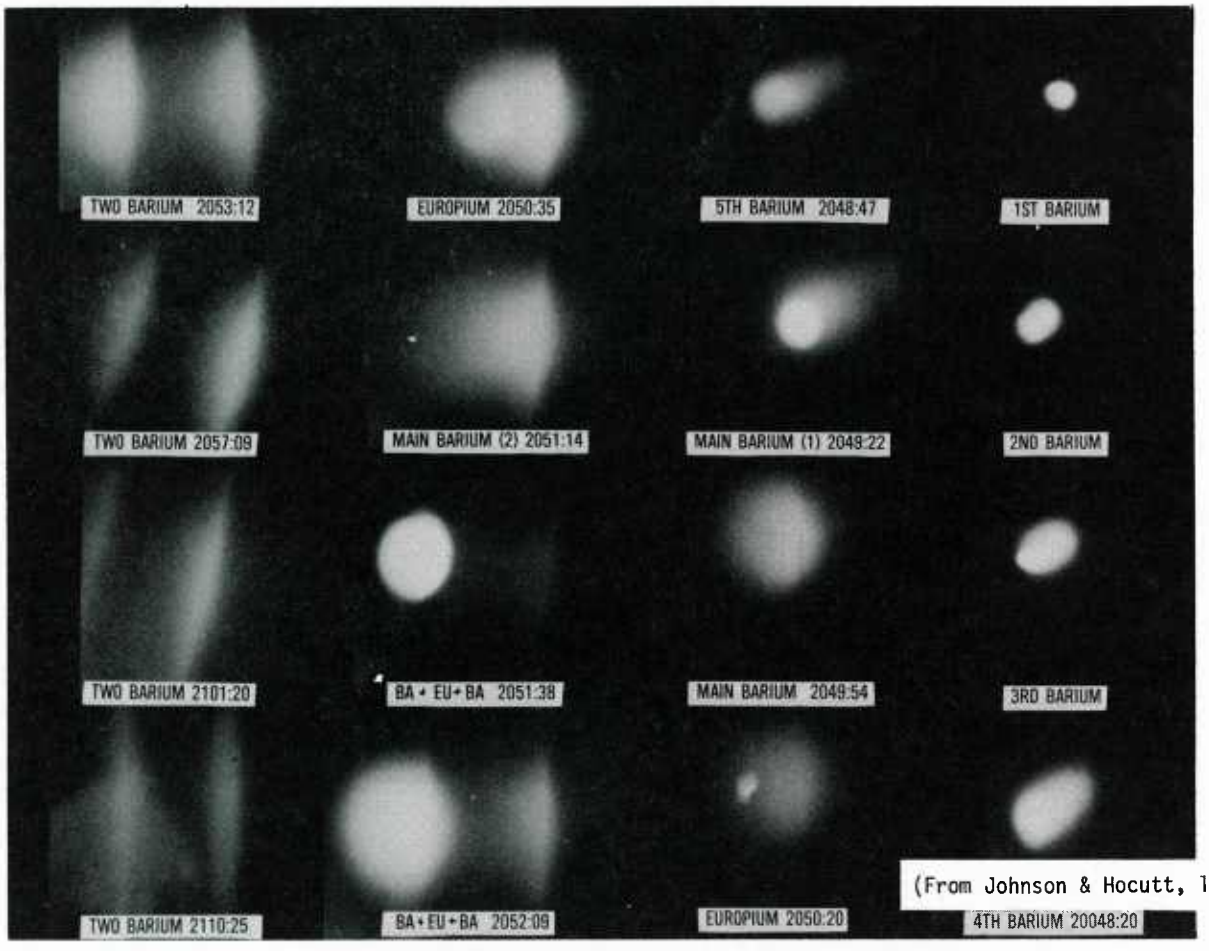
The aircraft was able to fly eight passes under the irregularities in Coloured Bubbles #2. Six minutes after the release, a 10 second long 1 dB disturbance was seen. Approximately 15 minutes after release, the irregularities caused 5 dB peak-to-peak fading, Figure 28. The irregularities continued to grow in width and intensity as the aircraft passed back and forth beneath it. Forty minutes after release, the fading on the FLTSATCOM signal was greater than 10 dB peak-to-peak. The fading was 25 dB peak-to-peak fifty-five minutes after release.

The velocities of the eastern and western edges of the irregularity region are shown in Figure 29, as well as the growth in the region's width. The velocity of the eastern edge was 30 m/s to the east about eight minutes after the second major barium was released. Twenty five minutes after the release, the velocity was 80 m/s. It seemed to level off to about 100 m/s to the east. On the other hand, the western edge of the irregularities moved westward first. This implies the region of irregularity was expanding in size faster than the entire region was drifting east. The western edge then began moving east with a fairly constant velocity of 45 m/s. About 45 minutes after the barium release, the western edge velocity increased to greater than 80 m/s. The second part of Figure 29 shows the irregularity region's width. Five minutes after the second barium release, the irregularities were 2.8 km wide. In 50 minutes, it grew to about 80 km. After that, growth slowed and the width appeared to reach a maximum of approximately 100 km. As with Coloured Bubbles #1, a zero vertical velocity was assumed in the width and velocity calculations for Coloured Bubbles #2.

The test results from Coloured Bubbles #1 clearly showed that a low density ion bubble was triggered in the unstable region between the two main barium releases. During the Coloured Bubbles #2 test, the aircraft made eight passes through the shadow of the bubble during the sixty minutes following the release, collecting data which showed the irregularities grew from a seed to a hundred kilometer width and caused 25 dB peak-to-peak fading. The fully developed bubble produced amplitude scintillation with fade depths and fade rates identical to that encountered with natural equatorial irregularities for over two hours. The continued growth of the irregularity region's width exhibited in Coloured Bubbles differed dramatically from the irregularities behavior exhibited by the STRESS and PLACES barium ionospheric modification experiments. Only the leading edges of the STRESS and PLACES barium clouds striated, growing to a width of approximately 20 kilometers in 15 to 30 minutes and remaining at that width for over two hours. Coloured Bubbles differed because the electric field between the two barium clouds triggered a natural depletion rather than directly forming the strong irregularity gradient at the leading edge of each of the barium clouds. The successful triggering of a bubble helps explain the formation process of the natural ionospheric irregularities. However, it leaves unanswered the question of what natural forces trigger equatorial ionospheric irregularities.

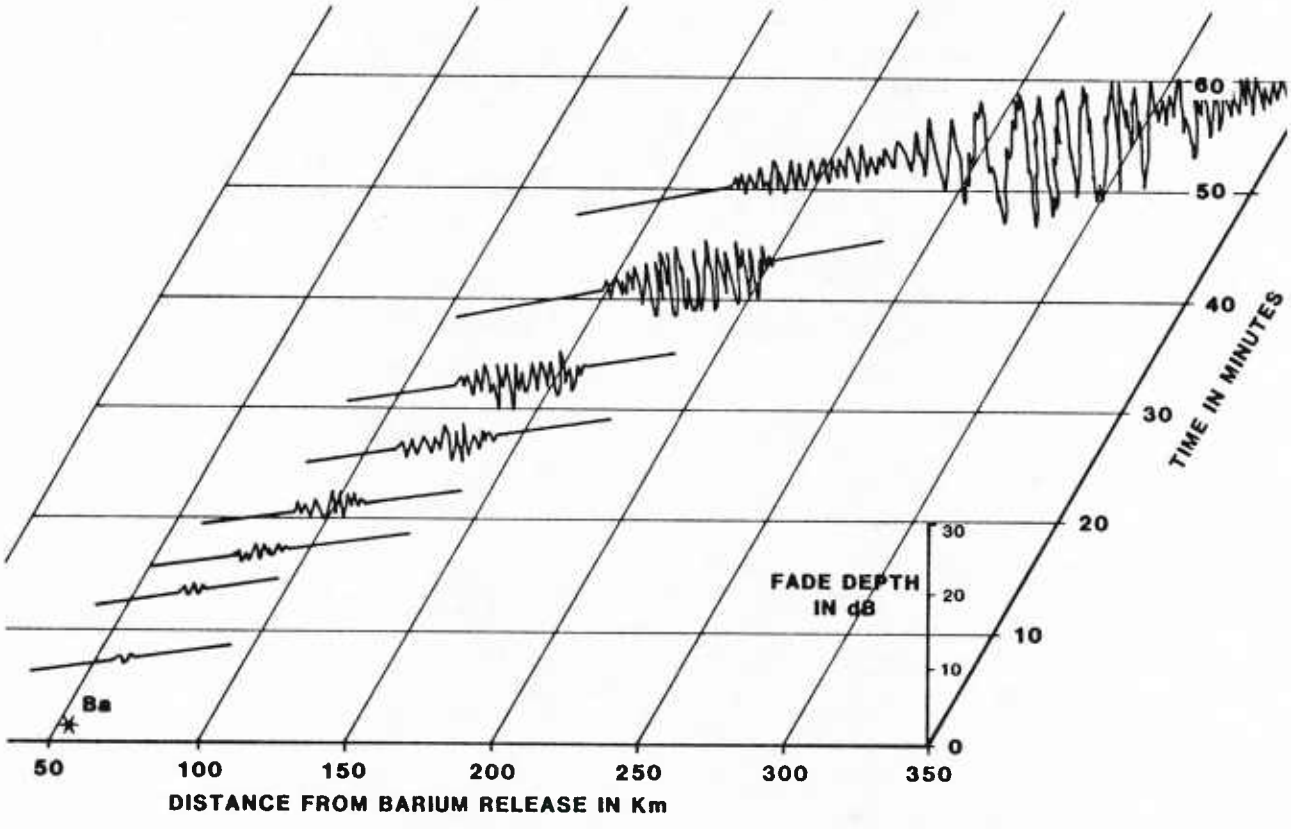
3. Mid-Latitude ionospheric Scintillation

The occurrence of ionospheric scintillation in the mid-latitude region is relatively rare, (Johnson, 1979-1). Occasional fades of several dB peak-to-peak at 300 MHz are observed, (Mass & Houminer, 1969). Fades caused by a single ionospheric irregularity also occasionally occur, as reported by Slack, 1968. These irregularities cause a ringing type defraction pattern that Slack refers to as "Quasi periodic amplitude scintillation," Figure 30. However, since the occurrence of scintillation is rare and the amplitude is small, mid-latitude ionospheric scintillation is not expected to cause any significant disturbance to an earth-satellite communications link.



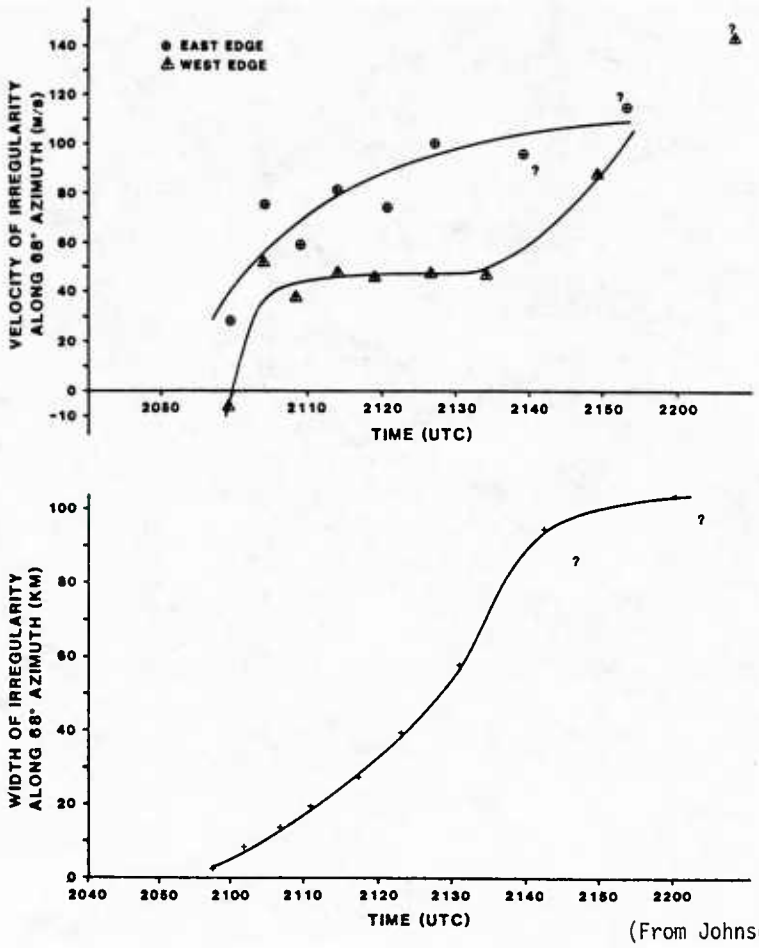
(From Johnson & Hocutt, 1984)

FIGURE 27. COLOURED BUBBLES RELEASE SEQUENCE



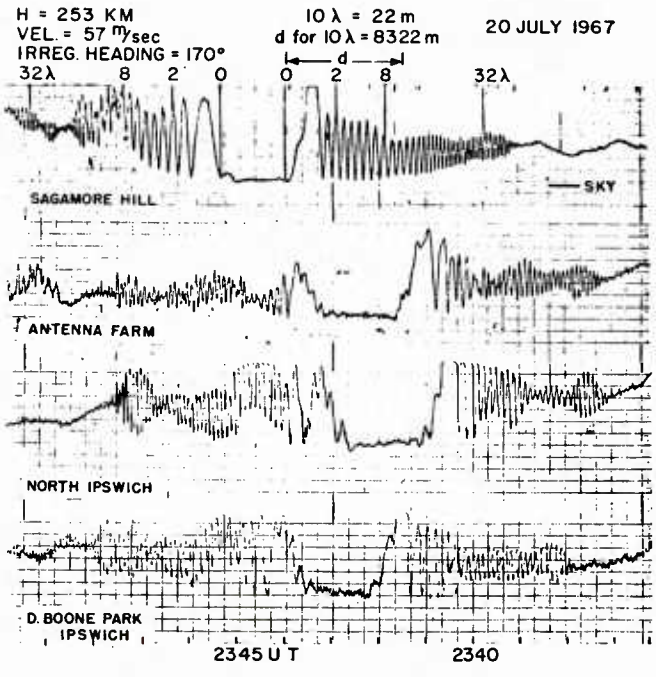
(From Johnson & Hocutt, 1984)

FIGURE 28. COMPOSITE FADING FOR COLOURED BUBBLES EXPERIMENT



(From Johnson & Hocutt, 1984)

FIGURE 29. IRREGULARITY VELOCITY AND WIDTH FOR COLOURED BUBBLES EXPERIMENT



(From Slack, 1968)

FIGURE 30. EXAMPLES OF RINGING IRREGULARITIES

4. High-Latitude Ionospheric Scintillation

The high latitude region has two distinct scintillation areas: the auroral area where visible aurora occurs in the D and E region, Figure 31; and the polar cap area from the upper boundary of the auroral area to the magnetic pole, (extracted from Johnson 1980).

The auroral region tends to be strongly affected by magnetic activity, time of day, and solar flux, while only moderately controlled by season. The polar cap tends to be strongly affected by solar flux, moderately affected season, and very slightly affected by magnetic activity and time of day.

The frequency dependence of the fading amplitude in both the auroral and polar cap is very similar to the equatorial region. There is slightly less fading experienced at the higher frequencies from 1 GHz and above, but the fading amplitude at VHF and UHF frequencies is very similar to the equatorial scintillation previously depicted in Figure 13. The rate of fading covers the same wide range experienced in the equatorial region, Figure 32.

The irregularity formation in the polar region, as in the equatorial region, tends to be elongated along the magnetic field lines. Elongations of five to ten times the irregularity width have been postulated from measurements (Aarons et al, 1980-2). The fact that these irregularities are at steep angles to the horizon make the observations very geometry sensitive. Measurements by Reno, et al (1978) indicate that observations which are L-shell aligned experience more scintillation than non-aligned observations.

Scintillation tests made during a high solar flux density period (S_f = 250), indicate that scintillation activity in the polar cap may be more strongly dependent on solar flux density than equatorial or auroral scintillation. Polar cap flight during January 1979 resulted in scintillation of 5 dB or more being experienced 96% of the flight time. Scintillations greater than 20 dB amplitude were experienced about 60% of the flight time. Fading rates varied from 15 to 30 fades per minute. The scintillation experienced appeared to be patchy with strong scintillation occurring for fifteen to thirty minutes, followed by brief, quiet period.

UHF phase scintillations were also measured in the polar region. During the 1979 tests, phase scintillations with variations up to 10 pi radians (1800°) per second were recorded during the extensive amplitude scintillation, Figure 33.

In the polar region ionospheric irregularity drift is much more complex than the almost linear irregularity drift exhibited in the equatorial region, (Johnson, 1985). Polar all-sky photometer 6300 Å images usually showed sun aligned polar arcs of fade-producing irregularities (Johnson et al, 1981). The arcs often drifted from dawn-to-dusk while individual irregularities are moving rapidly in a perpendicular, noon-to-midnight direction. These irregularities appear to be 100 to 300 kilometers in diameter. Measurements of the ionospheric F layer height using an airborne HF ionosonde showed the F layer undisturbed height to be four or five hundred kilometers prior to the onset of fading (Buchau, 1983). The F region height dropped to a few hundred kilometers during the scintillation period and then rose again after the irregularity had passed.

To evaluate polar irregularity drift, an arrangement of three UHF antennas placed in a triangle was used at Thule, Greenland; Sondrestrom, Greenland; and Goosebay, Labrador. By recording the signal amplitude received from a UHF polar satellite, individual irregularity fades could be identified at all three antennas at different times, Figure 34. By correlating the relative delay in the arrival of a particular fade at each antenna, one could calculate the direction and speed of the irregularities. Correlation of the data from the three antennas showed the pattern of irregularity flow repeated every 24 hours. Velocities from a few meters-per-second up to a thousand meters-per-second were measured. Typically, the data showed strong correlation over distances of 100 meters and time delays of a few seconds.

Occasionally, during extremely severe fading conditions, poor correlation was exhibited over short distances and times. This suggests most of the time the irregularities are relatively stable as they move through the ionosphere, but there are periods when the ionosphere is extremely dynamic with the irregularity changing shape in a few seconds. A 3-dimensional plot of correlation was made using sixty individual correlation plots taken at one minute intervals, Figure 35. This plot shows consistently strong correlation with no change in delay, indicating a stable irregularity velocity. An abrupt change in drift direction is apparent in Figure 36. The first few curves show a time lag for the arrival of the fade at two selected antennas of plus one second. A quarter of an hour later the correlation decreased and appeared erratic. Then the correlation again became strong but with a minus three-second time lag. The time lag can be seen more clearly in Figure 37. This data indicates a reversal of drift direction occurred in a ten-minute period.

The general movement of ionospheric irregularities on the polar cap is postulated to follow a two-cell model with irregularities flowing from the noon-time auroral oval across the pole to the midnight sector and then rotating back around the auroral oval to the noon sector (Heelis et al, 1976). Heppner, 1977, has postulated a slightly modified two-cell flow to include the disturbance produced by the rotation of the earth. Using a technique described by Whalen, 1970, the three antenna measurement drift velocities are plotted on Heppner's polar drift model, Figure 38.

During July 1979 another series of polar cap flights was conducted. Three eight-hour flights into the polar cap during July resulted in zero scintillation observations at 250 MHz. The solar flux density during July was at a temporary minimum of 150. These observations tend to support the theory that solar flux density plays a prominent role in polar cap irregularity formation.

5. Nuclear Ionospheric Scintillation

Increased ionization and ionospheric striations from high-altitude nuclear bursts produce signal absorption and scintillations that decrease in rapidity with increasing frequency, but remain potentially significant even at EHF (extracted from Frazier & Caster, 1983). Significant reduction of signal blackout effects can be achieved by using higher frequencies such as EHF. However, EHF links can be degraded by nuclear scintillation effects from high-altitude detonations as well as by attenuation and signal scintillation from dust that is lofted into the atmosphere after a low altitude detonation. Signal scintillation can be defined as signal fluctuations in both phase and amplitude that result from the interaction of a structured or striated ionospheric plasma region with the radiowave of a satellite link.

Predictions of the extent and duration of EHF scintillation are much more sensitive to uncertainties in nuclear phenomenology than predictions at lower frequencies such as UHF. The likelihood of encountering severe ionospheric scintillation is greatly reduced at EHF. However, based upon a reasonable worst case nuclear environment, an EHF SATCOM link could be subject to a scintillation region that may extend a few hundred kilometers. Scintillation disturbances are likely to persist up to an hour after high altitude burst, Figure 39.

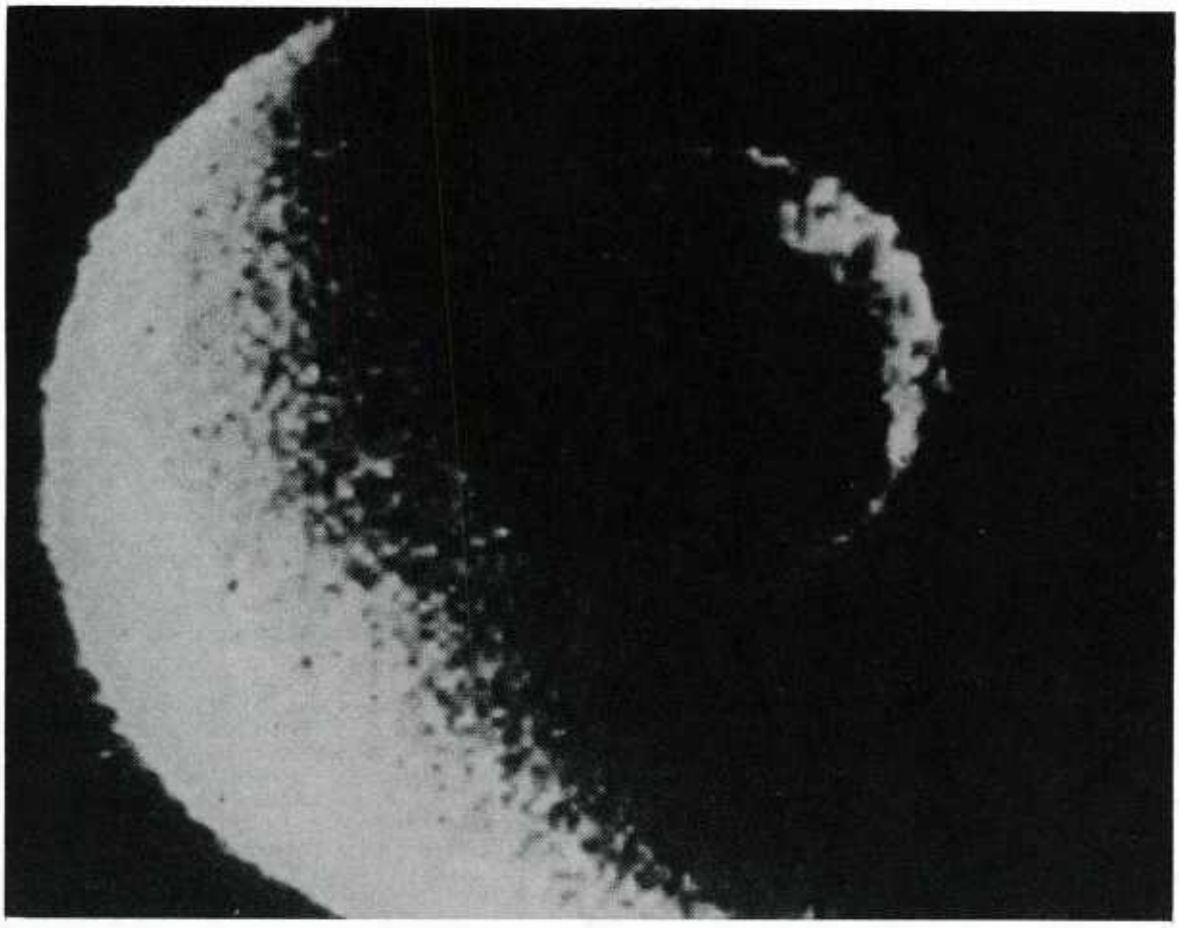


FIGURE 31. VISIBLE AURORAL OVAL

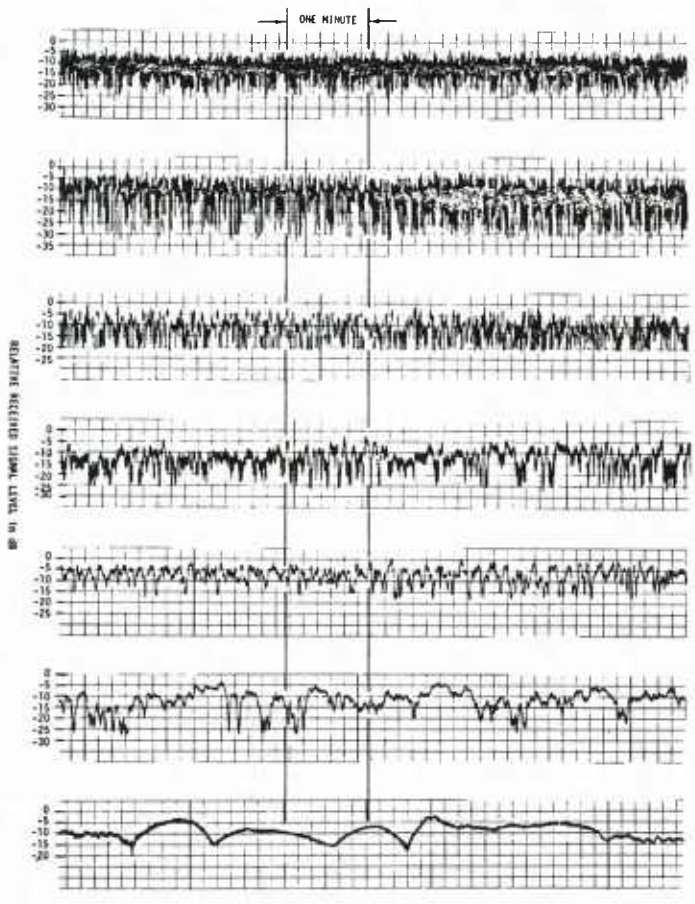
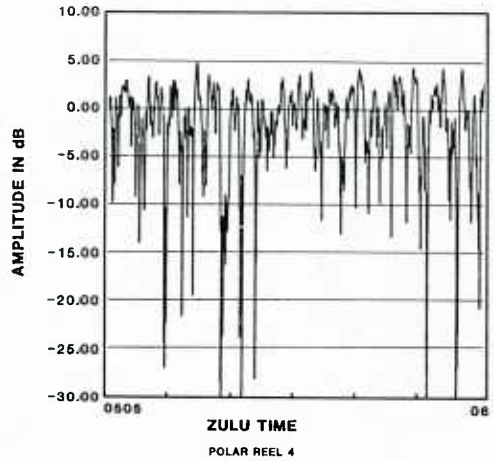
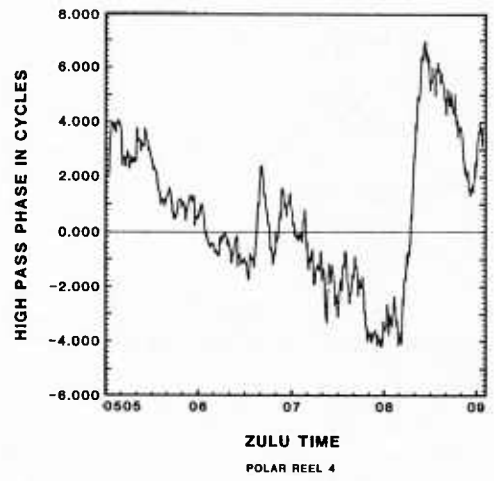


FIGURE 32. VARIATIONS IN HIGH-LATITUDE SCINTILLATION FADING RATE



AMPLITUDE OF SDS UHF SIGNAL
OBSERVED DURING STRONG FADING IN THE
26 JANUARY POLAR FLIGHT TEST



PHASE OF SDS UHF SIGNAL
OBSERVED DURING STRONG FADING IN THE
26 JANUARY 1979 POLAR FLIGHT TEST

FIGURE 33. PHASE AND AMPLITUDE OF POLAR UHF SCINTILLATION FADING

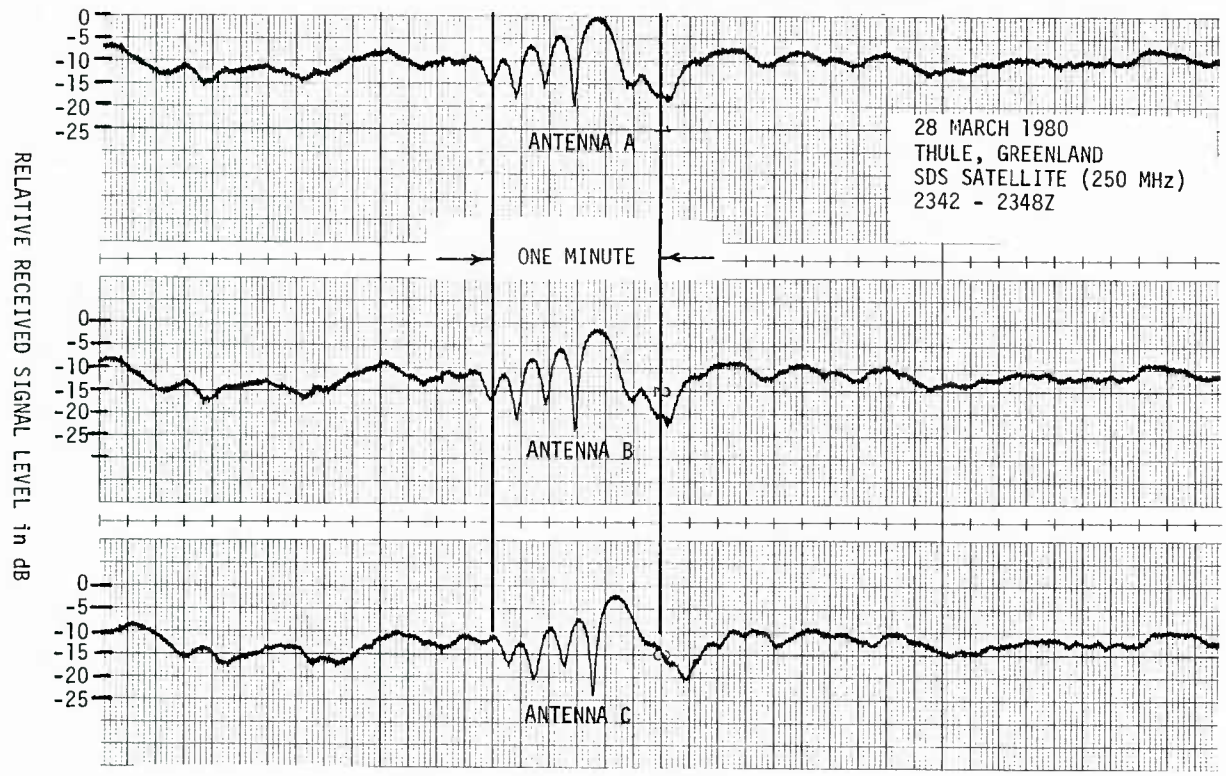
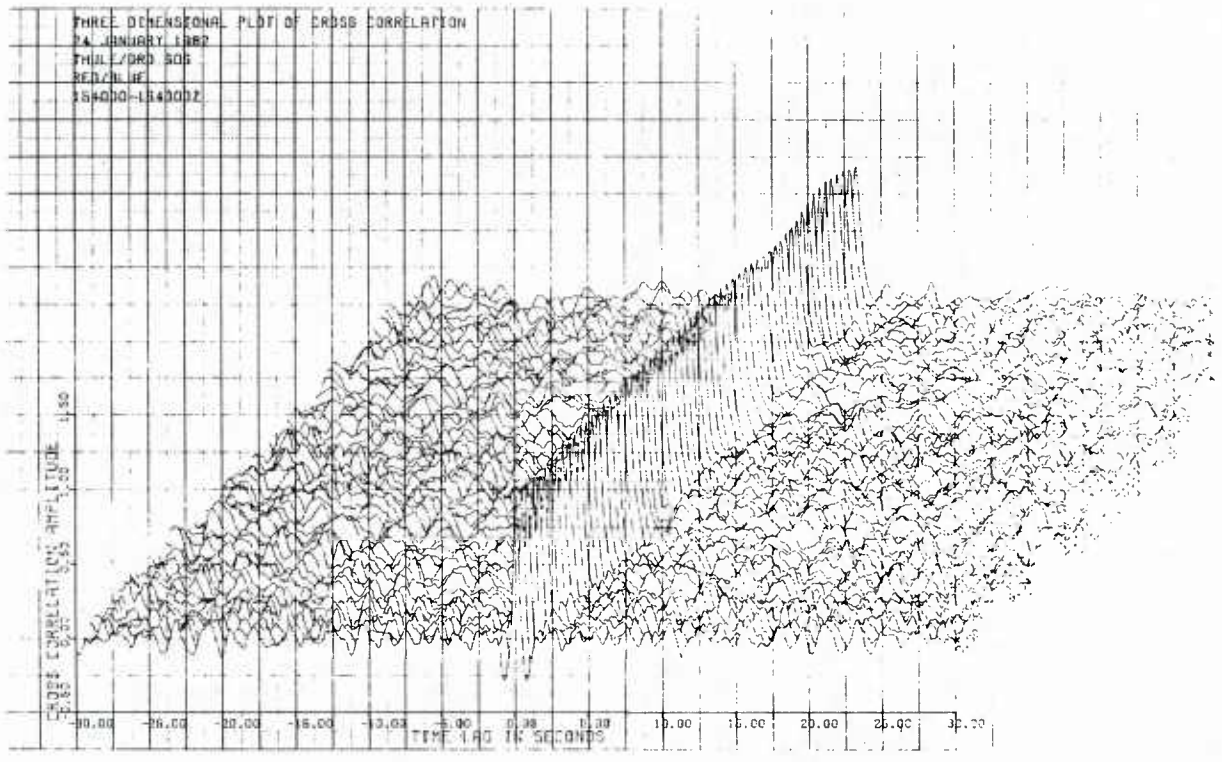
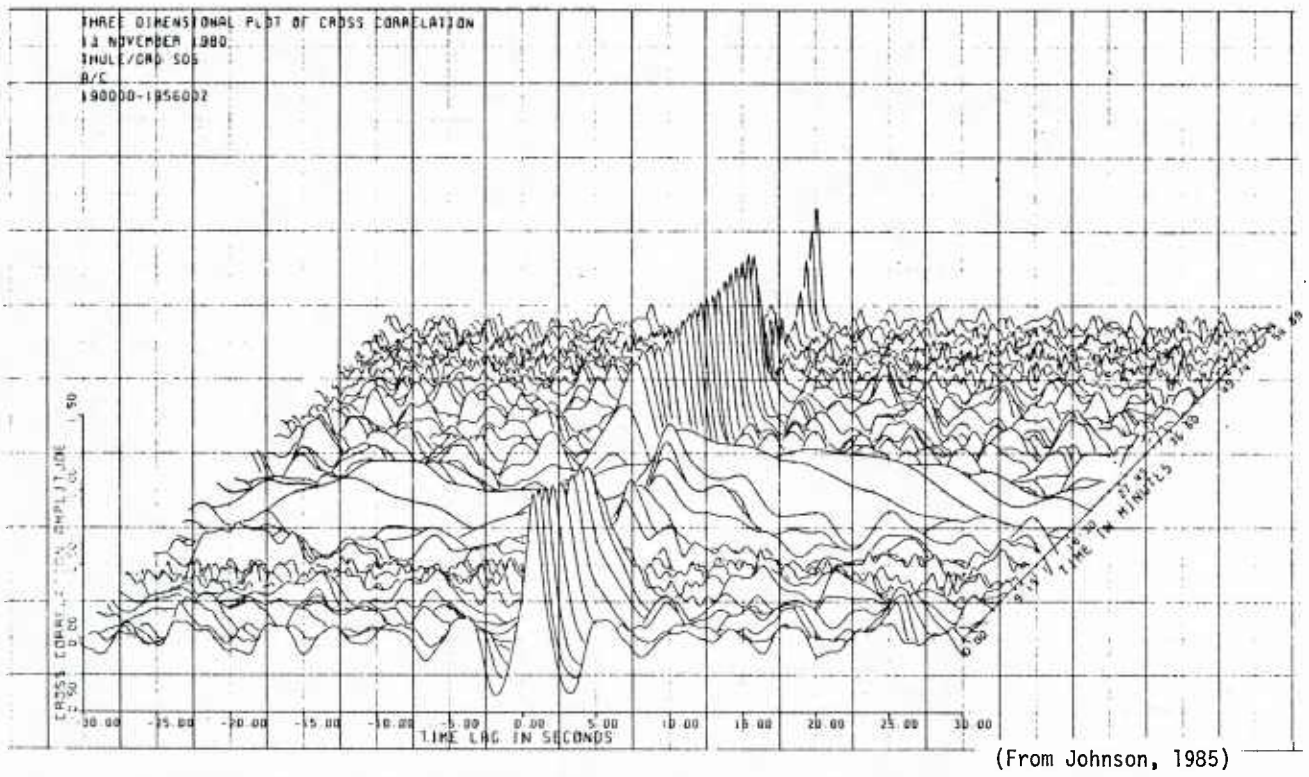


FIGURE 34. CORRELATION OF POLAR IONOSPHERIC SCINTILLATION FADING



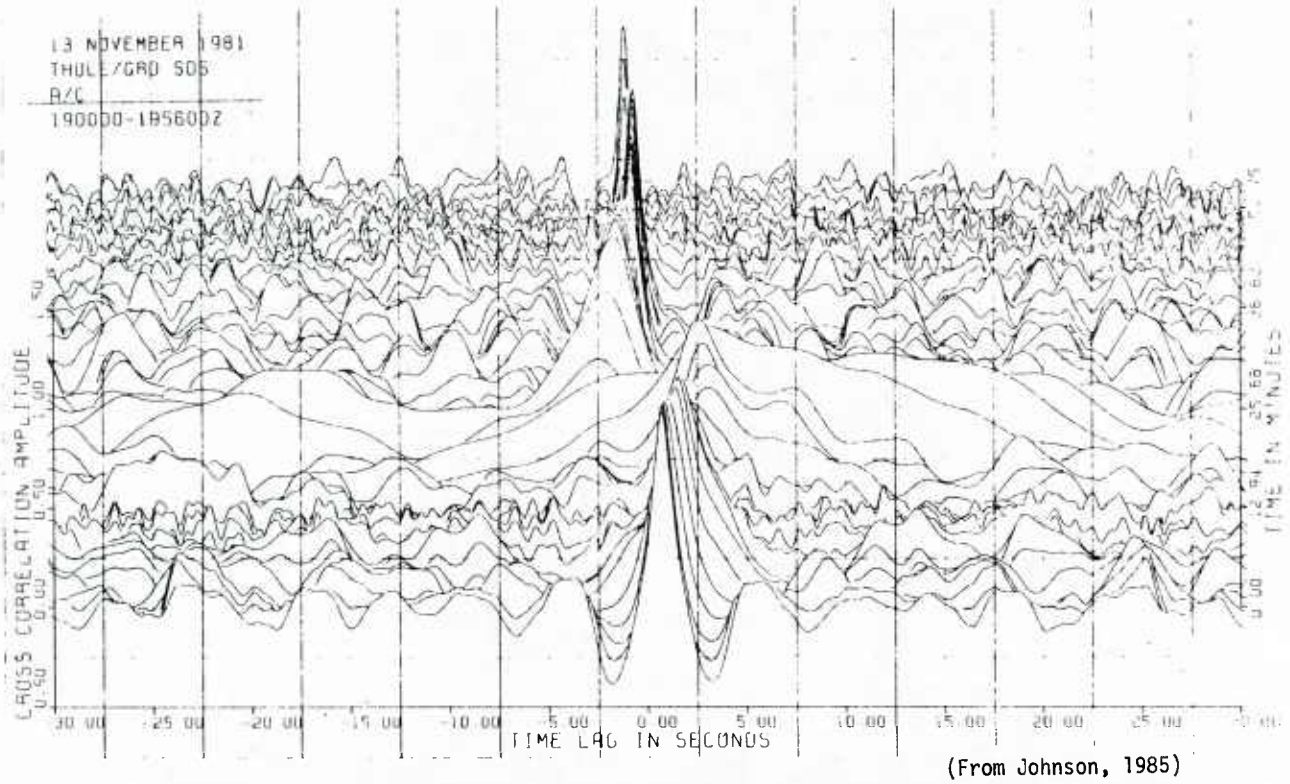
(From Johnson, 1985)

FIGURE 35. CROSS-CORRELATION OF FADING ON TWO ANTENNAS AT THULE



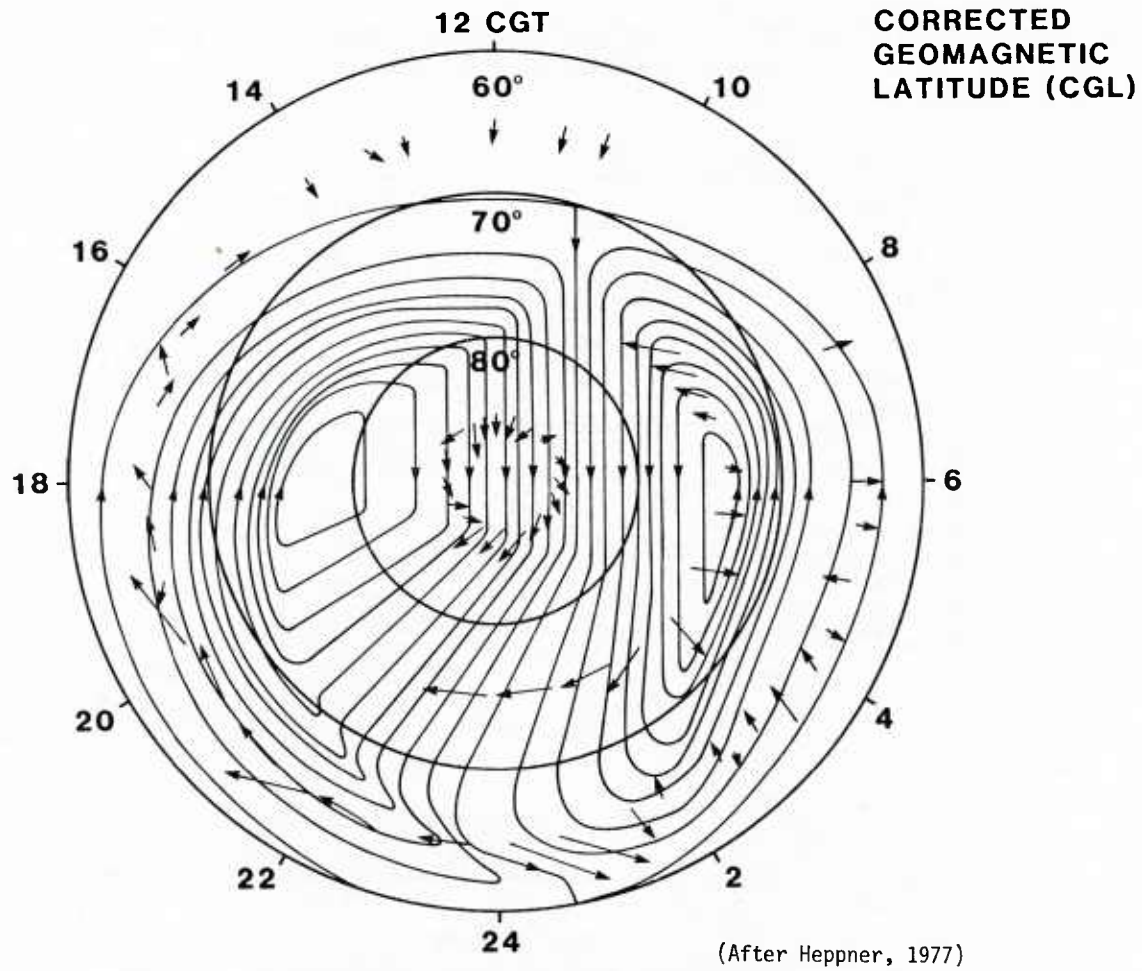
(From Johnson, 1985)

FIGURE 36. CROSS-CORRELATION OF FADING WITH CHANGE OF DIRECTION OF IRREGULARITY DRIFT



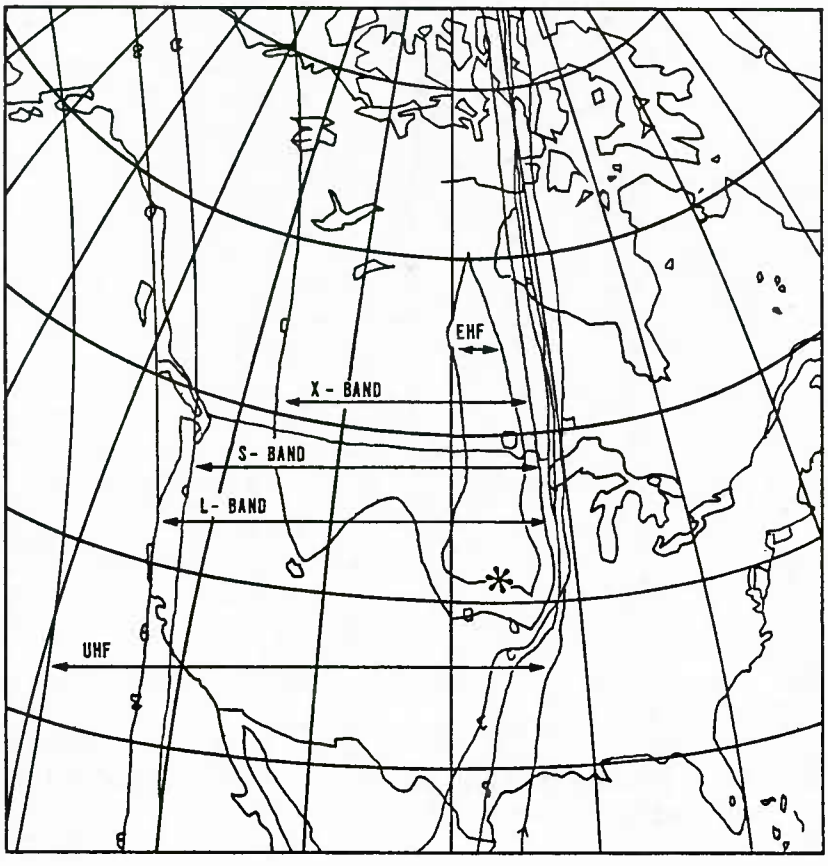
(From Johnson, 1985)

FIGURE 37. CROSS-CORRELATION OF FADING WITH CHANGE OF DIRECTION OF IRREGULARITY DRIFT



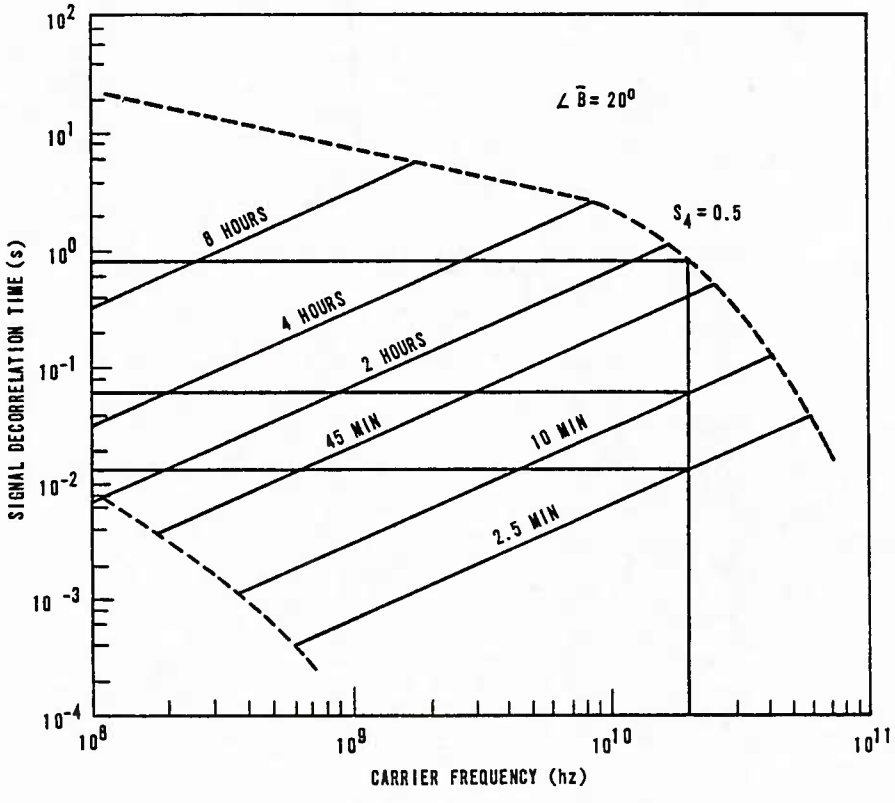
(After Heppner, 1977)

FIGURE 38. IONOSPHERIC IRREGULARITY CONVECTION PATTERN WITH DRIFT OBSERVATIONS



(From Frazier & Caster, 1983)

FIGURE 39. COMPARISON OF SCINTILLATION INTENSITY FOR VARIOUS FREQUENCY BANDS



(From Frazier & Caster, 1983)

FIGURE 40. RANGE OF SIGNAL DECORRELATION TIMES FOR VARIOUS FREQUENCIES

Probably the most important scintillation parameter driving the SATCOM link degradation is τ_o, the scintillation decorrelation time. The range of signal decorrelation times for various frequencies is shown in Figure 40. The τ_o can be viewed as the inverse of the scintillated signal fade rate. Large values of τ_o represent slow fading and smaller values of τ_o are applicable to rapid or fast fading. The τ_o is also important for characterizing a time-selective or flat signal fading environment.

The range of τ_o for a 20 GHz and a 40 GHz link are:

20 GHz $\quad 0.013 \text{ s} \geq \tau_o \geq 0.90 \text{ s}$
40 GHz $\quad 0.300 \text{ s} \leq \tau_o \leq 0.10 \text{ s}$

Interleaved coding links typically provide good message reception performance over a range of scintillation fade rates between the modulation rate and a rate corresponding to the interleaver length. For a 2400 bps link, interleaving can be used to a significant advantage.

Slow scintillation (large τ_o) causes the decoder performance to ultimately degrade due to loss of error randomization with finite-length interleavers. This degradation is virtually unrelated to the type of modulation except for its effect on code selection. The value of τ_o at which decoding performance degrades is dependent upon the time interval over which the code symbols are interleaved. This in turn depends on the interleaver length and the data rate. The manner in which the performance degrades in moderately slow scintillation depends upon interleaving and coding design parameters.

An interleaver is used to dispose the fading caused by burst errors into a random pattern which makes it easier for the Viterbi decoder to correct. When the fading τ_o of the channel approaches the time delay of the interleaver, there will be a loss in BER performance. This performance loss is due to the inability of the interleaver to mix the burst errors due to the signal fade in which a sufficient number of bits were not affected. It is, therefore, necessary to have enough interleaver storage to produce the desired BER performance for slower fading channels.

The DPSK chip combining is especially compatible with Frequency Hopping (FH) and typically provides additional jamming protection in an AWGN environment. The performance of a DBPSK link in scintillation, with or without chip repetition, depends on the value of τ_o and the interleaver length. When fading is too fast (small τ_o), or faster than the modulation rate, a DPSK link with chip repetition degrades slower than a link without chip repetition. In slow fading, a DPSK link without chip repetition will degrade slower. This is because a system without chip repetition coherently sums energy and, thus, performance is better than in a slow, Rayleigh fading environment. A system incorporating chip repetition and soft chip combining, noncoherently sums the energy. A summary of the link effects and mitigation methods is presented in Table 1.

C. Low Elevation Angle Reception

At low elevation angles between the earth terminal and the satellite, the primary propagation effects are multipath fading and ducting or radio holes in addition to the tropospheric attenuation and refractions which have already been discussed.

1. Multipath Fading

Multipath fading results from terrain reflections that can severely disrupt communications-navigation system performance. The disruption is caused by the destructive interference of the direct and reflected waves, Figure 41. The severity of the multipath fading is dependent upon the ratio of the amplitude of the direct and reflected signals. Multipath fading distorts the signal waveform by causing intersymbol interference and frequency dispersion. Observed errors cannot always be reduced simply by increasing transmitter power, but sometimes can be reduced using time, space, frequency, and polarization diversity or combination of these. Multipath fadings results are dependent upon the type of surface the reflected wave encounters. Examples of the reflecting surfaces include:

Broken - mountainous terrain,
Flat - smooth terrain,
Water - lakes, ocean, ice, snow.

UHF multipath fading on an air-to-satellite link over mountainous terrain is minimal. Since mountainous terrain does not present a good reflecting surface, multipath is seldom observed, regardless of the elevation angle between the aircraft and the satellite. There may be an occasional glint which causes fading of a few dB for a few seconds, but, in general, the fading is less than 1 dB for mountainous terrain.

Multipath fading from flat, smooth terrain follows the diffuse reflection model. The multipath signal is a combination of many separate, coherent rays. The resulting fading has a Rayleigh appearance with an average depth of 1 or 2 dB and an occasional glint causing a fade of 5 or 8 dB. The duration of the deeper fade is usually one or two seconds and occurs less than 1% of the time.

On flights over large lakes, frozen fresh water or frozen salt water (Arctic and Antarctic), and snow covered Arctic tundra, multipath fading has been observed that is quite similar to that experienced on the open ocean. The extent of the multipath fading is limited by the size of the lake.

Over water, the fading model is primarily a specular reflection resulting in a two-ray multipath condition. As the distance from the satellite to the aircraft changes, the multipath fading is very cyclical as the direct and reflected rays oscillate between in phase (enhancement) and out of phase (null), Figure 42. At low elevation angles, the water looks very smooth to the UHF radio wave. Even a rough ocean looks smooth because at the low angle only the tops of the waves and not the troughs are seens.

For circularly polarized satellite signals, the fading of approximately 3 dB is experienced at very low elevation angles. The fading increases to 8 or 10 dB at an elevation angle between 10° and 15° and then drops off to 1 dB or less at high elevation angles. An estimate of the fading characteristics versus elevation angle is given in Figure 43, (Johnson, 1979-2). This is a heuristic chart developed from thousands of hours of airborne testing performed since 1967.

The degree of multipath fading experienced is dependent upon the antenna used and its exact placement on the aircraft. If the antenna has poor gain in the direction of the multipath reflection, the fading will be minimized. In recent antenna tests it was noted that one antenna experienced multipath while the other six antennas under test saw no multipath at the particular elevation/azimuth angle. Not only is the antenna gain toward the horizon important, but also the polarization. All antennas have an axial ratio, and the exact relation of horizontal polarization to vertical polarization is important with respect to

TABLE 1. SUMMARY OF SATELLITE LINK EFFECTS AND MITIGATION METHODS

Frequency	Systems and Missions	Extent and Duration of Major Effects	Dominant Degradation Effects*	Mitigation Methods
UHF (225-400	Strategic Missions AFSATCOM FLTSATCOM	Absorption over hundreds of kilometers lasting tens of minutes	Loss of E_b/N_o below demodulation threshold	Increased EIRP (not very effective); multiple redundant links
		Scintillation over thousands of kilometers for hours in striated regions for high-altitude bursts	Severe phase and amplitude fading with slow to very fast fade rates	Optimal demodulator design, coding, interleaving, repetition, increased EIRP
		Scintillation over CONUS-sized regions for hours in striated ambient F region from bursts at any altitude	Moderate to severe phase and amplitude fading with slow to moderate fade rates, comparable to severe natural scintillations	Optimal demodulator design, coding, interleaving, repetition, increased EIRP
L, S Bands (1.2-2.3 GHz)	GPS Navigation DSP Data Links	Absorption over tens of kilometers lasting few minutes	Loss of E_b/N_o	Increased EIRP, Link redundancy
		Scintillation over thousands of kilometers for hours in striated regions from high-altitude bursts	Severe phase and amplitude fading with slow to very fast fade rates	Optimal demodulator design, coding, interleaving (GPS), increased EIRP, multiple links, spaced antennas (DSP)
		Scintillation over CONUS-sized regions for hours in striated ambient F region from bursts at any altitude	Moderate to severe phase and amplitude fading with slow to fast fade rates	Optimal demodulator design, coding, interleaving (GPS), increased EIRP, multiple links, spaced antennas (DSP)
X-Band (7.2-8.4 GHz)	DSCS Other proposed links	Absorption over tens of kilometers for few seconds	Transcient loss of E_b/N_o	Increased EIRP, link redundancy
		Scintillation over hundreds to thousands of kilometers for few hours in striated regions from high-altitude bursts	Moderate to severe phase and amplitude fading with slow to fast fade rates	Optimal demodulator design, coding, interleaving, repetition (some links), increased EIRP, multiple links
		Dust and/or water vapro (or ice) from surface or near-surface bursts over tens of kilometers for hundreds of seconds	Potential deep amplitude and phase scintillations with moderately fast fade rates	Optimal demodulator design, coding, interleaving, repetition (some links), increased EIRP, multiple links
EHF	AFSATCOM (SSS/NFCS) Other proposed systems	Scintillation over hundreds of kilometers for about an hour in striated regions from high-altitude bursts	Moderate to severe phase and amplitude fading with moderate fade rates along some paths	Optimal demodulator design, coding, interleaving, repetition (some links), increased EIRP, multiple links
		Dust and/or water vapor (or ice) from surface or near-surface bursts over tens of kilometers for tens of minutes	Moderate attenuation and severe phase and amplitude scintillation	Optimal demodulator design, coding, interleaving, repetition (some links), increased EIRP, multiple links
		Tropospheric effects (clouds, rain, etc.)	Attenuation of E_b/N_o scintillations	Optimal demodulator design, coding, interleaving, repetition (some links), increased EIRP, multiple links

*Possible degradation due to dispersion, noise, and other effects may be a problem, depending on specific system characteristics.

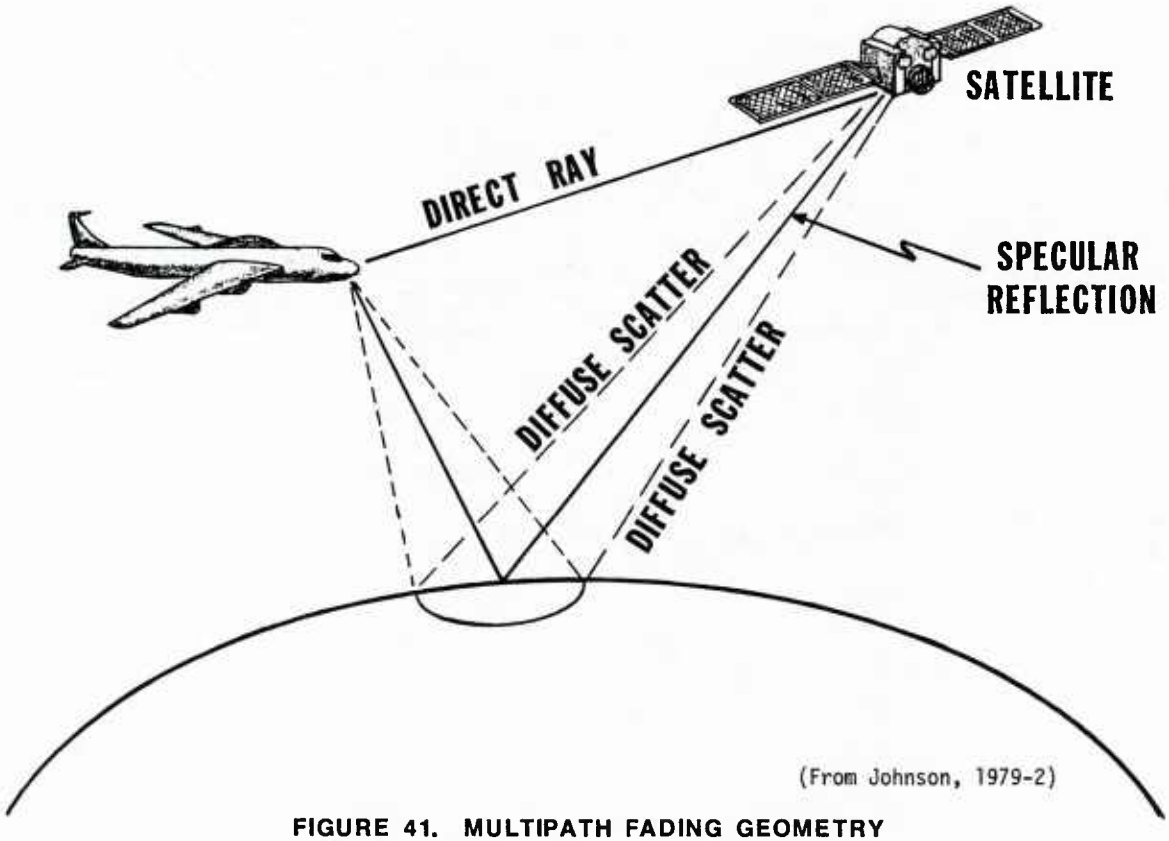


FIGURE 41. MULTIPATH FADING GEOMETRY

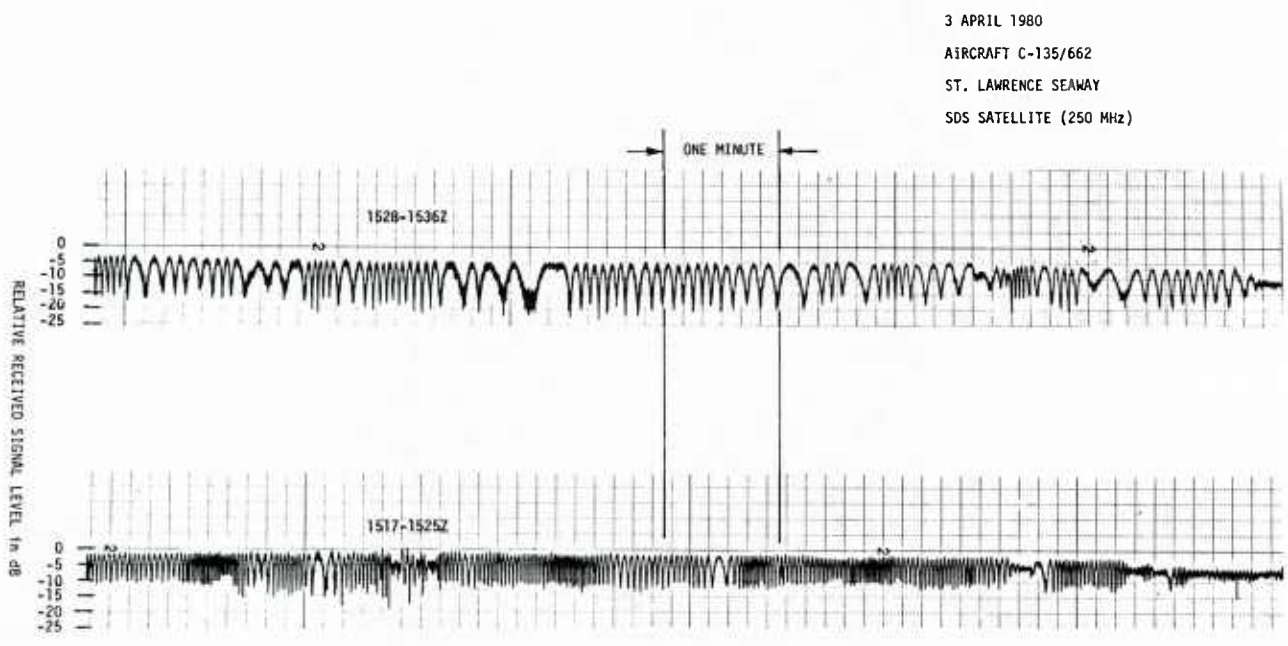


FIGURE 42. VARIATIONS IN AIRBORNE UHF MULTIPATH FADING OVER WATER

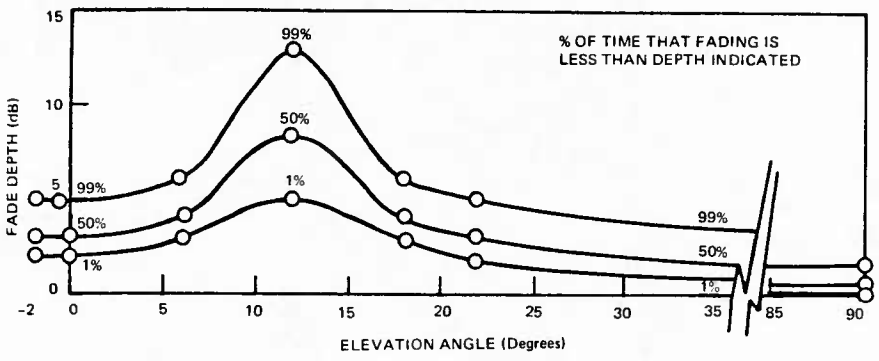


FIGURE 43. MULTIPATH FADING DEPTH

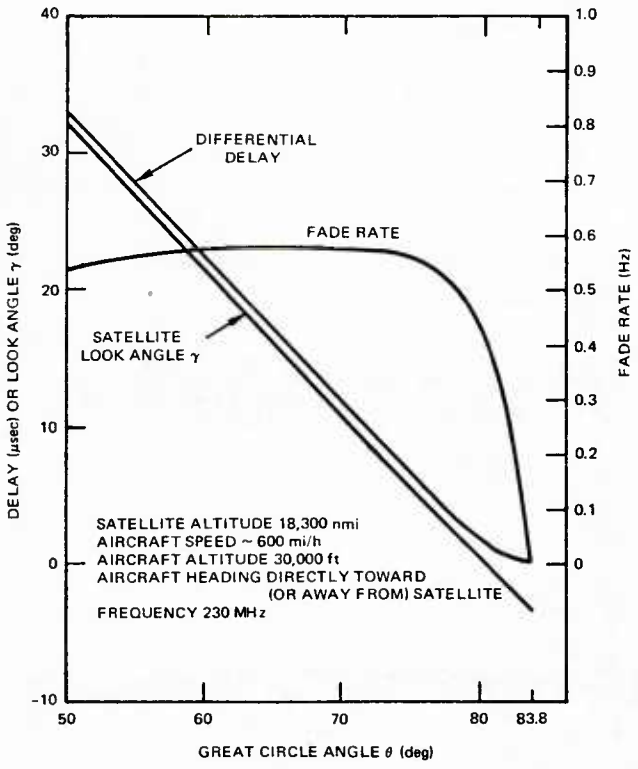


FIGURE 44. TWO-RAY MULTIPATH FADING RATE

multipath. Therefore, the multipath fading depths in Figure 43 are an average for a wide variety of antennas (blades, crossed dipoles, spirals, crossed slots, turnstile, double-tuned stubs and dual mode) and a wide variety of antenna placements on the aircraft.

A plot of the expected two-ray multipath fading rate is shown in Figure 44 and the differential delay in Figure 45. The fade rate for the fading signal shown in Figure 46 can be measured by taking the power spectral density of the signal, Figure 47. While the diffuse power spectral density decreases at approximately f^{-2}, a very strong specular component can be seen at approximately 0.75 Hz. The autocorrelation function of this same signal shows a repetitive correlation with a spacing of approximately 1.3 seconds, Figure 48.

a) PRN Measurement Technique

One of the measurement techniques used to separate the direct and reflected signal component is the use of psuedo-random (PRN) sequence and correlation technique (extracted from Johnson, 1979-2). One PRN system that has been tested, utilized a 127 bit direct psuedo-random sequence transmitted through the satellite. At the receiving terminal the matching 127 bit psuedo-random sequence was correlated with the receive signal and with the time-delayed reflected signal, which has traveled a further distance, Figure 49. The relative amplitude between the direct and the reflected correlation indicates the amplitude difference of the direct and reflected signals. The time difference indicates differential path delay.

The results of the PRN correlation technique with the satellite at a high elevation angle are shown in Figure 50 for an overhead type antenna (Johnson, 1979-2). The reflected signal is delayed approximately 50 microseconds and is approximately 20 dB lower in amplitude than the direct signal. The results in Figure 51 are for a horizon coverage antenna at a satellite elevation angle of 24°. Here the reflected signal is only approximately 1 dB below the direct signal. A rather long tail of diffuse energy is evident with delays up to 15 microseconds later than the specular component. The energy reflected from ice, Figure 52, is approximately 5 to 10 dB below the direct signal. The reflections from land, even at relatively low elevation angles (Figure 53) are down significantly from the water reflected multipath.

b) Directive Antennas

Another technique which provides information on the reflected component is the use of a directive, bottom-mounted antenna and a separate antenna mounted on top of the aircraft (Johnson, 1979-2). This technique allows the measurement of the direct and reflected components for medium to high elevation angles, as shown in Figure 54. Over water the bottom antenna yielded a strong reflected signal at high elevation angles with the average reflected energy approximately 6 dB less than the direct received energy. The bottom received signal, Figure 55, appears noise-like with fade rates of 10 to 100 Hz and a peak-to-null amplitude of 10 to 15 dB. Over land, at a high elevation angle, the bottom antenna yielded a less consistent reflected signal, Figure 56. The average signal energy was approximately 10 dB less than the direct signal.

2. Radio Holes and Ducting

As an earth terminal transmits radio signals to a satellite at low elevation angles it is possible for the transmitted rays to get trapped in an inversion layer due to the abnormal refractive index of the layer (extracted from Johnson, 1974).

a) Radio Holes

Radio waves traveling through an ideal, or "Standard Atmosphere," bend slightly due to the change in the refractive index with altitude, (Johnson, 1974). This change in the refractive index can extend communications to distances beyond optical line-of-sight. To compensate for this bending effect a correction factor often used is 4/3 the actual earth radius. However, a "Standard Atmospheric" gradient of refractive index is not always encountered. In certain areas of the world inversion layers are common and can cause a sharp discontinuity in the atmospheric refractive index gradient. These inversion layers are usually a few hundred feet thick and can extend for hundreds of miles.

As depicted in Figure 57, if energy is being radiated by a source above the inversion layer, say from either an airplane or a satellite, the energy that passes above the inversion layer travels in effectively a straight line. The rays that strike the layer at a shallow angle can be bent or refracted and bunched together with waves that pass through the layer uneffected. In this case an energy void is cause by the bending of the rays. This energy void area is called a "Radio Hole." The bunching of energy below the radio hole is called an "Anti-Hole."

If an aircraft is moving away from the signal source at a constant altitude, as depicted in Figure 57, the received signal strength would normally decrease at the rate of approximately 6 dB per octave due to the free space loss until the aircraft reaches the Radio Hole. The signal strength would then drop quite rapidly until the aircraft enters the Anti-Hole. Data taken from an air-to-air flight test experiment illustrates this effect, Figure 58. It can be seen that the received signal strength follows closely the free space loss to a separation of approximately 220 miles. Between 220 and 235 miles the signal strength drops below the free space condition. This is the Radio Hole region. Upon reaching the Anti-Hole the signal recovers to slightly above the free space condition. At about 260 miles the aircraft is beyond line-of-sight and the signal strength decreases rapidly.

The Radio Hole can be modeled if the refractive index gradient is known and is essentially consistent. Information on refractive index is available from the meteorological services. Normal measurements are taken by weather balloons, but more accurate information can be obtained from airborne or air dropped microwave refractometer.

b) Ducting

Another atmospheric phenomena closely related to the Radio Hole is ducting (Johnson, 1974). In the case of the duct, the energy is trapped between an inversion layer and the earth, or between inversion layers. The signal in the duct can travel far beyond the horizon with little attenuation. Ducting is a phenomena that can be observed over approximately 35% of the earth's surface with fair regularity. A belt known as the Trade Wind Duct exists in the area centered approximately 20° north and 20° south of the equator. Warm air rising from the equator flows towards the poles and falls in the vicinity of these latitudes. This warm, moist air encounters cool, dry air near the ocean and an inversion layer is formed. Specific areas in these trade wind ducts are well known for their observed ducting effects. For example, this ducting effect is frequently used for VHF communications between Brazil and Africa. Similar ducting effects have been observed between southern California and Hawaii in the VHF spectrum. Strong layering exists south of Bermuda in an area known as the "Bermuda High."

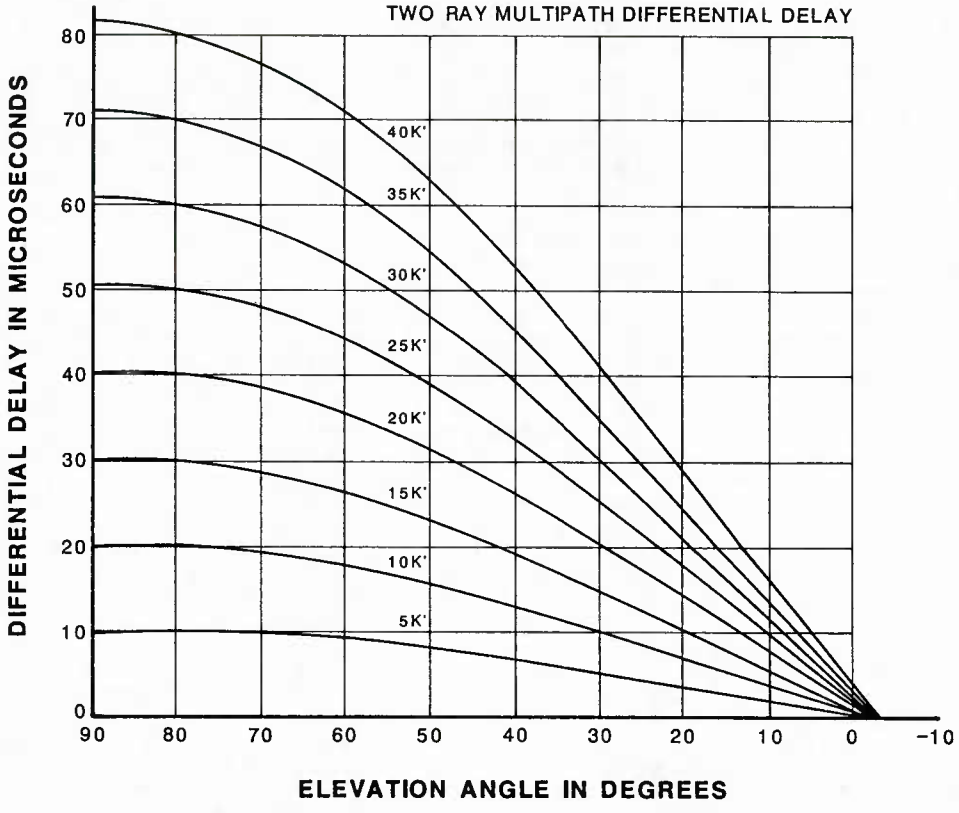
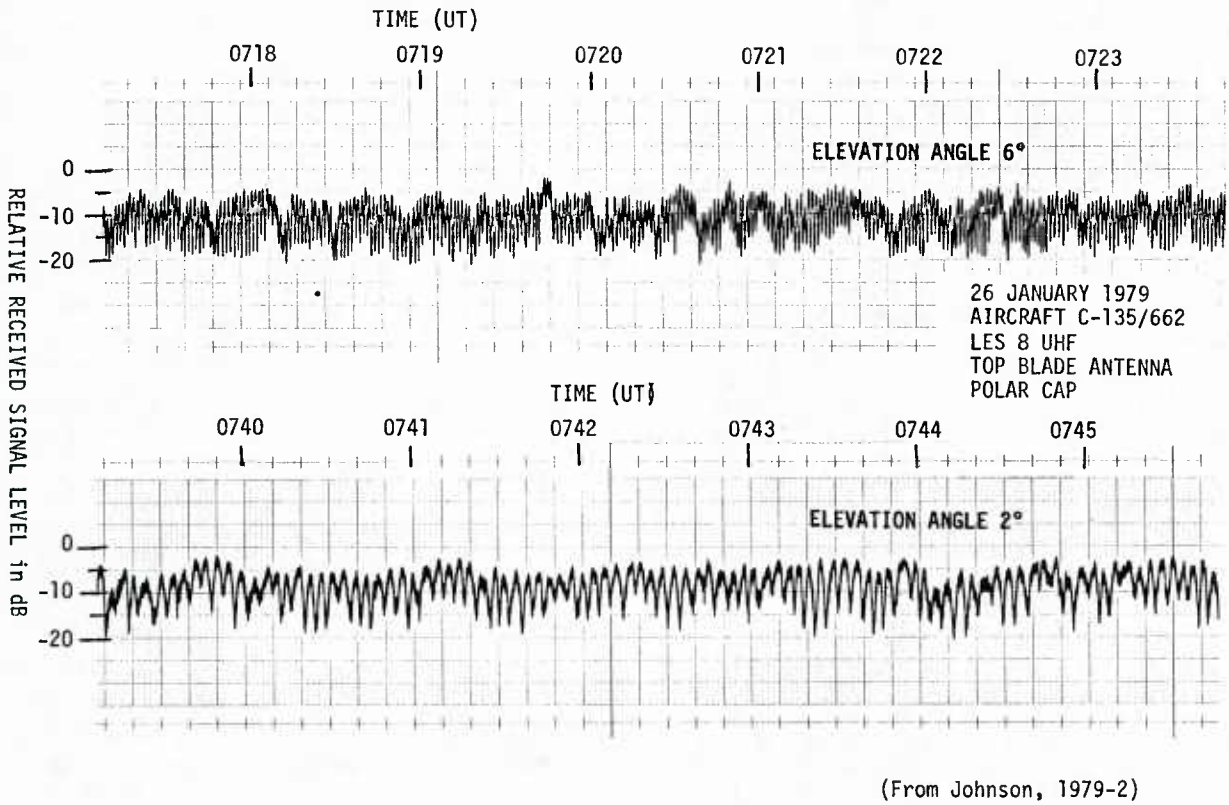
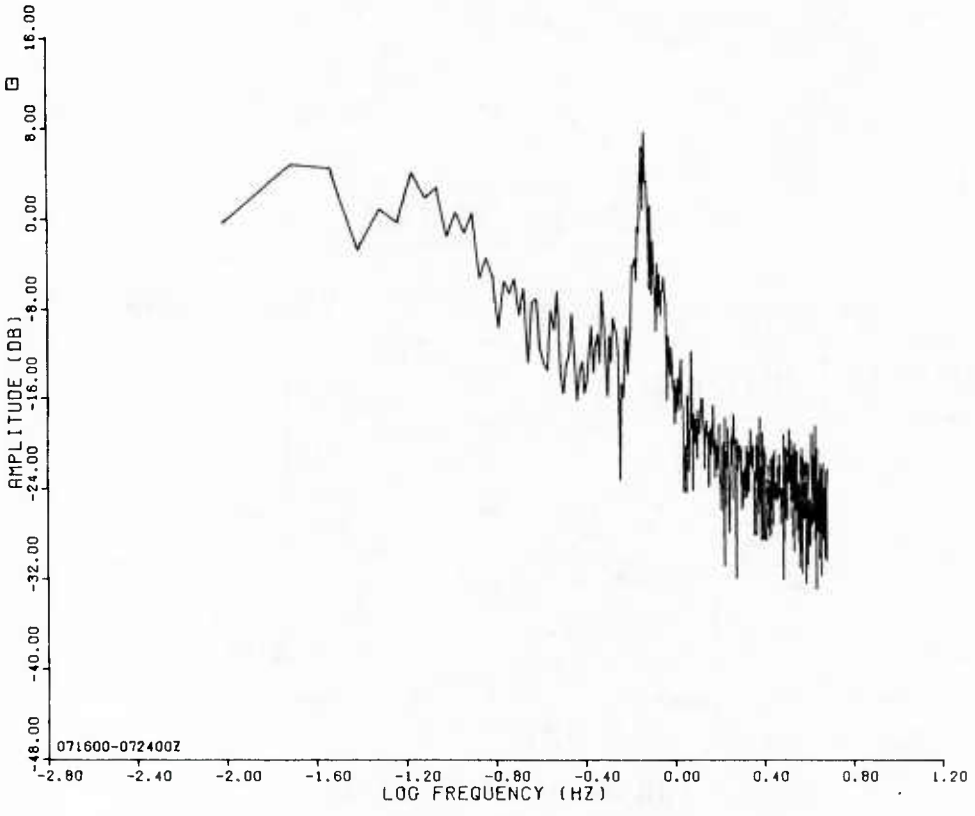


FIGURE 45. DIFFERENTIAL DELAY VERSUS ELEVATION ANGLE



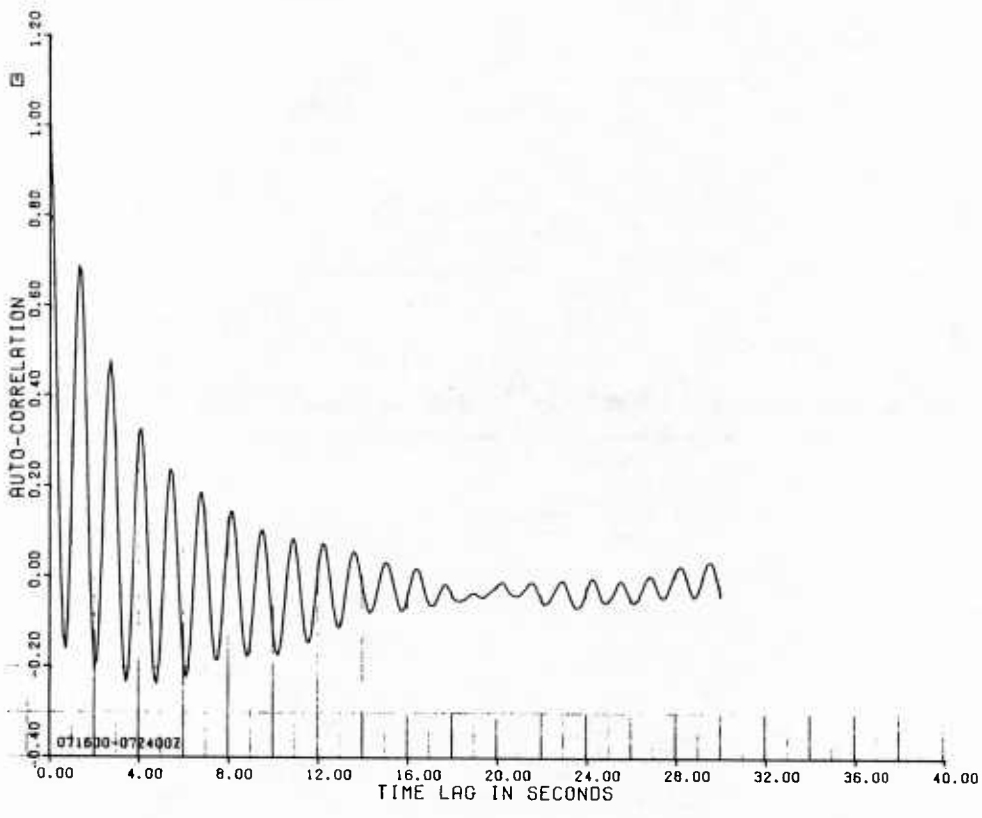
(From Johnson, 1979-2)

FIGURE 46. EXAMPLE OF LOW ELEVATION ANGLE UHF MULTIPATH FADING



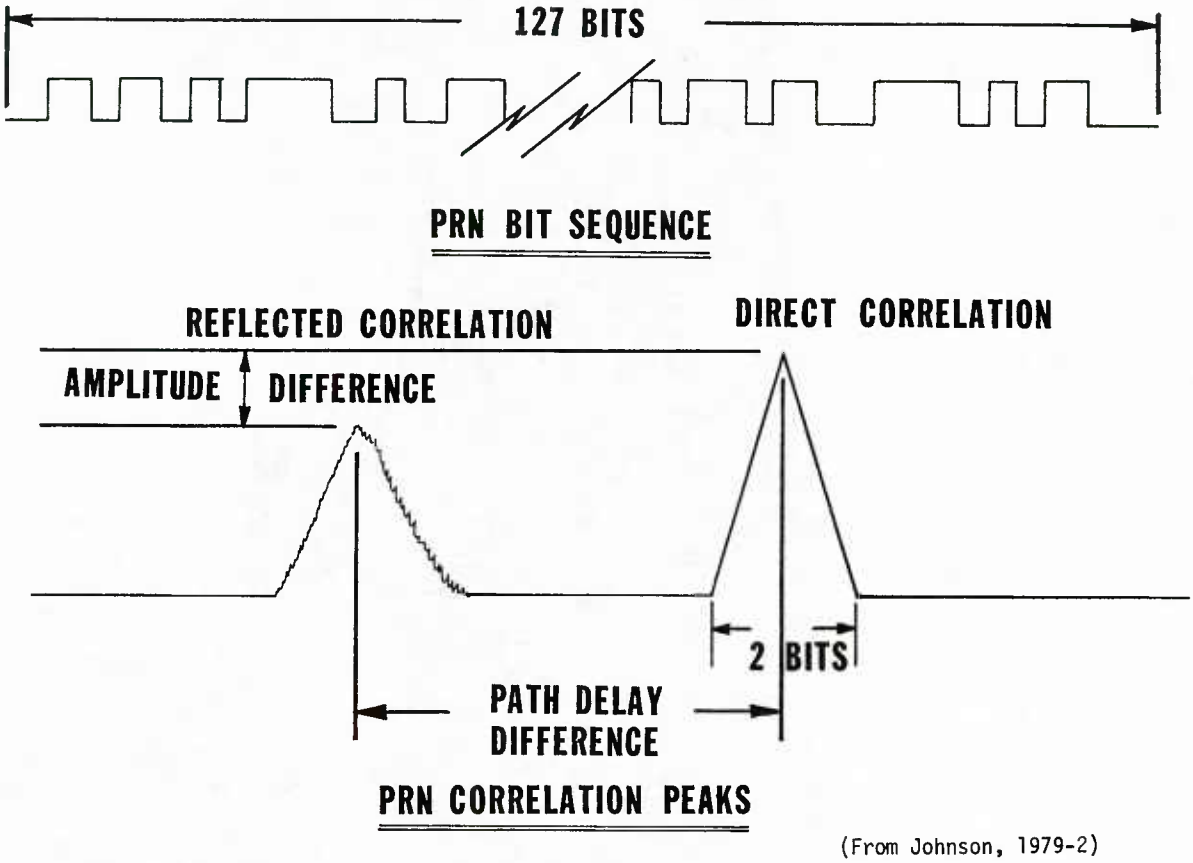
(From Johnson, 1979-2)

FIGURE 47. POWER SPECTRAL DENSITY OF MULTIPATH SIGNAL



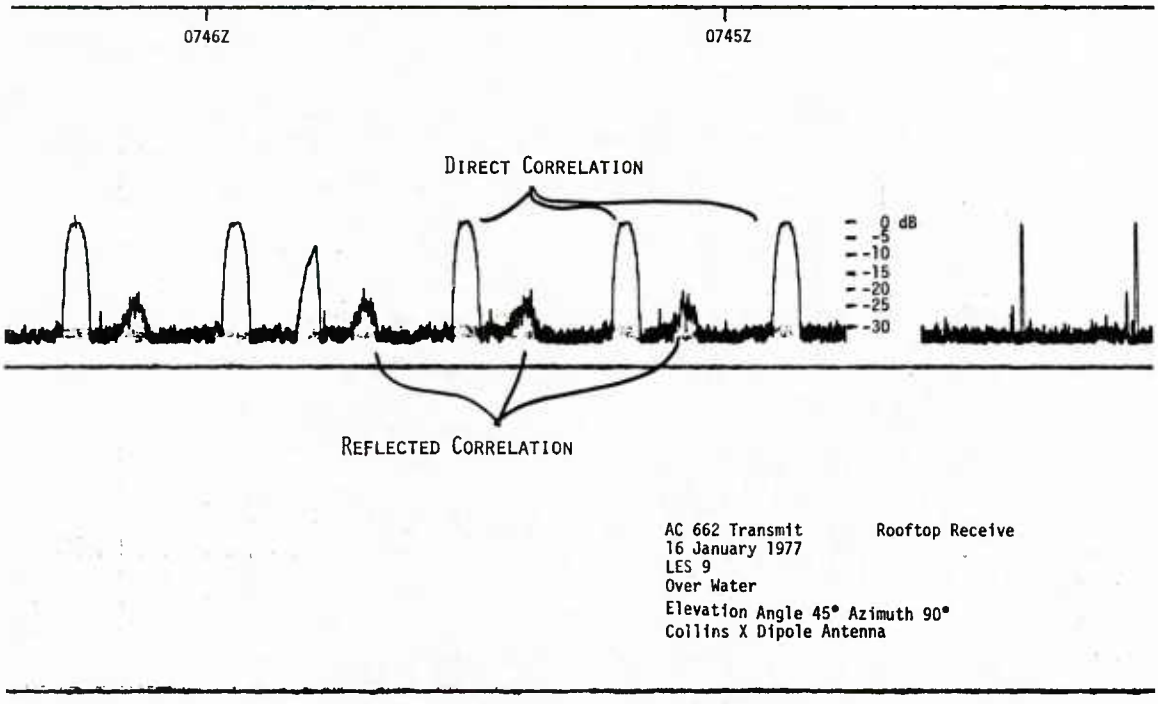
(From Johnson, 1979-2)

FIGURE 48. AUTOCORRELATION OF MULTIPATH SIGNAL



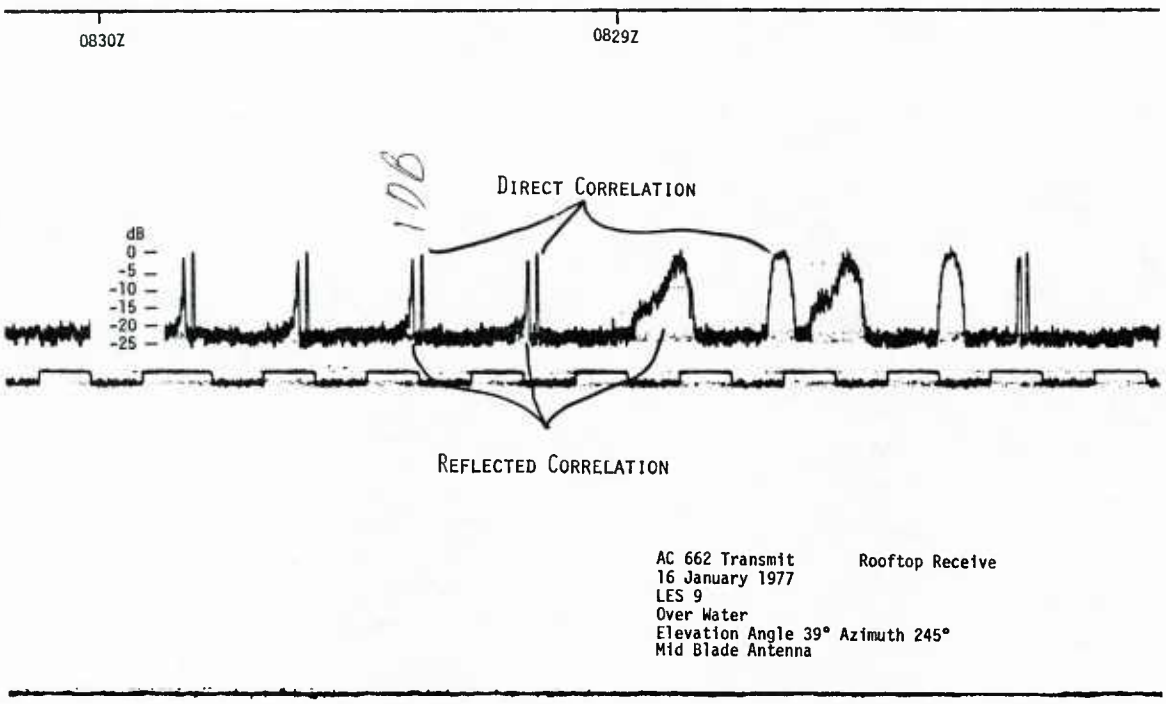
(From Johnson, 1979-2)

FIGURE 49. PSUEDO RANDOM SEQUENCE MEASUREMENT TECHNIQUE



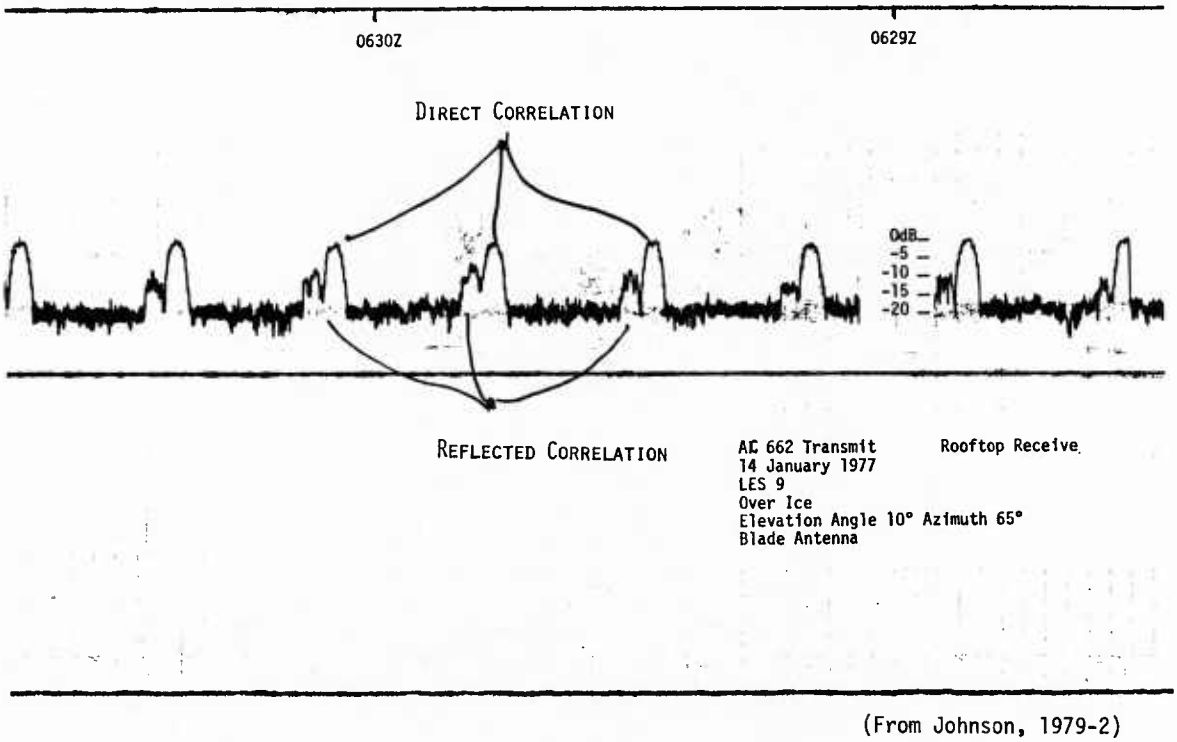
(From Johnson, 1979-2)

FIGURE 50. RESULTS OF PRN CORRELATION WITH OVERHEAD ANTENNA



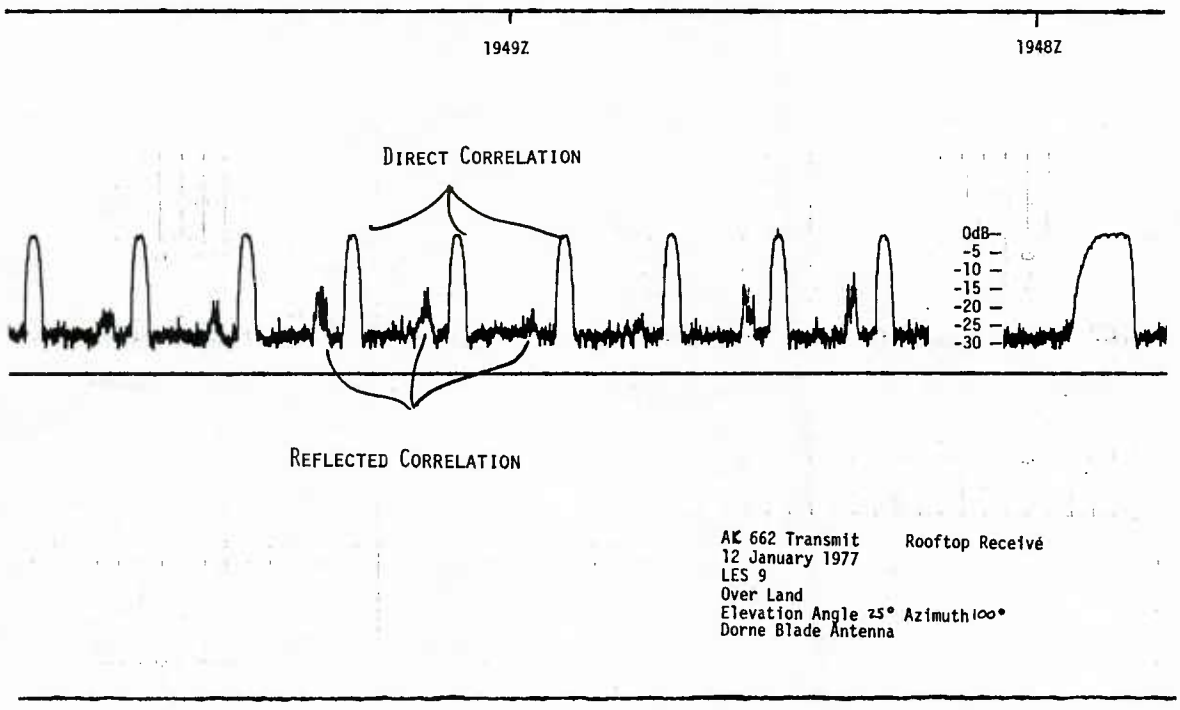
(From Johnson, 1979-2)

FIGURE 51. RESULTS OF PRN CORRELATION WITH HORIZON ANTENNA OVERWATER



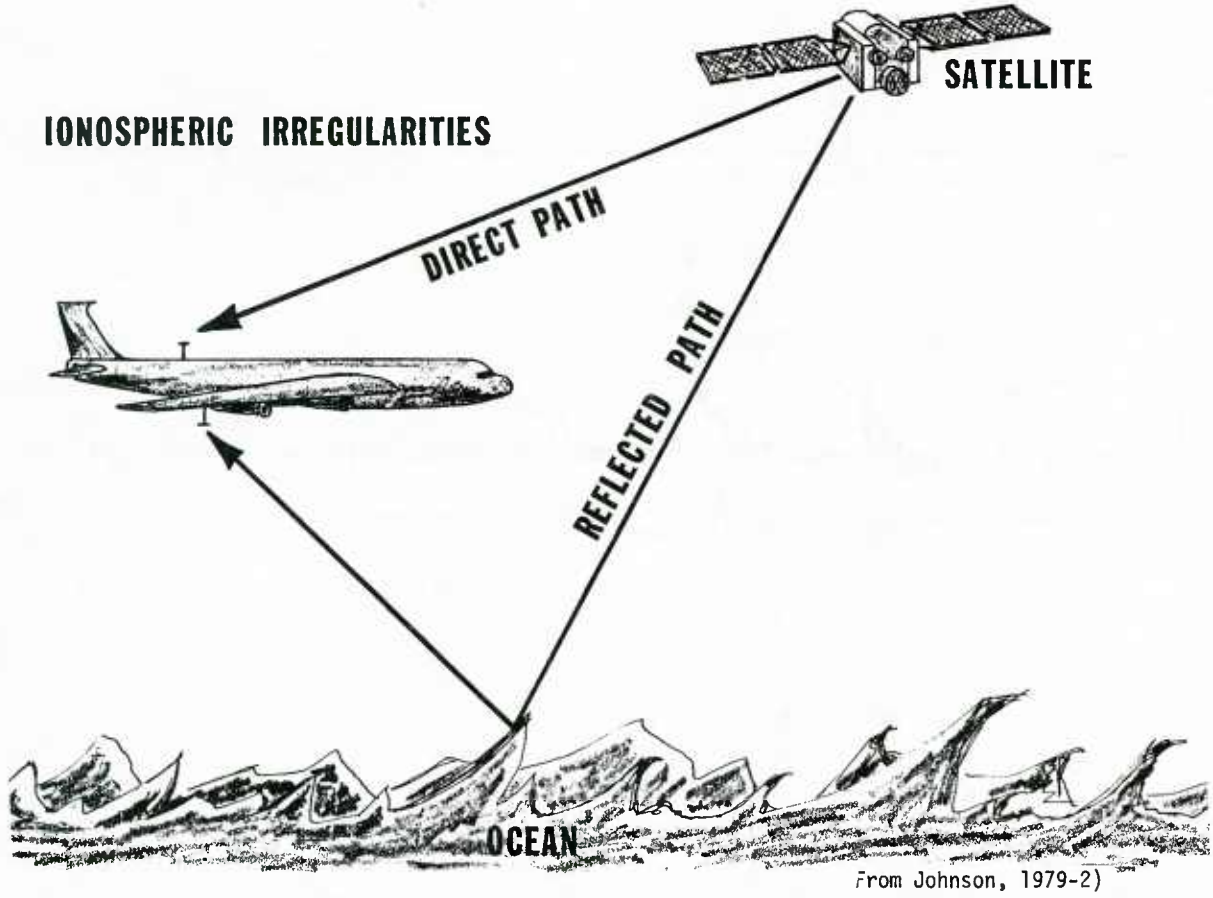
(From Johnson, 1979-2)

FIGURE 52. RESULTS OF PRN CORRELATION WITH HORIZON ANTENNA OVER ICE



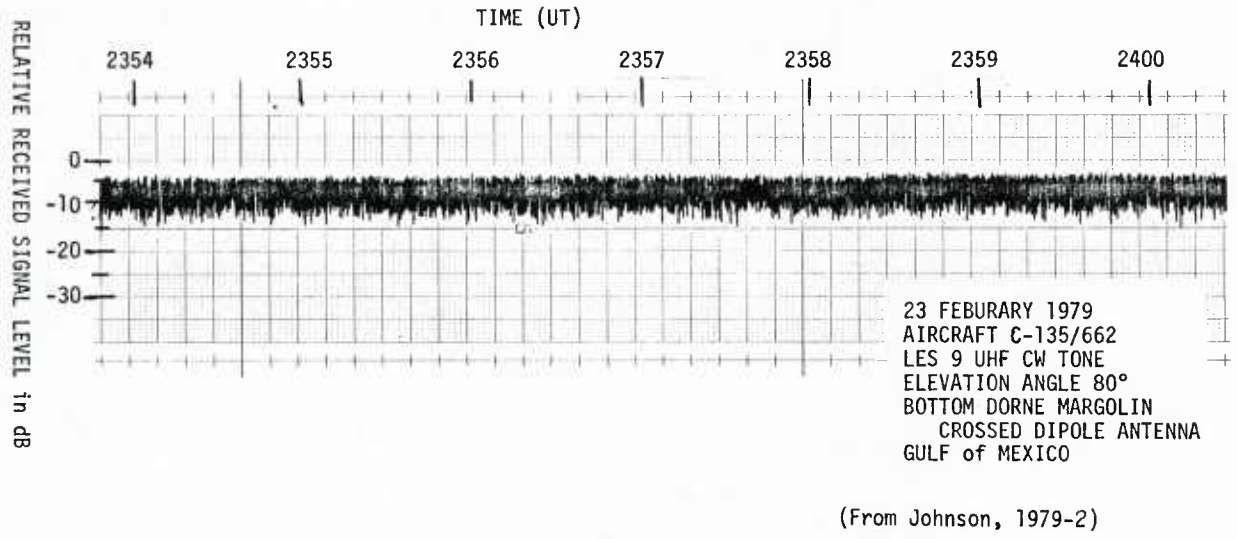
(From Johnson, 1979-2)

FIGURE 53. RESULTS OF PRN CORRELATION WITH HORIZON ANTENNA OVER LAND



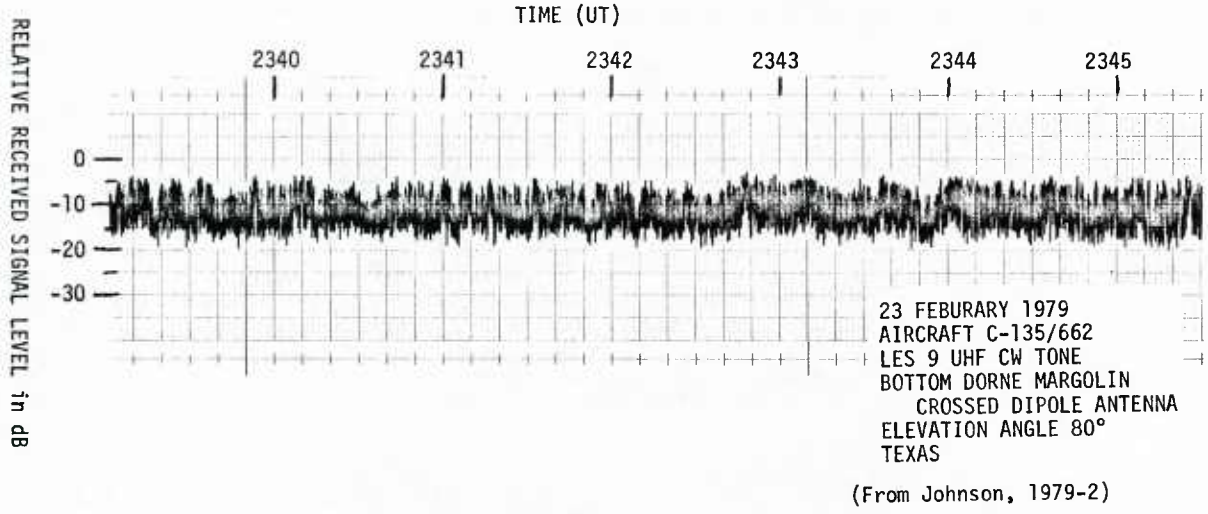
From Johnson, 1979-2)

FIGURE 54. USE OF DIRECTIVE ANTENNA TO MEASURE MULTIPATH COMPONENTS



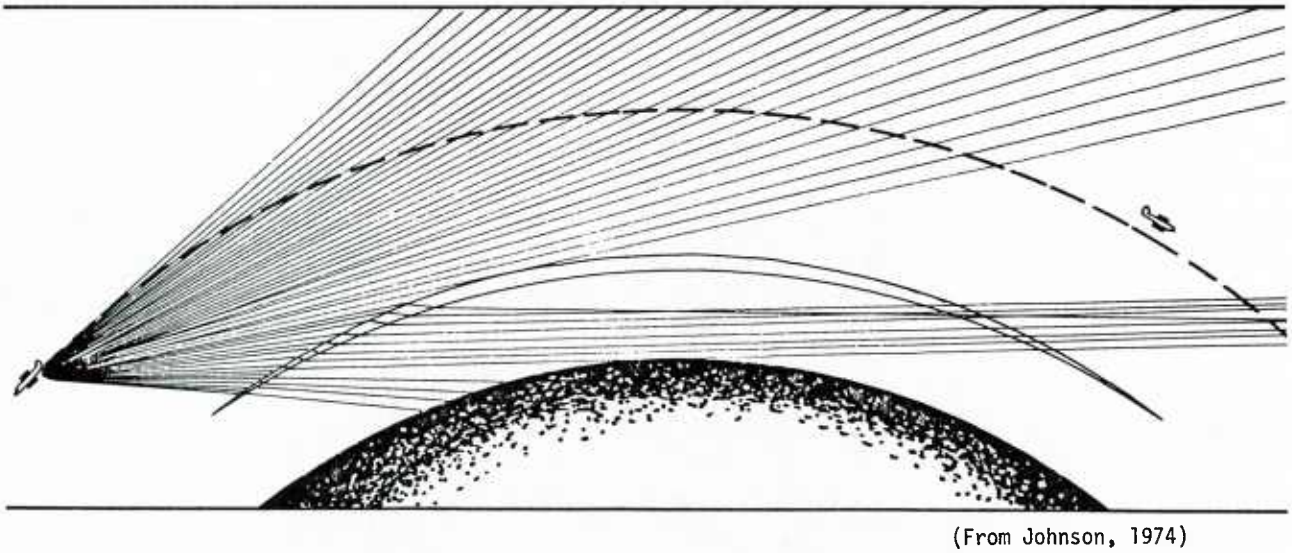
(From Johnson, 1979-2)

FIGURE 55. RECEIVED SIGNAL FROM BOTTOM ANTENNA OVERWATER



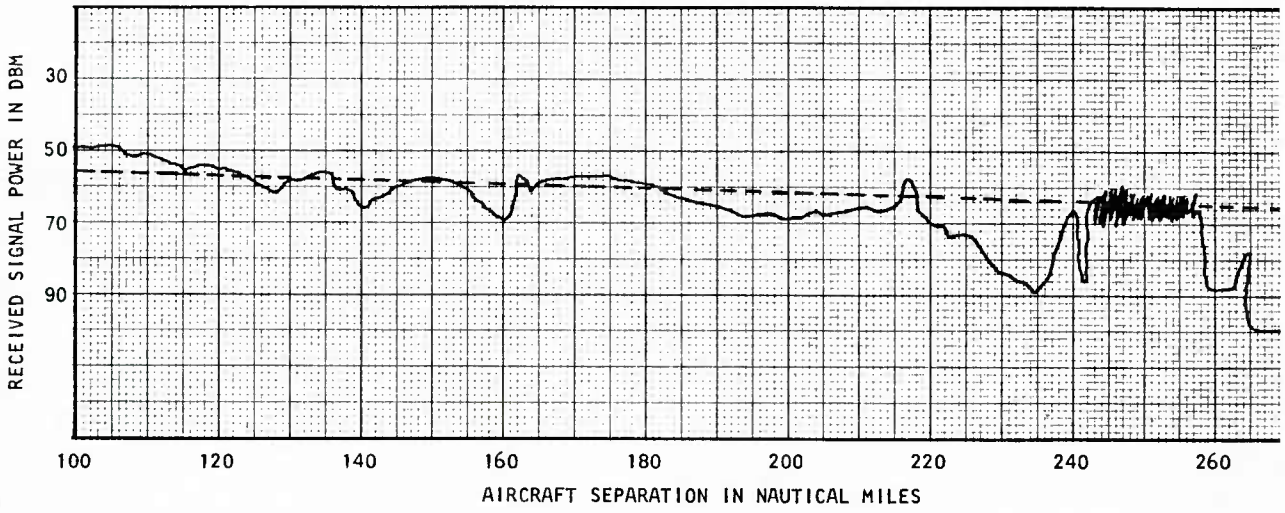
(From Johnson, 1979-2)

FIGURE 56. RECEIVED SIGNAL FROM BOTTOM ANTENNA OVER LAND



(From Johnson, 1974)

FIGURE 57. GEOMETRY OF A RADIO HOLE



(From Johnson, 1974)

FIGURE 58. RECEIVED SIGNAL STRENGTH IN THE VICINITY OF A RADIO HOLE

By observing the altitude and shape of the duct through current weather information it is possible to position two aircraft within the duct and maintain communications far beyond line-of-sight. Such a test was run in the West Indies area. In this test two aircraft, each flying at a 15,000 feet altitude and separated by 700 nautical miles, could communicate high-quality wide-band television-type signals. The normal line-of-sight for aircraft at this altitude is approximately 300 miles.

Extensive utilization of ducting and beyond the line-of-sight communications is complicated by the narrowness and irregularity of the ducts. The modeling of ducting communications is well understood and computer program models have been developed to predict the communications ducts at various frequencies, once a consistent and continuous refractive index gradient information is fed into the computer system. In the future, more extensive reporting of refractive index information could help to make the duct a feasible propagation media.

D. Implication of Mobile Terminals

Mobile terminals are subjected to a variety of effects that are normally not experienced by fixed terminals, such as Doppler frequency shift and antenna pointing errors.

1. Doppler Effects

The Doppler experienced by the mobile terminal can be expressed as:

$$D = F \cdot V/C$$

where D = Doppler in Hertz,
F = frequency in Hertz,
V = velocity in meters per second
and C = speed of light in meters per second.

The velocity of interest is the radial velocity between the satellite and the mobile terminal. Since all satellites are not equatorial synchronous, a fixed ground station can also experience Doppler due to the relative motion of the satellite. The mobile terminal must search out a larger frequency uncertainty because of the Doppler and find the downlink beacon signal from the satellite. Once the satellite signal has been received, the mobile receiver can track Doppler, which for an SHF airborne system may change at the rate of a few hundred Hertz per second. The Doppler is strictly a geometric consideration and can be calculated. Computer-aided antenna-pointing systems can calculate the Doppler by knowing the satellite orbital information and the aircraft velocity vector. The Doppler shift is not sensitive to propagation anomalies and ionospheric effects tends to be negligible at UHF and SHF frequencies.

2. Antenna Effects

Pointing antenna pointing is another consideration for the mobile terminal. With UHF antennas the problem is usually the selection of an antenna providing horizon or overhead antenna. For the SHF mobile terminal the beamwidth may be as small as a few degrees and an active tracking system is required to keep the mobile terminal antenna pointed towards the satellite. Errors in the antenna tracking algorithm will degrade the antenna pointing and reduce the link margin. Active tracking systems using a nutating sub-reflector have been extensively evaluated and provide antenna pointing with negligible errors, even during propagation anomalies. Computer-aided antenna-pointing system usually do not consider ray bending in their calculations and therefore can be in error by the amount of the ray bending. At SHF, the antenna beam width is usually wider than the ray bending and therefore the error is negligible.

II. System Design Aspects of Satellite Communications

The design of a satellite communications system should take into account the propagation characteristics of the communications channel. A balance is required to obtain high efficiency along with a degree of robustness needed to assure reliable communications.

A. Modulation Techniques

The process by which information is impressed on a suitably chosen RF carrier frequency for transmission is referred to as modulation (extracted from DCA Handbook, 1972). The two forms of modulation that are normally used to convey information are analog and digital modulation processes. Analog modulation methods encompass those techniques which vary the amplitude, phase, or frequency of the carrier over a continuous range of values in response to source information. Digital modulation methods encompass those techniques that cause discrete changes in the amplitude, phase, or frequency of the carrier after the source information has been digitized into a stream of marks (ones) and spaces (zeros) prior to the carrier modulation process. Since analog modulation is not normally employed on satellite links this section will be limited to digital modulation.

A good digital modulation system is one which communicates the maximum amount of data reliably. The performance of a digital system is bounded theoretically by Shannon's classical results, which state that given sufficient processing, data at rate R in bits per second can be communicated with arbitrarily small error through a channel of capacity C, provided $R < C$. More specifically if a channel is linear, restricted to bandwidth B, has average power S_i, and is perturbed by additive white Gaussian noise (with single-sided noise power density N_o), then:

$$R \leq C = B \log_2 1 + \frac{S_i}{N_o B}$$

This equation can be rewritten in terms of the system bandwidth expansion factor ($m = B/R$) and the available energy per bit-to-noise density ratio, $E_b/N_o = S_i/N_o R$, as follows:

$$R/B \leq \log_2 1 + \frac{E_b R}{N_o B}$$

This leads to an expression giving the minimum possible energy per bit to noise density ratio required to reliably communicate with bandwidth expansion factor m.

In a binary digital communications system (higher-order or M-ary systems are possible but will not be discussed), the transmitter selects and sends one of two basic waveforms depending on whether a binary zero or one is delivered by the digitized (i.e., sampled and quantized) information source. For example, if the Nth term of the binary digital sequence is a zero, the system transmits, during interval $NT \leq t \leq (N+1)T$, the signal waveform $S_0(t)$; if it is a one, the system transmits $S_1(t)$. Since T seconds are required to transmit one symbol of the sequence, the transmission rate is $1/T$ symbols per second. If the binary data source selects the bits of the sequence with equal probability and independently, then each symbol conveys one bit of information and hence the information rate is $1/T$ bits per second. Because of the addition of noise (e.g., the receiver will always have a finite temperature and thus will add a finite noise power density to the received signal), there will be a nonzero probability that the receiver will make an error when attempting to decide whether $S_0(t)$ or $S_1(t)$ was transmitted. The performance of the digital system therefore is measured in terms of the probability that the receiver will err in making this and subsequent decisions.

To calculate the performance of a particular digital modulation technique requires knowledge of the form of the waveforms selected for transmission, the transmission channel characteristics and the exact structure and decision rule employed at the receiver. The signal waveform characteristics of interest are the energy per signal, the spectral occupancy and a measure of the basic difference in the signal structures used to represent transmitted zeros and ones (this latter characteristic is referred to as the signal's normalized inner product and is denoted as p).

1. Phase Shift Keying

The channel characteristics of interest pertain to its linearity, phase response, and noise properties. The receiver structure details pertain to a priori knowledge assumed to exist regarding which signal was sent, knowledge of the signal structure (i.e., the signals epoch, amplitude, phase, and frequency), and the decision rule employed. For example, if information is conveyed by shifting the carrier phase 180° to represent a one (mark) and by 0° to represent a zero (space) and all other aspects of the waveforms are identical, the system is usually referred to as binary antipodal modulation, or simply binary phase-shift keying (PSK or BPSK). For an additive white Gaussian noise channel that permits retention of carrier phase information, the optimum receiver can be shown to be a coherent matched filter receiver with one filter matched to each of the phase states sent at the transmitter. Exact knowledge of received signal time, phase, and frequency is required to implement this ideal detector, and the decision rule is simply to look at the outputs of the matched filters every T seconds and decide in favor of the filter with the largest output (this assumes ones and zeros are equally likely to have been transmitted and that ties are resolved by simply flipping a coin). For this binary PSK system with a coherent ideal PSK detector, the performance is given by:

$$P_E = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \quad \text{(coherent, antipodal PSK signals)}$$

where

$$Q(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} \exp(-y^2/2)dy$$

$$E^* = \text{signal energy} = \int_0^T S_1^2(t)dt = \int_0^T S_0^2(t)dt$$

and the normalized inner product a measure of the correlation of the binary signals, is:

$$p = \int_0^T \frac{S_0(t)S_1(t)dt}{E_b} = -1 \quad ; \text{(antipodal PSK signaling)}$$

2. Frequency shift Keying

In the event that the channel does not permit a coherent detector to be built; i.e., the received signal's phase is a random variable, $-\pi \leq \phi \leq \pi$, an optimum noncoherent receiver may be built. In this case decisions can be based on the magnitude of the envelope of the received signal (since the phase information has been destroyed by the channel).

The transmitter can elect to send one of two frequencies (frequency-shift keying, FSK), to represent the binary source data. Such a selection causes the normalized inner product, p, of the signals to be zero (referred to as orthogonal signaling) and the performance is given by:

$$P_E = \frac{1}{2} \exp\left(\frac{-E_b}{2N_o}\right) \quad ; \text{(noncoherent, orthogonal FSK signals)}$$

*E_b = energy per bit = E for these systems, since there is one bit per symbol.

For coherently detected orthogonal signal sets, such as coherently detected frequency shift-keyed signals, the performance becomes:

$$P_E = Q\left(\sqrt{\frac{E_b}{N_o}}\right) \quad ; \text{(coherent, orthogonal FSK signals)}$$

For all but high P_E cases the E_b/N_o required to achieve a given P_E with noncoherent reception is about twice that required with coherent reception, ie 3 dB performance difference.

3. Differently Coherent PSK

An intermediate case, referred to as "differently coherent PSK," is also of interest with binary digital systems. Unlike the preceding systems, in differentially coherent PSK (DPSK), a binary one dictates a change in phase in the transmitted signal and a zero dictates no change in phase. The receiver requires two sets of detectors with overlapping operation; while one operates on the received signal during even intervals forming decisions on even-numbered data symbols, the other operates during the odd intervals forming decisions on the odd-numbered data symbols. In effect the phase of the preceding symbol serves as a reference for the present symbol. The performance of DPSK is given by:

$$P_E = \frac{1}{2} \exp\left(\frac{-E}{N_o}\right) \quad ; \text{(DPSK)}$$

It should be noted that for a given P_E exactly one half the energy is required in a DPSK system relative to the noncoherent orthogonal signal case (noncoherent, FSK).

Performance curves for each of the binary digital systems discussed above are shown in Figure 59. Finally, the bandwidth expansion factor for these signaling schemes can be shown to be:

$$m = \frac{B}{R} = \begin{cases} 2; \text{ orthogonal (FSK)} \\ 1; \text{ antipodal (PSK \& DPSK)} \end{cases}$$

Generalizations of these basic concepts to higher-order alphabet systems (M-ary) are conceptually easy. Digital communications systems of this type use K consecutive symbols of the data sequence to select and transmit one of $M = 2^K$ stored signals (K = 1 for binary systems). The performance obtained requires less E_b/N_o than for the binary case but requires an exponential increase in bandwidth and receiver complexity relative to the binary signaling schemes.

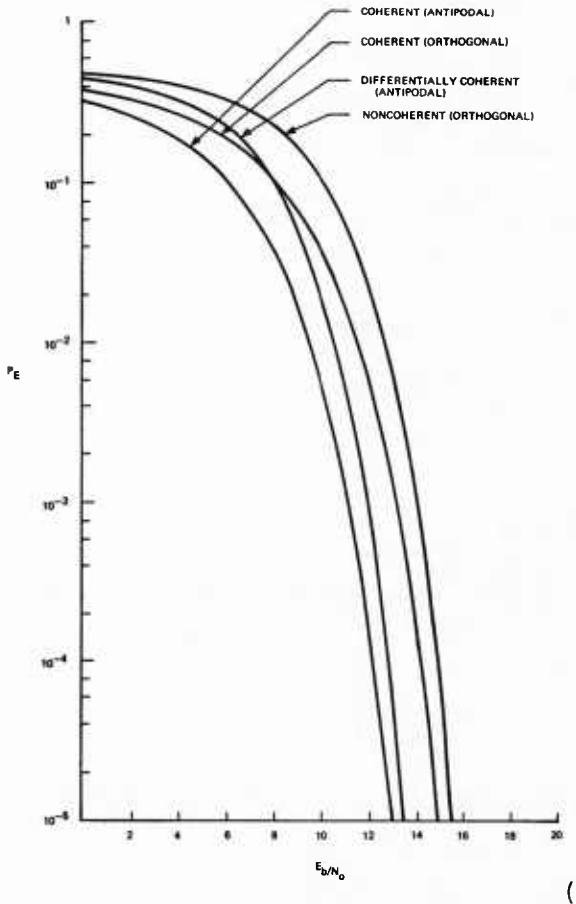
4. Choice of Modulation

The characteristics of the propagation channel can influence the choice of modulation. PSK modulation offers better channel efficiency in a high quality channel than FSK. PSK can provide the same bit-error-rate at a lower received signal level than FSK or it can provide a better error rate at the same received signal level as FSK.

However, FSK modulation tends to be more robust than PSK when the channel suffers phase distortion such as during severe ionospheric scintillation. Especially in the case of nuclear induced ionospheric scintillation, the phase of the earth-satellite link is expected to be perturbed to such a degree that both coherent and differential PSK will encounter an unacceptable error rate. Noncoherent FSK will suffer some degradation but not nearly as severe as PSK. Analysis of natural ionospheric by Prettie, 1981 showed that during severe polar scintillation, low data rate PSK would suffer an irreducible error rate (regardless of signal strength) due to the rapid phase changes.

B. Coding Techniques

In the preceding discussions on the performance of digital modulation/demodulation (modem) techniques, the probability of bit error (P_E) for a given energy per bit-to-noise-density (E_b/N_o) ratio was used as a measure of how well the modem could operate (extracted from DCA Handbook, 1972). A modulation technique that satisfied the system probability of bit error (P_E) specification without requiring an excessive E_b/N_o ratio (E_b = ratio of average received carrier power to data rate, P/R_a) or bandwidth allocation was deemed an acceptable choice. Furthermore, since the performance of an ideal digital system could be calculated under certain conditions, it was possible to assess how any



(From DCA, 1972)

FIGURE 59. ERROR PROBABILITY VERSUS Eb/No FOR VARIOUS MODULATIONS

particular choice of modem deviated from the ideal and to quantitatively measure how much loss was incurred relative to the ideal system.

To specify the ideal system's performance it was necessary that the channel, or modem-transmission medium combination be linear, restricted to bandwidth B, average power P, and be perturbed only by additive white Gaussian noise (AWGN), with single-sided noise power density N_0 (watts/Hertz). These considerations are of more than academic interest since many satellite communications systems can be modeled as an AWGN channel. It is precisely for this reason that Forward acting Error Correcting (FEC) coding techniques can and should be applied to satellite communications system designs. These FEC techniques allow a system design that better approximates the ideal system's performance.

In general, the coding technique changes each information bit--input from the source--into a sequence of binary digits or symbols before transmission over the satellite channel. The coded sequence or representation of the information bit has a structure that permits the complementary decoding technique at the receiver to properly identify the transmitted information bit despite errors that occur in the coded sequence because of the channel noise conditions.

The set of rules used to represent an information bit by a sequence of symbols at the transmitter is referred to as an encoding technique and the corresponding set of rules used to unscramble the received sequence to an information bit is referred to as a decoding technique or algorithm. When the encoding-decoding technique does not rely on retransmission strategies to decode an information bit, it is referred to as an FEC coding strategy or technique. These FEC techniques are distinctly different from Automatic Repeat Request (ARQ) techniques that rely on detecting the presence of an error in the received sequence (only detecting not correcting) and automatically requesting a retransmission of any bit or sequence that has been detected to contain errors. In satellite communications system applications, ARQ techniques are of only limited utility since the long propagation delays (¼ second) and occasionally high error rates due to propagation problems can severely limit the throughput of ARQ systems. Consequently, only FEC techniques will be treated in the remainder of this discussion.

1. Block Encoding

The placement of an FEC encoder-decoder in a typical communications system configuration is indicated in Figure 60. Thus, from the FEC coder's standpoint, the channel includes the items shown between the dotted lines of Figure 60. There are two broad classes of error-correcting codes; namely, block codes and convolutional codes. With block coding techniques each group of K consecutive information bits is encoded into a group of N symbols for transmission over the channel. Normally, the K information bits are located at the beginning of the N symbol block code and the last N-K symbols correspond to parity (check) bits formed by taking the modulo 2 sum of certain sets of the K information bits. Block codes exhibiting this property are referred to as systematic block codes.

The term block code stems from the fact that each block of N symbols corresponds to a particular group of K information bits. The encoded symbols for the $K+1^{st}$ bit and beyond are completely independent of the symbols generated for the first K information bits, and hence cannot be used to help decode the first group of K information bits at the receiver. Because N symbols are used to represent K information bits, the (code) rate (R) of such a block code is K/N bits per symbol (R = K/N). The encoder structure for the (N = 7, K = 4) binary code is shown in Figure 61. The information bits are stored in the K = 4 storage

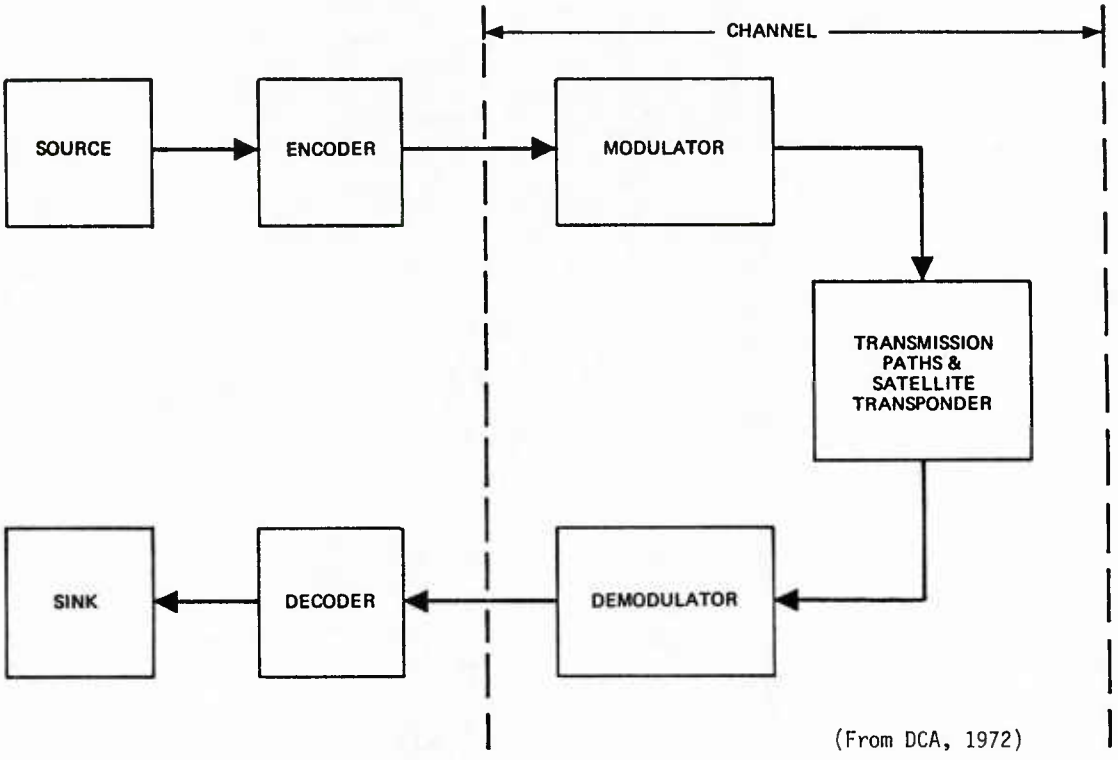


FIGURE 60. TYPICAL CODED COMMUNICATIONS SYSTEM BLOCK DIAGRAM

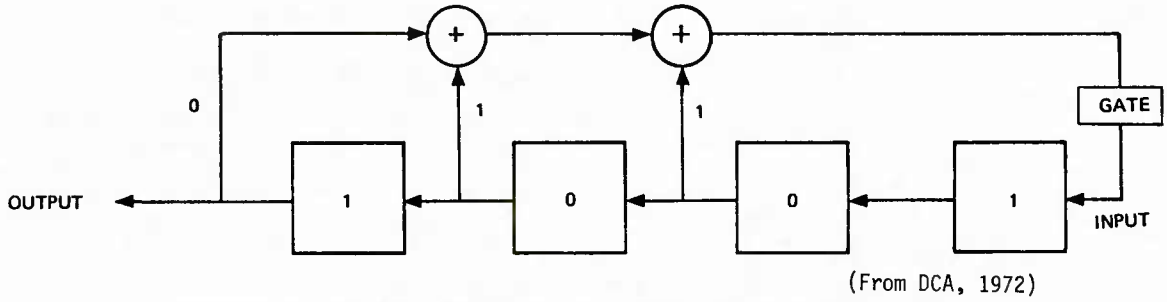


FIGURE 61. SHIFT REGISTER FOR ENCODING THE (7, 4) CODE

devices, and then the device is made to shift N = 7 times. The first K symbols that come out will be the information symbols, and the last N-K symbols will be a set of check symbols that form the whole N symbol word. A block code is often denoted with the symbols (N, K, t); N corresponds to the block length, K to the number of information symbols in the work, and t to the number of errors in a block of N symbols that the code is guaranteed to correct. The code of Figure 61 is a (7, 4, 1) code in this notation.

2. Convolutional Encoding

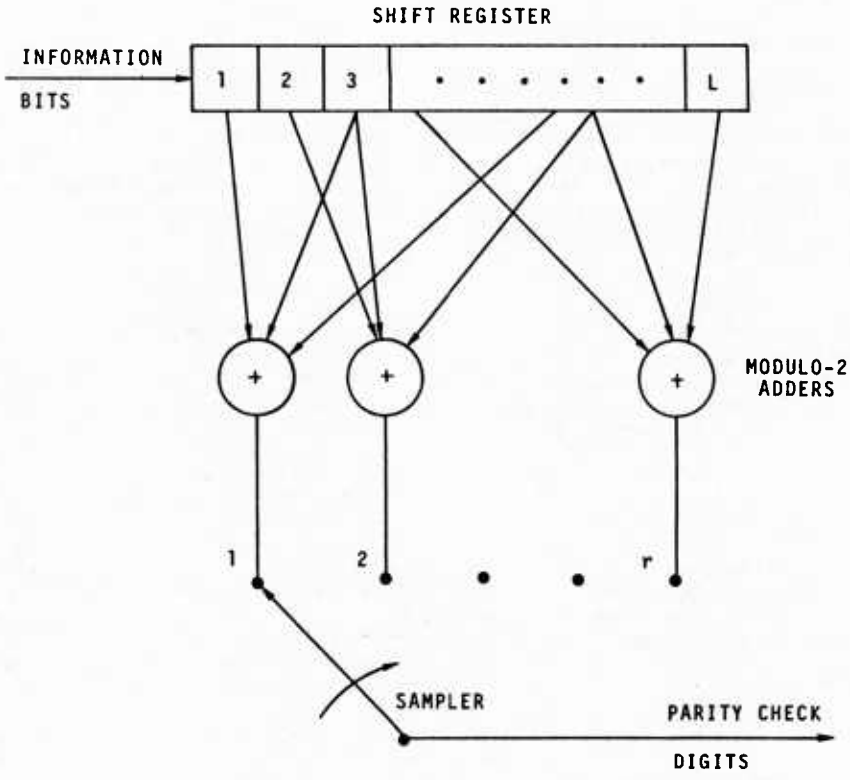
The block diagram of a typical convolutional encoder is shown in Figure 62. Information bits are shifted to the right through the L stage shift register as new information bits enter from the left. Bits out of the last stage of the shift register are discarded. The bits are shifted one position each T seconds, where 1/T is the information rate in bits per second. The modulo-2 adders are used to form the parity check bits, each of which is a binary function of a particular subset of the information bits in the shift register. The parity check bits are seen to depend on a sequence of L information bits, so that the constraint length of the code is L.

Once each T seconds, the terminals labeled 1, 2..., r are sampled in succession. Thus for each information bit fed into the encoder, there are r parity check bits at the output. These output parity check bits are referred to as symbols. Since each symbol carries an average of 1/r information bits, the code is said to have rate 1/r.

When the first modulo-2 adder is replaced by a direct connection to the first stage of the shift register, the first symbol becomes a replica of the information bit. Such an encoder is termed a systematic convolutional encoder.

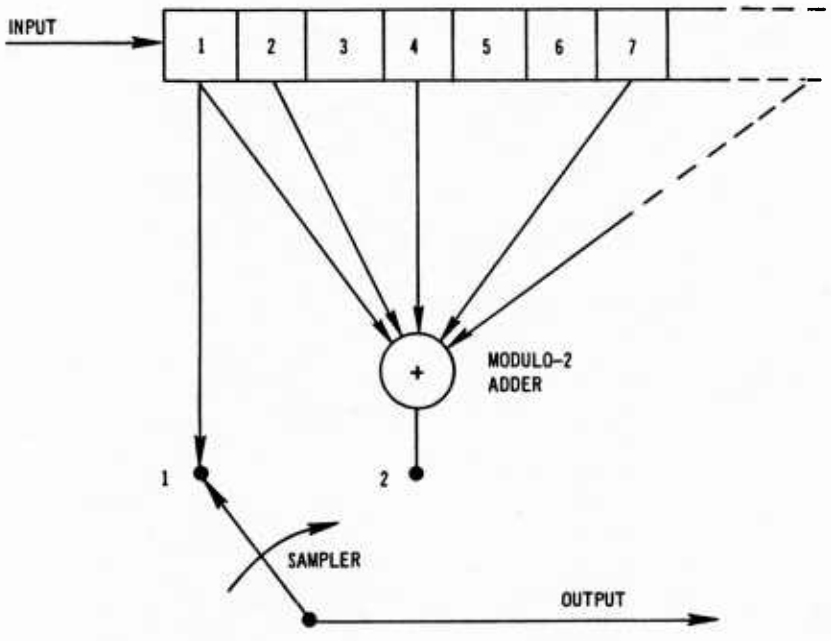
The code can be thought of as forming a tree structure. At each node the information bit determines which direction will be taken; up for a 1 and down for a 0. The r digits occurring on the branch selected to correspond to the output symbols. A particular sequence of information bits then describes a particular path through the code tree. This will be illustrated by the following example.

Consider the rate ½ encoder of Figure 63 and let the shift register contain all zeros initially. Now assume that the input sequence of information bits is 1100.... The path defined by this sequence is



(From DCA,1972)

FIGURE 62. CONVOLUTIONAL ENCODER



(From DCA, 1972)

FIGURE 63. ENCODER USED TO GENERATE CODE TREE

illustrated as the heavy line in the code tree of Figure 64. Since the code generated by this encoder is a systematic code, the first digit of the output at each branch is the information bit, as shown in Figure 63. The second digit is the output of the modulo-2 adder in Figure 63. After the first information bit is fed into the shift register, the contents are 1000000..., so that the output of the adder is a 1 and is shown as the second digit on the first branch of the code tree. After the second input to the shift register, the contents are 1100000... and the adder output is a 0; this is easily seen since the modulo-2 sum of an even number of 1s is a 0. Continuing in this fashion, the remaining portion of the tree can be constructed. It should be noted that unlike block codes, convolutional codes have no formal block structure in the generated code words. That is, past information bits do have an influence on the symbols used to represent a present information bit. The code can be constructed by taking the convolution of the shift register tap connection set (code generating polynomial) with the information bit pattern; hence the name convolutional codes.

3. Decoding Techniques

Decoding algorithms for block and convolutional codes are usually quite different. The formal structure of the block encoded words permits decoding by taking advantage of the known structural properties of the words or the algebraic nature of the constraints among the symbols used to represent an information sequence. Examples of decoding algorithms for block codes are the algebraic decoding techniques of Peterson, Chien and Berlekamp for the Bose-Chaudhuri (BCH) codes. Decoding of convolutional codes, on the other hand, is often done using the probabilistic techniques of Wozencraft (sequential decoding) or Viterbi (Viterbi decoding). These latter techniques rely not on the algebraic structure of the code words but on the ability to home in on the correct sequence by designing efficient search procedures that discard unlikely sequences very quickly. As an example, consider a sequential decoding algorithm for convolutional codes. The symbols presented to the decoder are the symbols generated as described in Figure 62 after they have been corrupted by noise. In a communications system, the output of the demodulator would constitute the decoder input. Assume that the symbols presented to the decoder are 1s and 0s, each symbol being in error with probability p (i.e., assume a binary symmetric channel).

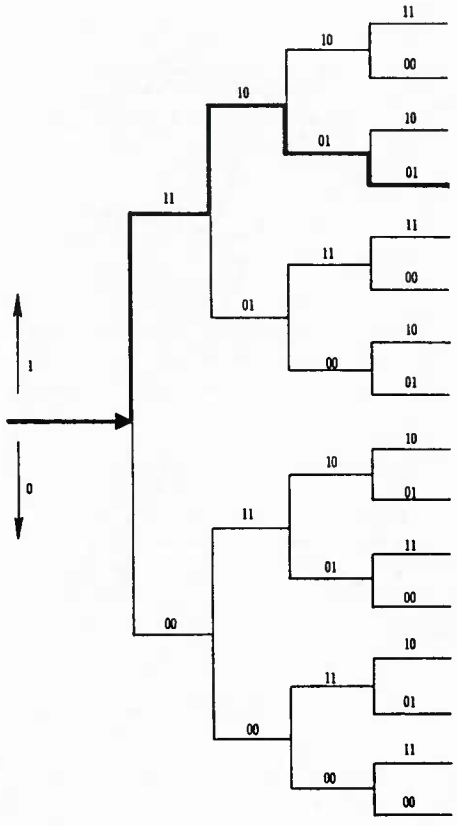
The decoder contains a duplicate of the encoder, whose shift register contains the received information bits. By comparing the input-code-plus-noise symbols at the input to the decoder with the symbols found at branch points of the tree as in Table 2 the decoder attempts to find the most probable path through the code tree.

The sequential decoder differs from most other types of decoders in that, when it finds itself on a wrong path in the tree, it has the ability to search back and forth, changing previously decoded information bits, until it finds the correct path. The frequency with which the decoder has to search back, and the depth of these backward searches, is dependent on the value of the channel error probability (p).

Table 2. Sequential Decoding Procedure

Message	1	1	0	0		Notes
Code	11	10	01	01		Fig. 64
Noise	00	10	00	00		Assumed
Input	11	00	01	01		Code + Noise
Decision					Total	
1. Go To	11					
	00				0	determinate
2. Try	11	10				
	00	10			1	
Try	11	01				
	00	01			1	same as above
3. Try	11	10	10			
	00	10	11		3	
Try	11	10	01			
	00	10	00		1	lowest possible
Try	11	01	11			
	00	01	10		2	check
Try	11	01	00			
	00	01	01		2	check
4. Go To	11	10	01	01		
	00	10	00	00	1	lowest possible
Try	11	10	01	10		
	00	10	00	11	3	check

NOTE: Total is sum of digits in sum modulo 2 of Input and line above Total.
⊖ = modulo 2 addition; e.g., 1 ⊖ 1 = 0



(From DCA, 1972)

FIGURE 64. A CODE TREE

One of the most important properties of a sequential decoder is that, if the chosen constraint length is large enough, the probability that the decoder will make an error becomes negligible (e.g., less than 10^{-9}). The type of error that becomes significant is the occurrence of an overflow, which is defined as being a situation in which the decoder is unable to perform the necessary number of computations in the performance of the tree search. To be more precise, a computation is defined as having occurred when the decoder examines a path through the decoding tree. Since the decoder is able to search back and forth through the tree, and does so according to how the errors arrive at the decoder input, the number of computations the decoder must make to decode one information bit is a random variable. An important parameter in the system then becomes the average number of computations per decoded information bit. As long as the channel error probability (p) is not too high, the probability of decoder overflow will be acceptably low and thus satisfactory performance results.

The performance of block decoding algorithms, however, is determined by the number of errors that the code is guaranteed to correct (t) in a block of N received symbols. If the channel error probability (p) is too high then the probability of obtaining t - 1 or more errors in a block of N symbols becomes significant and hence the decoder will fail in its attempt to identify the transmitted code word.

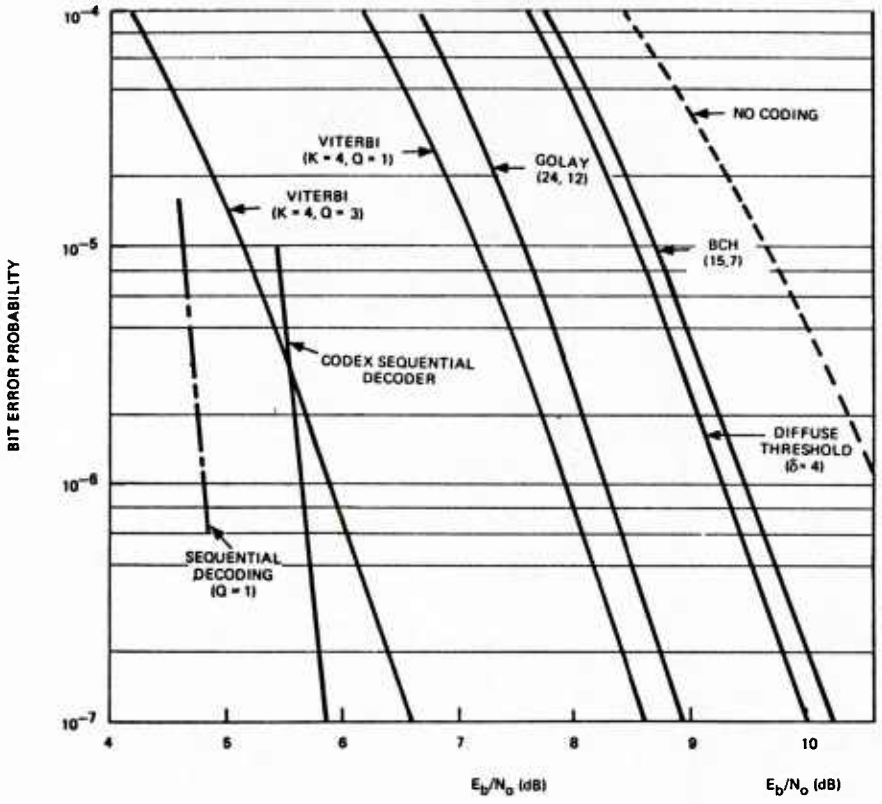
Up to this point, it has been assumed that the demodulator supplies only hard decisions (that is, the demod output has been hard-limited) to the decoder. If instead the demodulator output is quantized into 4 or 8 levels (2- or 3-bit quantization, respectively), certain decoding algorithms can use this additional information to obtain a lower probability of output bit error than if supplied only with hard decisions. Both the sequential and Viterbi decoding algorithms can effectively use this soft-decision demodulator information, giving these algorithms a distinct advantage over algebraic decoding techniques that cannot easily account for this added information in making a final decoder decision.

To gain some insight into the performance of error-correcting codes on random error channel's Figure 65 is plotted. Five codes of moderate complexity are shown. These are: BCH (15, 7), Golay (24, 12), Viterbi (K=4), diffuse threshold (K=4), and sequential decoding.

The curve for the sequential decoder is based on a rate ½ code operating on a hard-decision Binary Symmetric Channel (BSC). The curve is shown as a nearly vertical line to emphasize the relative insensitivity to undetected errors. One of the curves for a Viterbi decoder is for 3-bit quantization (Q = 3) and the other is for the hard-decision case (Q = 1). The remaining curves are based on calculations. Several observations from Figure 65 are worthy of note.

For the parameters shown, the Viterbi decoder with Q = 1 provides performance exceeding that of the block codes and the threshold decoder. The advantage of the Viterbi decoder is even more apparent when quantization is considered. Using a quantized version of the demodulator output, in lieu of hard decisions, improves performance on the order of 2 dB. This improvement may not be directly realized with block codes.

Another pertinent observation from Figure 65 is the increased slope of the sequential decoder curve. This may be particularly important in regard to performance requirements on the order $P_E\ 10^{-10}$.



(From DCA, 1972)

FIGURE 65. ERROR CORRECTING PSK CODE PERFORMANCE

As an example of how effective FEC techniques can be in a satellite communications system, consider the performance of the coded systems in Figure 65 relative to the no-coding (ideal PSK demodulation) system also shown in this Figure. For a 10^{-5} output bit error specification, the improvement in required E_b/N_o for these systems is shown in Table 3. As shown, the largest coding gain is supplied by the sequential decoder, 5 dB. This means that for a fixed received average power to thermal noise density ratio (P_r/N_o), the sequential decoder system can support a 5-dB higher data rate (3.2 times more data) than the uncoded system, or correspondingly, the same data rate as the uncoded system, or with 5 dB less average power. Furthermore, if even lower output probability of bit error rates are required than 10^{-5}, the gain of the coded systems (particularly sequential decoding) relative to the uncoded system is even more impressive. Finally, it should be emphasized that this sizable gain comes at the expense of only a 2 to 1 increase in channel bandwidth. All of the coded designs shown operate at a code rate of ½ (bits/symbol) and hence the bandwidth required, for a fixed modulation method, is twice that of the uncoded system design. That is, the symbol rate is 2 times the data rate with a rate ½ bit per symbol code. This is a very moderate price to pay for the dramatic savings in satellite power obtained with sequential or Viterbi decoding.

Table 3. E_b/N_o Coding Gain for Several Coded Systems

System	Required E_b/N_o at 10^{-5} (dB)	Gain (E_b/N_o Uncoded - E_b/N_o Coding) dB
No Coding	9.6	-
(15,7) BCH	8.65	1
Threshold	8.5	1.1
Golay (24,12)	7.5	2.1
Viterbi (Q=1)	7.1	2.5
Viterbi (Q=3)	5.1	4.5
Sequential (Q=1)	4.6	5

C. Antenna Characteristics

One of the major factors the system designer have control over is the choice of antenna. The proper choice can improve the communications link margin, reduce the effect of multipath fading, and minimize self-jamming or mutual interference. Four generic types of antennas are illustrated in Figure 66, (Johnson 1979-2).

1. Horizon Coverage

A typical antenna that provide horizon coverage is a ¼ wave stub or blade antenna. The blade provides a donut-shaped coverage, illuminating the horizon with a null overhead, Figure 66. Since the blade has most of its gain toward the reflecting surface it exhibits a large illumination factor and is subject to severe multipath fading. A typical azimuth pattern for a blade mounted on an aircraft is shown in Figure 67. The primary advantage of the blade antenna is it simplicity, ease of installation, and low cost.

2. Over-Head Coverage

Another common satellite communications antenna for mobile terminals is the over-head coverage crossed dipole antenna. Its coverage pattern is a cone centered on zenith, Figure 66. The overhead antenna pattern characteristics of a cross-dipole antenna give considerable protection to the multipath reflective component when the satellite is at a high elevation angle. The azimuth coverage pattern of the crossed-dipole antenna is shown in Figure 68. The crossed-dipole tends to be a little larger and more costly than the blade antenna but is a good choice for high angle coverage on a mobile terminal.

3. Omni-Directional Coverage

A compromise antenna that provides both horizon and over-head coverage can be use if it is not possible to install two separate antennas on the mobile terminal. An antenna such as the MIT Lincoln Laboratory designed crossed-slot antenna or the Rockwell designed drooping dipoles give reasonable coverage over the upper hemosphere, Figure 66. The azimuth pattern of the crossed-slot antenna is shown in Figure 69. This type antenna is subject to the severe multipath fading of the blade antenna and provides less over-head gain than the crossed dipole. This antenna tends to be larger, more expensive, and more difficult to install than the blade or crossed dipole.

4. Directional Coverage

A fourth generic antenna type is the directive antenna, primarily used at the microwave frequencies. The directive antenna gives good protection against the reflected signal unless the satellite is right on the horizon. The major disadvantage of the directive antenna for mobile applications is the need to continuously point the antenna. Directive antennas tend to be heavy, expensive, and difficult to install. An airborne microwave satellite communications antenna may cost $500,000 to purchase and another $500,000 to install in an aircraft. However, for microwave satellite communications applications there are usually no cheaper choices.

D. Network Considerations

In order to interconnect the various communications satellite users, network protocols must be established (extracted from DCA Handbook, 1972). Historically, the first network protocols used were single carrier per channel with frequency spacing between channel to separate users. Increased demand for the limited satellite capacity prompted the development of more efficient multiple access techniques. One of the major considerations was to minimize the interference with other users of the transponder. The primary multiple-access (MA) techniques that are used to accomplish the above purposes in communications satellites are:

* Frequency Division Multiple Access (FDMA)
* Time Division Multiple Access (TDMA)
* Code Division Multiple Access (CDMA - sometimes called Spread Spectrum SSMA)

Table 4 summarizes the characteristics, advantages and disadvantages of these techniques.

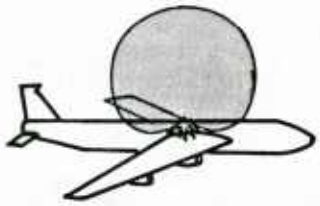
1. Frequency Division Multiple Access

In FDMA each earth terminal is assigned a carrier frequency within the transponder bandwidth. The carrier frequencies must have a minimum spacing large enough so that the modulated spectra do not overlap. An earth terminal receiver is tuned to the downlink frequency corresponding to the uplink frequency of the desired transmitter. The transponder power is shared by all the links simultaneously transmitted. Two important considerations in the use of FDMA are efficient use of the available transponder power and avoidance of interference among the simultaneous users. If a linear transponder is used, then the carrier frequencies can be spaced at the minimum spacing referred to previously. However, individual earth terminal power and/or maximum number of users must be controlled to avoid driving the output power device into saturation. If this happens, nonlinearity in the power output device will cause intermodulation products of the signals to rise to an unacceptable level.

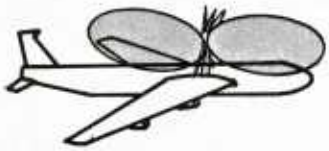
If a hard-limiting transponder is used, intermodulation products will be formed in the limiter. The carrier frequencies may be spaced so that the intermodulation products do not fall on any of the carriers. It has been shown that for equally spaced carriers and a limiting transponder the signal-to-intermodulation noise power ratio (measured in the bandwidth of the signal) will vary from 9 to 11 dB depending on the spacing of the carriers. In either case, the number of accesses possible in the transponder bandwidth is reduced. In addition, the intermodulation products use some output power of the transponder, reducing the useful output power. Finally, the small signal suppression effect of a limiter makes power control of the earth transmitters necessary.

2. Time Division Multiple Access

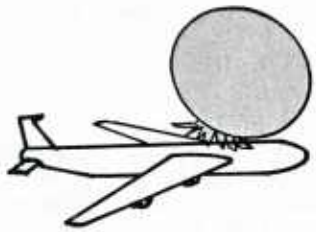
In TDMA multiple access is accomplished as a time-gating function that correctly locates each earth terminal transmission burst relative to those of the other terminals. Thus each earth terminal is assigned exclusive use of the transponder during specified time slots and the transmissions do not overlap. An access channel in TDMA designates a particular sequence of time slots. The simplest sequence would allow the time slot of a channel to occur periodically at a definite repetition frequency, call the frame repetition rate of the system. All other channels would have the same frame rate but different times of occurrence, and possible different slot durations.



OMNI COVERAGE ANTENNA
(AFSAT)



HORIZON COVERAGE ANTENNA
(BLADE)



OVERHEAD COVERAGE ANTENNA
(CROSSED DIPOLE)

DIRECTIVE ANTENNA
(DISH)

(From Johnson, 1979-2)

FIGURE 66. TYPICAL ANTENNA PATTERN COVERAGE

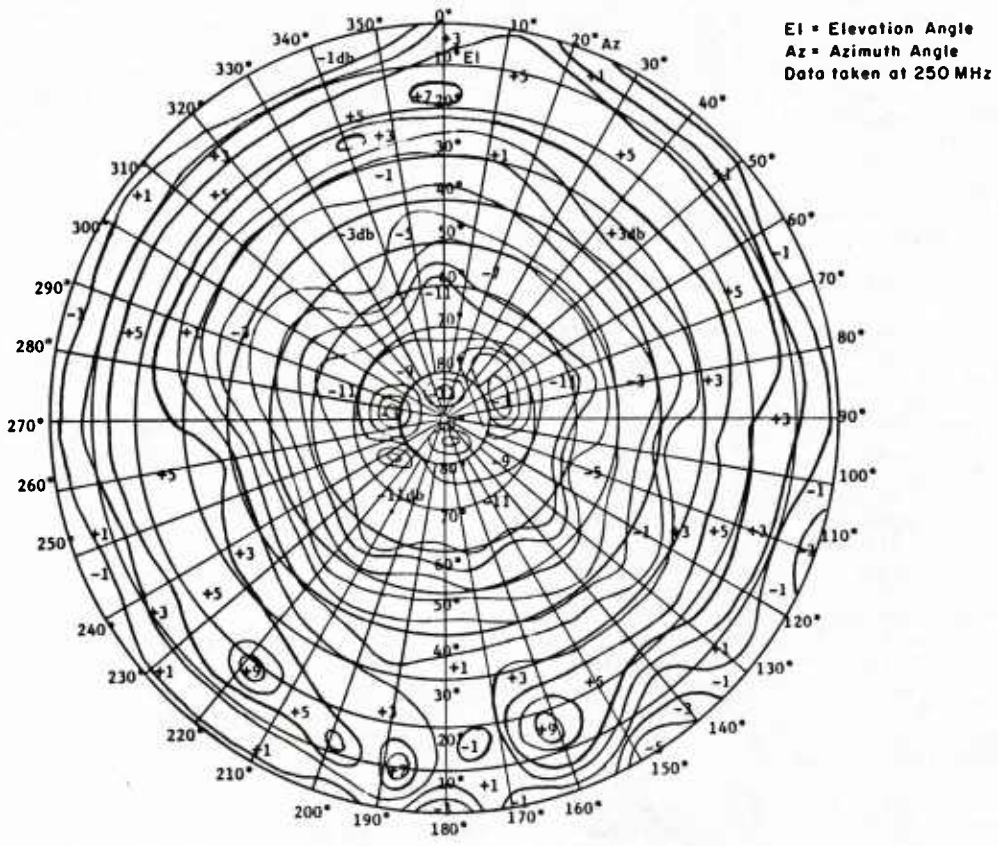


FIGURE 67. CONSTANT GAIN CONTOURS FOR UHF BLADE ANTENNA

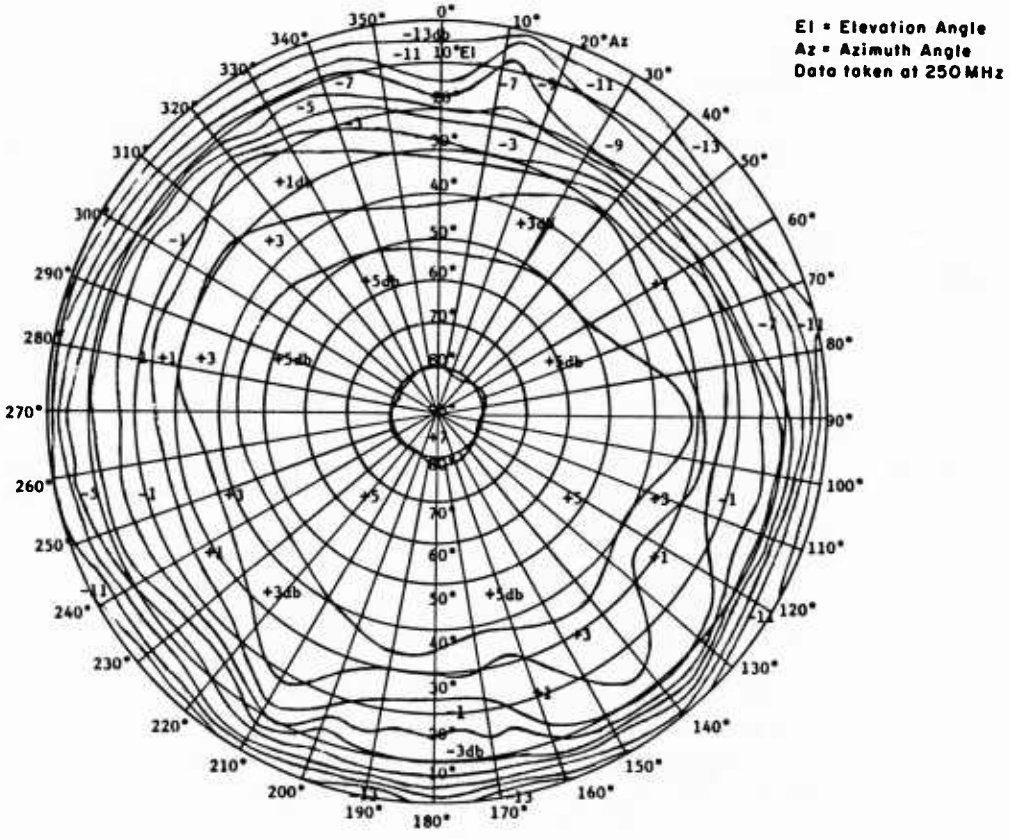


FIGURE 68. CONSTANT GAIN CONTOURS FOR UHF CROSS-DIPOLE ANTENNA

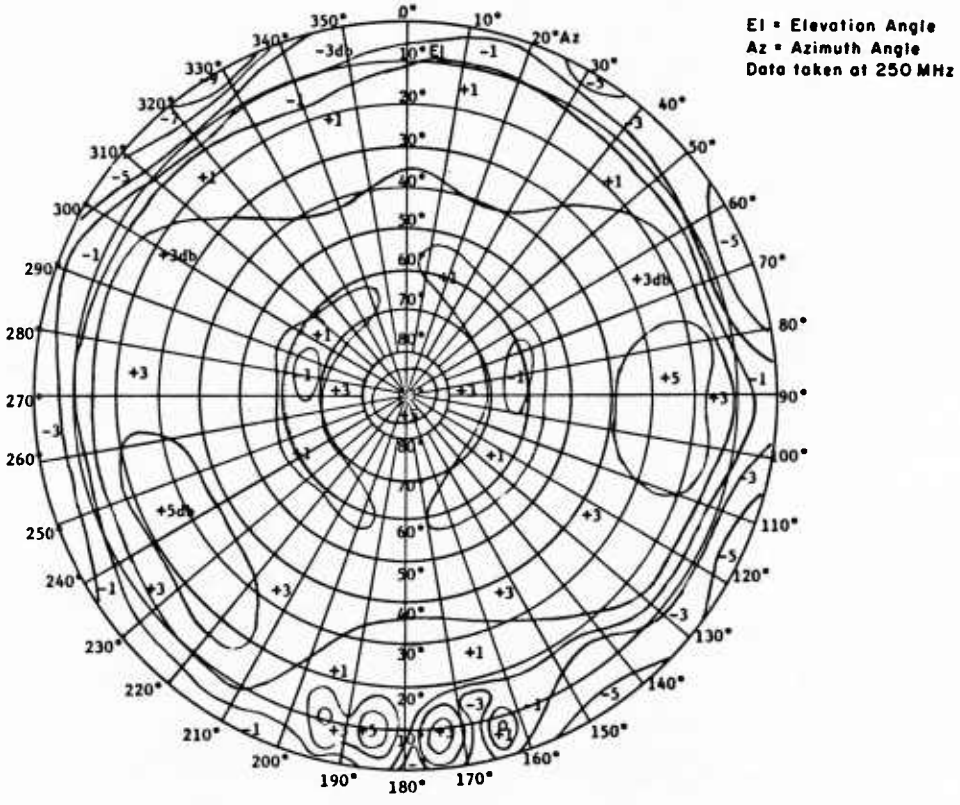


FIGURE 69. CONSTANT GAIN CONTOURS FOR UHF CROSS-SLOT ANTENNA

Table 4. Summary of Multiple-Access Techniques

Techniques	Characteristics	Advantages	Disadvantages
FDMA	Constant envelope signals Signals confined to non overlapping frequency bands MA demultiplexing by filtering Message information by angle modulation	Uses existing hardware No network timing	Intermodulation in repeater Requires uplink power control
TDMA	Signals from different links never present simultaneously in satellite MA demultiplexing by time gating Message information by angle modulation within carrier burst	High efficiency in using satellite power Does not require uplink power control Peak power transmitter	Network timing required Analog to digital conversion required
SSMA	Constant envelope carriers Transmitted spectrum is spread over satellite bandwidth MA demultiplexing by correlation with local replica of code Message information by angle modulation	No network timing required Can use fixed-address assignments Minimizes suppression in hard limiter	Link synchronization required Uplink power control required

TDMA offers higher performance and more flexibility than FDMA or CDMA. Since only one carrier is present in the transponder, there is no power loss or interference due to intermodulation. The power output device can be run at saturation, providing higher output power (typically 3 dB) and efficiency. There is no need for earth terminal power control.

TDMA places stringent requirements on system timing. System timing accuracy affects required guard times between adjacent transmission bursts from earth terminals. Guard time is one of the components of overhead time in a TDMA system. The other components of guard time are in the preamble of each burst, carrier synchronization time, bit synchronization time, carrier phase ambiguity resolution time, and word synchronization time. In the DSCS Phase II TDMA system, the goal for time utilization efficiency is 95 percent, where the efficiency is the ratio of time during which message information is transmitted to total frame time. TDMA rates are limited by transponder bandwidth, burst acquisition times, and practical data storage capabilities.

3. Code Division Multiple Access

In CDMA (or SSMA), each multiple-access carrier signal usually occupies all or a large part of the transponder bandwidth. In this technique a carrier is modulated with two different signals. The bandspreading modulation can be either analog or digital. When the bandspreading is accomplished by phase-shift keying it is often referred to as Pseudo Random Noise (PRN) and when frequency shift keying is employed the term Frequency Hopping (FH) is used. To date the greatest consideration has been given to digital techniques. The PRN bandspreading modulation results in a noise-like carrier with over 90 percent of its power in a bandwidth twice the modulation rate. The information modulation, at a rate typically one-tenth or less than that of the bandspreading modulation, is impressed as an additional phase or frequency variation.

The pattern of the bandspreading modulation is called its code. At the intended receiver a code identical to the transmitted code is generated. This is synchronized with the received code and allows detection of the information. By proper choice of codes other SSMA transmission simultaneously present at the receiver will cause relatively small interference. The code are then said to exhibit low cross-correlation.

For frequency hopping spread spectrum systems the carrier frequency is hopped in a pseudo random manner across the RF bandwidth. The FH signal is hopped at a rate of a few hops per second to thousands of hops per second depending on the system design.

The CDMA technique through a transponder satellite requires uplink power control as in FDMA; otherwise one or more signals will use too much of the transponder power output. However, neither use of a limiter

nor a saturated output device will result in interference from intermodulation products. The receiver must use an exactly matching replica of the transmitted code which presents the problem of synchronizing the receiver before message transmission can begin. This requires an initial synchronization searching procedure by the receiver and is one of the critical aspects of CDMA.

4. Fixed Channel Assignments

The simplest multiple access protocol is to provide fixed channel assignments to each user. If a user has a fixed frequency, time slot, or code assignment, the resource is always available and other users can always reach him. In high priority applications such as the US Navy's fleet broadcast, or the US Air Force's force direction network, this protocol meets the 100% availability need. Other situations where fixed channel assignments may be warranted include lightly used satellite systems and for disadvantages users.

5. Demand Assignment Multiple Access

An efficient channel assignment protocol is Demand Assignment Multiple Access, DAMA (Spilker 1977). DAMA may be used with FDMA, TDMA or CDMA systems. The commercial carrier satellite companies make use of an FDMA demand assignment system called SPADE (Single-channel per carrier Pulse-code modulation multiple-Access Demand assignment Equipment). The user requests service over an order-wire channel and is given a frequency assignment. To improve the efficiency, the frequency is reassigned during long pauses or when the user is listening. This rather complicated protocol is warranted to increase the efficiency and reduce the cost for commercial telephone calls.

A TDMA application of DAMA is being implemented by the US Navy. In the Navy DAMA scheme, order wire time slots are set aside in the DAMA frame. A tactical terminal requests service, indicating its priority and approximate message length. The DAMA controller reviews the current assignment and if service is available at the users priority level, a channel (time slot) assignment is sent back over the order wire time slot. The users equipment recognizes the assignment and transmits his message when the time slot occurs.

DAMA offers a significant increase in channel loading efficiency when large numbers of users with occasional, short messages are interconnected. Compared with fixed channel assignment, a DAMA system can provide timely communications for 10 to 100 times more users than a fixed channel assignment system. The primary penalty for this efficiency improvement is increase equipment complexity and equipment cost.

E. Prediction Techniques

Propagation modelling is well developed in the areas of ray tracing, multipath prediction, and atmospheric attenuation. The prediction or modelling techniques for ionospheric scintillation are still in their infancy. The current modelling knowledge will be reviewed in this section.

1. Scintillation Modelling

Over a period of years, starting from available data and from weak scintillation theory, a model of scintillation termed WBMOD has been developed by Fremouw and others with attempts to satisfy propagation theory and incorporate available observations (Sagalyn et al, 1974; Fremouw & Bates, 1971; Fremouw & Lansinger, 1981; Fremouw & Rino, 1978; Fremouw, 1980-1; Fremouw & Rino, 1976; Fremouw et al, 1977), (extracted from Aarons, 1982). The program provides for phase and amplitude information. Input parameters include frequency, location, local time, sunspot number, and planetary magnetic index K_p. The user also must specify the longest time the system needs phase stability. Scintillation indices are the output. A model of the irregularity drift velocity is contained in the program.

Program WBMOD permits a user to specify his operating scenario. The code returns the spectral index p for power-law phase scintillation, the spectral strength parameter T, the standard deviation σ_ϕ of phase, and the intensity scintillation index S_4, as functions of a changing independent variable chosen by the user.

The theory employed in WBMOD is based on the equivalent phase-screen representation of Booker et al, 1950, formulated to account for three dimensionally anisotropic irregularities (Singleton, 1970) described by a power-law spatial spectrum. The formulation employed was developed by Rino, 1979, in the infinite outer-scale limit, but a means for dealing with the effect of a finite outer scale on phase scintillation has been incorporated in WBMOD. Similarly, a means has been provided for accommodating multiple-scatter effects on intensity scintillation that should suffice for practical applications.

The descriptive irregularity model is based on numerous observations (Fremouw & Bates, 1971; Fremouw & Rino, 1978), but most particularly on observations of phase scintillation performed in the DNA Wideband Satellite Experiment (Sagalyn et al, 1974). The most significant caveat about use of WBMOD, however, is that it has been calibrated quantitatively against data from the Wideband satellite obtained from only a single station in the northern auroral zone (Poker Flat, Alaska). The descriptive model was developed by iterative comparison with most of the Wideband data population from Poker Flat, with a portion of the population reserved for final comparative tests.

The basic calculations are made of two central quantities T and p. T is the spectral strength of phase at a fluctuation frequency of 1 Hz, and p is the power-law spectral index of phase. T is highly variable, unlike p. The program calculates T and p and the two commonly used indices of scintillation activity based on them, one for phase σ_ϕ and one for intensity S_4.

In order to calculate T, p, σ_ϕ, and S_4, one must have values for eight parameters describing ionospheric irregularities. They are 1) the height h; 2) vector drift velocity V_d of the irregularities; 3) an outer scale a; 4, 5, 6, 7) four "shape" parameters describing the irregularities' three dimensional configuration and spatial "sharpness", a, b, δ, and v; and 8) the height integrated spectral strength C_sL. Program WBMOD contains models for the foregoing eight parameters, but the degree of detail is very much less for some than for others.

The most variable and, probably, the most important of the eight is the height-integrated strength C_sL. The irregularity strength is modeled by:

$$\sqrt{C_sL} = E(\lambda_m, \lambda_g, T, D, \bar{R}) + M(\lambda_m, T) + H(\lambda_m, T_m, K_p, \bar{R})$$

where λm = geomagnetic invariant latitude
λg = geographic latitude

T = local meridian time
D = day of the year
R = smoothed Zurich sunspot number
T_m = geomagnetic time
K_p^m = planetary geomagnetic activity index.

The three terms in this equation respectively describe the strength of equatorial, middle-latitude, and high-latitude irregularities. The first two have not been tested extensively against Wideband data but H, the high latitude term has.

The high-latitude term is based on the observation that often there is a more or less abrupt boundary (Aarons et al, 1969), between the middle-latitude region of relatively smooth ionosphere and the high-latitude region. It is located, typically, equatorward of discrete-arc aurora in the general vicinity of the diffuse auroral boundary. The underlying form of H stems from the supposition that the instantaneous boundary latitude is normally distributed about a mean value λ_b for a given set of T_m, K_p, and R. This supposition, together with other considerations to be discussed shortly, yields the following form for H:

$$H = C_h(1 + C_r\bar{R}) \left[1 + \mathrm{erf}\left(\frac{\lambda_m - \lambda_b}{\lambda_h} \right) \right]$$

where the C's are constants to be established by iterative testing against scintillation data, and where the error function arises from integration over the normal distribution of instantaneous boundary location, whose distribution has standard deviation λ_h (Fremouw & Bates, 1971).

The outer scale, height, spectral index, and drift velocity are established by simple models in the program. The parameters a, b, and δ describe the three-dimensional configuration of the irregularities. As described previously, these have been established for the auroral zone by utilizing Wideband observations at Poker Flat.

Once these eight parameters have been established, the model will provide T, p, σ_ϕ, and S_4. Since our interest is in the scintillation indices, we will concentrate on them.

The scintillation index for phase is simply its standard deviation σ_ϕ which may be calculated by integrating the phase-scintillation temporal spectrum $\Phi_\phi(f)$ as follows:

$$\sigma_\phi^2 = \int_{f_c}^{\infty} \Phi_\phi(f)df = \int_{f_c}^{\infty} \frac{Tdf}{(f_0^2 + f^2)^{p/2}}$$

where $f_0 = V_e/2\pi\alpha$ (V_e being the effective velocity of the satellite).
[The outer scale α is measured in radians per meter in the field normal reference direction.]

In this equation, f_c is the lowest phase-fluctuation frequency to which the system is sensitive. For instance, in the Wideband satellite experiment with normal processing, f_c was 0.1 Hz (Sagalyn et al, 1974), as set by phase detrending. In a coherently integrating radar, it would be the reciprocal of the time over which phase coherence is required. For systems not sensitive to phase instability in the propagation medium, f_c is effectively infinite, and the effective σ_ϕ is zero.

The scintillation index for intensity is the ratio S_4 of the standard deviation of received signal power to the mean received power (Briggs & Parkin, 1963). Unlike σ_ϕ, its relation to T is set not by a system or an ionospheric parameter, but by the diffraction process that gives rise to intensity scintillation. For weak to moderate levels of intensity scintillation, S_{4w}^2 is very well approximated (Rino, 1979; Fremouw, 1980-1) by:

$$S_{4w}^2 = C(V)\frac{T}{V_e^{(2v-1)}} \frac{F}{G} Z^{v-1/2}$$

where C(v) is a normalization factor. The Fresnel filter factor F(a, b, δ, v) describes the geometrical enhancement of intensity scintillation. It also accounts for diffraction, together with the Fresnel-zone size.

$$Z = \frac{\lambda z \text{ seconds } \Theta}{4\pi} \cdot$$

in which z is the effective "reduced height" (including correction for wavefront curvature and curved-earth geometry) of the irregularities. G describes the static geometrical factor.

While the above equation is a weak-scintillation formula, it may be generalized for practical purposes, to include the well-known saturation of S_4 at unity by writing

$$S_4^2 = 1 - \exp(-S_{4w}^2)$$

which is exact for scintillating signals that obey Rice statistics (Fremouw & Rino, 1976).

2. Scintillation Display

In the past few years the trend toward operational reliance on satellite communications has increased the need for forecasting communications outages caused by propagation anomalies. Studies over the past ten years have shown that ionospheric irregularities can severely disrupt earth-space communications, (Aarons, 1977; Paulson, 1981; Nichols, 1974; Frazier & Caster, 1983).

Hardware techniques, such as increased link margin, coding, and interleaving, can be used to reduce the errors caused by ionospheric scintillation fading. However, until current satellite communications hardware is improved, there are a number of operator-generated procedures which can increase message throughput probability during ionospheric scintillation fading. Techniques, such as sending short messages and message repeating, should be employed selectively since these redundancy techniques result in a lower message throughput rate during unfaded conditions.

To determine which communications links have a high probability of fading, an ionospheric scintillation prediction model and a display system have been developed for AFWAL by Linkabit (Rothmuller, 1983).

The Command Post Modem/Processor (CPM/P) in the AN/ASC-30 satellite communications system uses a plasma display for its operator interface. One display mode includes a world map to provide satellite coverage information to the operator. This map can also be used to display areas of the world having a high probability of ionospheric scintillation fading. Displays that are available include a natural (ambient) scintillation display, nuclear scintillation display, and the probability of scintillation over a point-to-point link.

To display natural scintillation, the operator chooses the scintillation screen and the ambient scintillation mode display. The operator then selects the time of interest and input the solar activity as described by sunspot number and magnetic index. If the operator does not know the solar conditions, he can use the default numbers stored in the display memory. Once time and the solar conditions are entered, the areas that have a high probability of scintillation fading will appear as a dotted region, Figure 70. In general, ambient scintillation fading affects the northern and southern polar caps, the auroral belt, and a small zone beginning at local sunset along the magnetic equator.

Another display available is for nuclear scintillation fading. The frequency band to be displayed (UHF, SHF, or EHF) is chosen first. Then the time and location of the nuclear detonations are entered by the operator, along with the number of minutes between display updates. As time is stepped forward, the area of scintillation caused by nuclear disturbance can be seen to grow, stabilize, and slowly decay, Figures 71 - 75.

The third display mode is for point-to-point calculations which gives scintillation at a point on the earth or along a particular link. In this mode, the operator selects the time of interest and the point or link coordinates. The resultant scintillation index, S_4, and the probability of the scintillation fading occurring, P, are then displayed, Figure 76.

It is possible to measure scintillation fading at various locations just as wind speed, temperature, and humidity are measured by small unmanned weather stations. The current scintillation fading conditions could be sent back through the satellite to a command post terminal, either on a random basis or in response to an operator-generated poll. This information could then be used to update the scintillation display. If automatic scintillation measuring and transmitting stations were located in the polar, equatorial, and auroral regions, it would significantly enhance scintillation prediction capability. Just as general predicted weather data is substantiated by measured meteorological conditions, the prediction models for the scintillation could be improved or validated with measured data.

Systems which provide location and yield of nuclear detonations, such as IONS, could provide inputs to the scintillation model to automatically introduce the nuclear scintillation events as they occur, without operator intervention.

The ionospheric scintillation models used in the CPM/P scintillation display are a simplified heuristic model intended to introduce operators with the concept of ionospheric scintillation fading rather than give highly accurate predictions. If the scintillation display concept proves to be a useful tool,

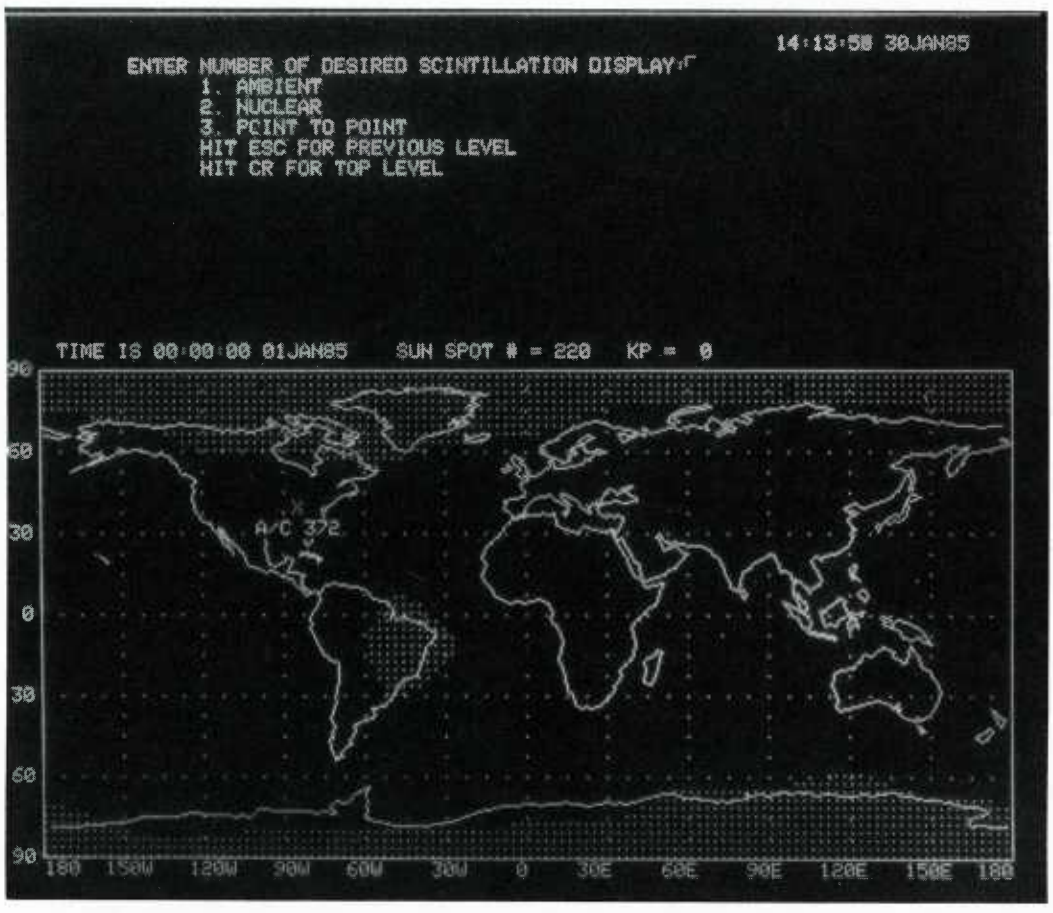


FIGURE 70. AMBIENT SCINTILLATION AREAS

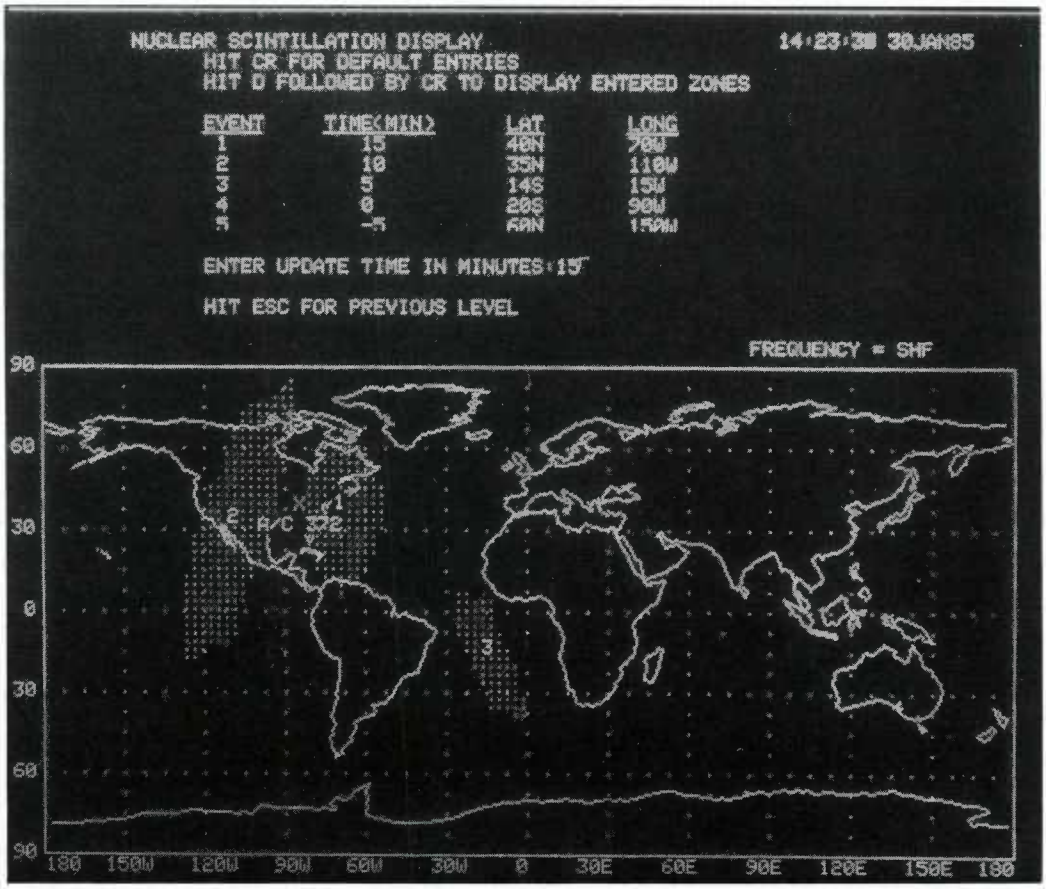


FIGURE 71. SHF NUCLEAR SCINTILLATION DISPLAY AT FIFTEEN MINUTES

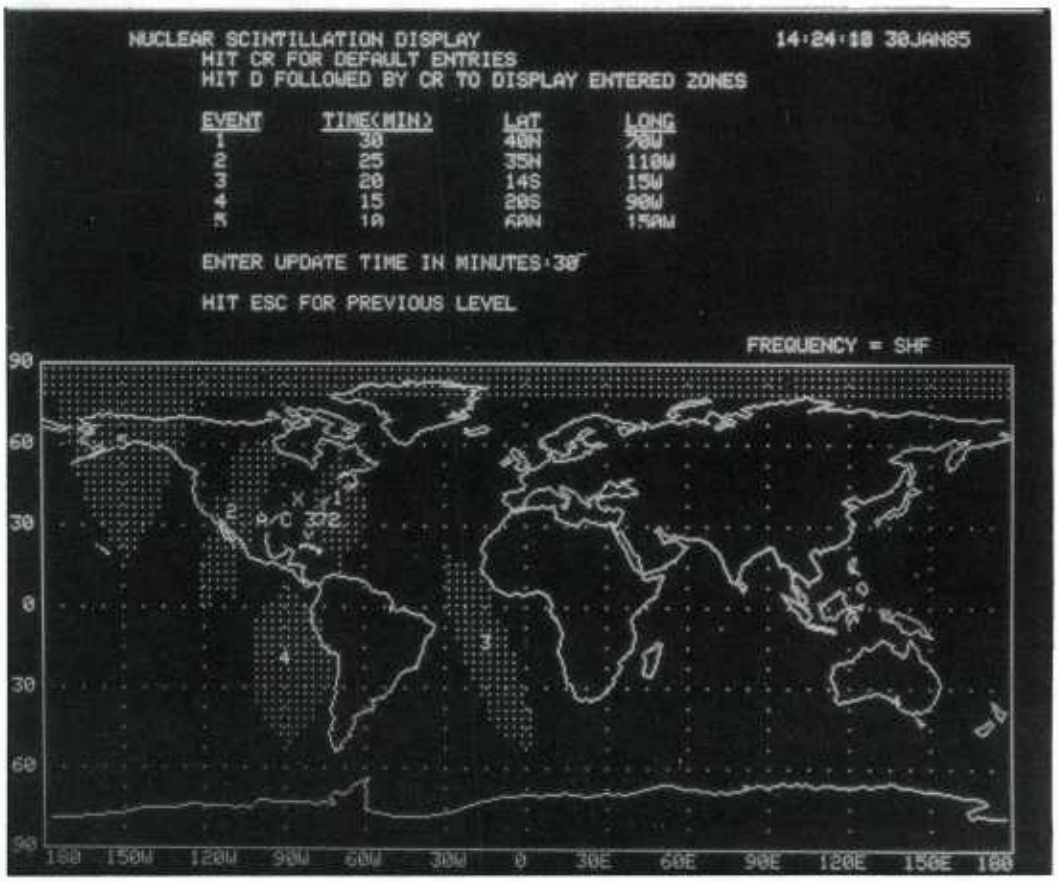


FIGURE 72. SHF NUCLEAR SCINTILLATION DISPLAY AT THIRTY MINUTES

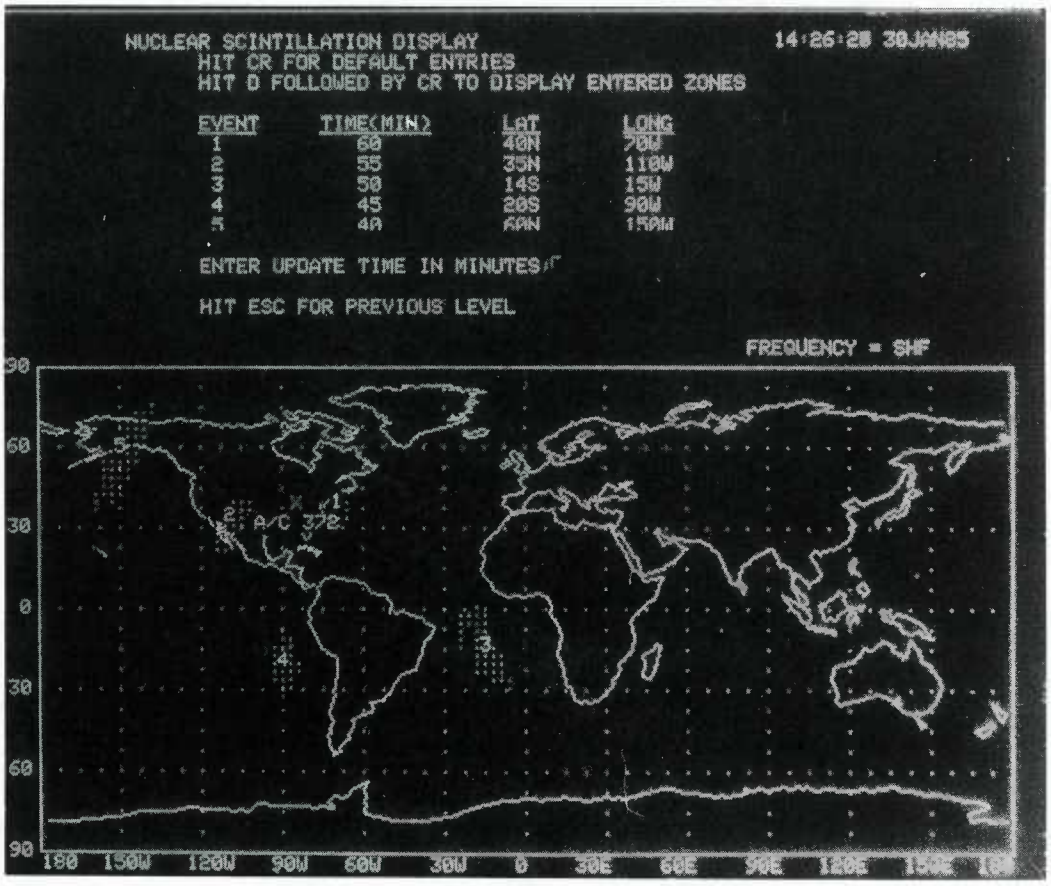


FIGURE 73. SHF NUCLEAR SCINTILLATION DISPLAY AT SIXTY MINUTES

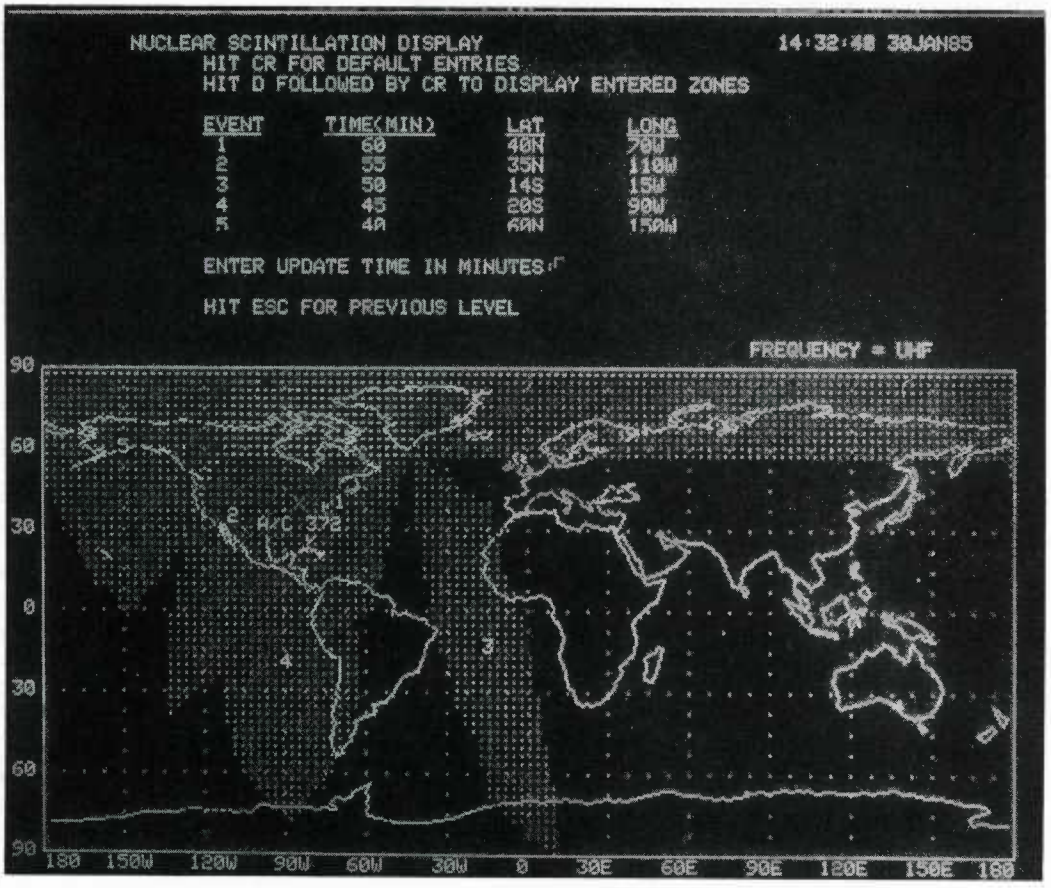


FIGURE 74. UHF NUCLEAR SCINTILLATION DISPLAY AT SIXTY MINUTES

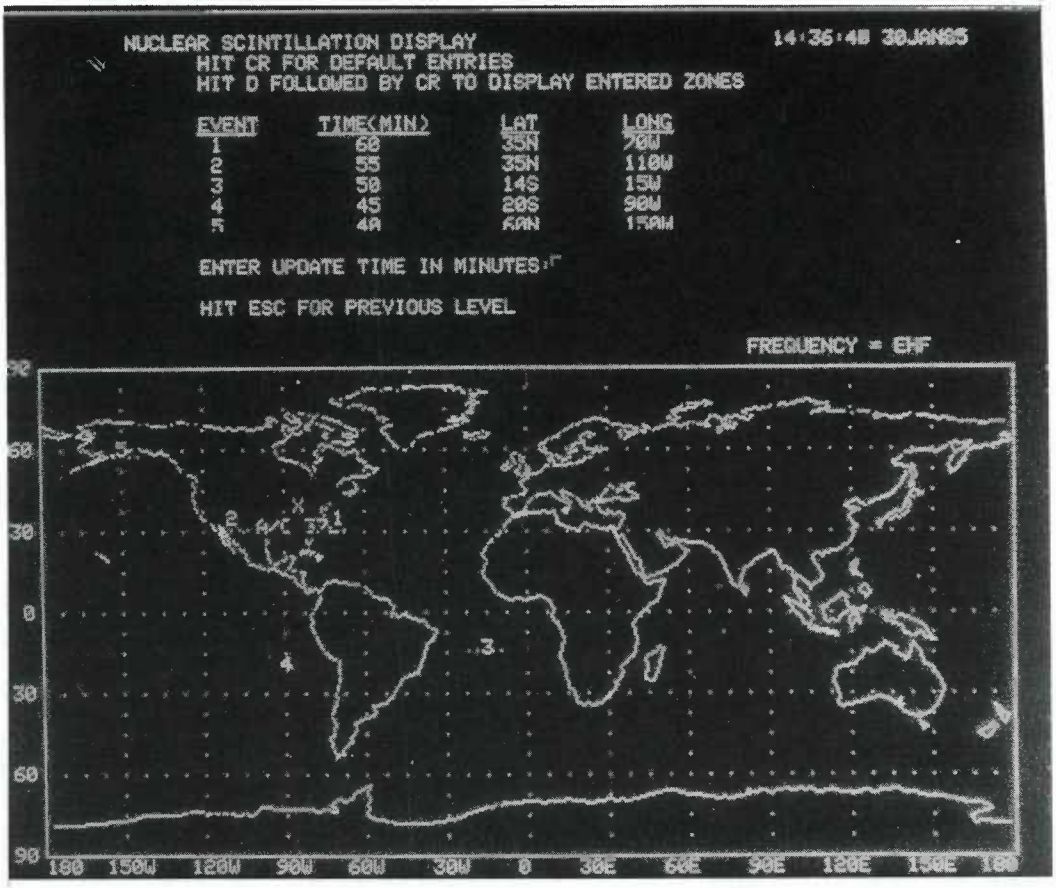


FIGURE 75. EHF NUCLEAR SCINTILLATION DISPLAY AT SIXTY MINUTES

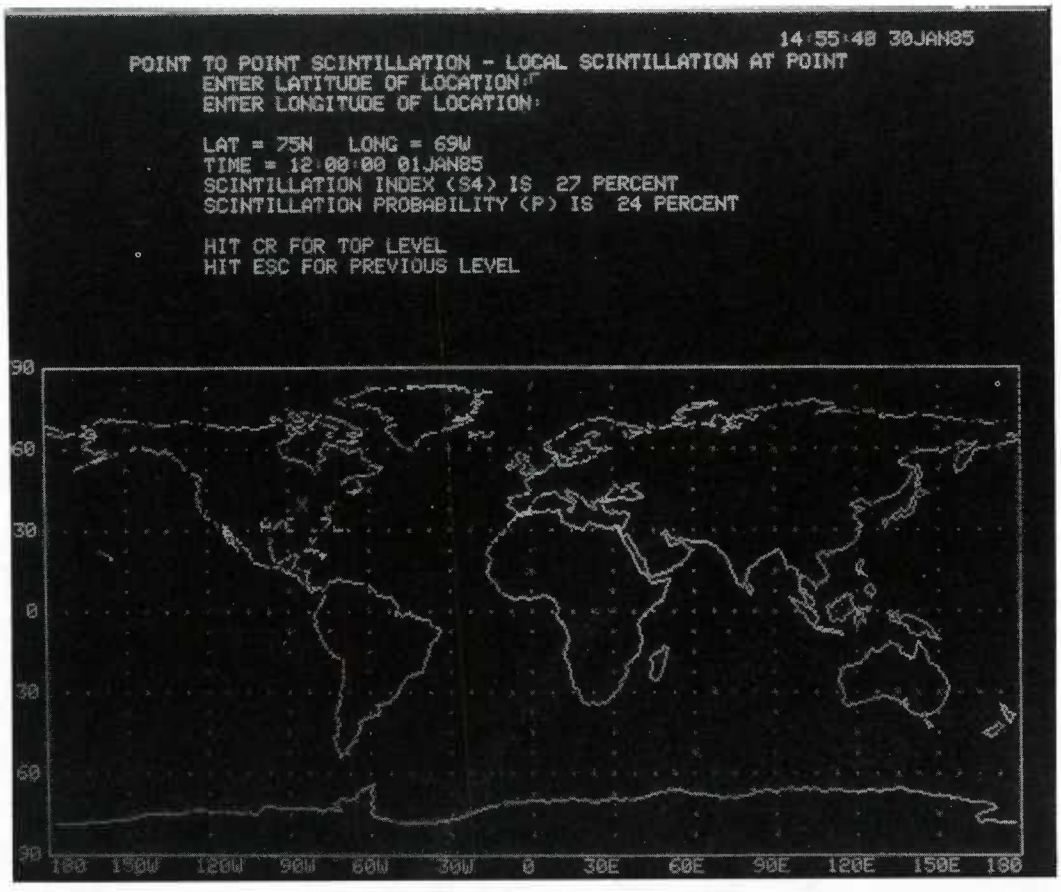


FIGURE 76. POINT-TO-POINT SCINTILLATION RESULTS

the more accurate scintillation modeling which is underway at dozens of facilities (Aarons, 1982; Fremouw, 1980-2; Rastogi, 1980; Rino, 1979; Tsunoda, 1983) could be used to upgrade the scintillation prediction model used in the display.

3. Multipath Modelling

Multipath fading caused by terrain reflections can disrupt an airborne communications or navigation system, (extracted from Johnson, 1979-2). The severity of the multipath is dependent upon the antenna illumination factor, surface reflection coefficient, and the divergence of the reflected signal.

The multipath model presented in this section consists of a direct component and two reflective components, a specular and a diffuse, Figure 41. The severity of the effect of multipath fading is dependent upon the antenna pattern, the reflection coefficient, and the curvature of the reflecting surface. The depth of multipath fading is given by (Franklin, S. B., 1973):

$$\text{Fade depth in dB} = 20 \log (1 - K^2[R|D)$$

where K^2 is the antenna illumination factor (ratio of antenna gain in direction of reflection to gain in direction of direct path); R is the reflection coefficient of the reflecting surface; and D is the divergence factor for the spreading from the curved earth.

The antenna pattern, or antenna illumination factor, has a major influence on the multipath severity. Four types of antennas are illustrated in Figure 66. The first is a blade type antenna which has a horizon coverage pattern. The horizon coverage pattern tends to illuminate the reflecting surface and generally results in a large illumination factor.

The antenna illumination factor is dependent upon the azimuth direction. Any real antenna mounted on an aircraft has nulls and peaks in its pattern, and the gain toward the satellite or toward the reflecting surface varies with the aircraft heading. The average illumination factor, K^2, from a measured blade antenna pattern on an aircraft is given in Table 5. A range of K^2 for the best and worst case aircraft directions is given in Table 6.

An omni-directional type antenna exhibits an antenna illumination factor similar to a horizon antenna, Table 5.

The overhead antenna pattern characteristics of a cross-dipole antenna give considerable protection to the multipath reflective component when the satellite is at a high elevation angle. This discrimination results in a low antenna illumination factor, Table 5.

A fourth antenna type is a directive antenna primarily used at the microwave frequencies. The directive antenna, again, gives good protection against the reflected signal unless the satellite is right on the horizon, Table 5.

TABLE 5: Antenna Illumination Factor (K^2) for Various Antennas*
(For Aircraft at 30,000 Ft.)

ELEVATION ANGLE	AVERAGE HORIZON ANTENNA	AVERAGE OMNI ANTENNA	AVERAGE OVERHEAD ANTENNA	AVERAGE DIRECTIONAL ANTENNA
90°	.18	.10	.026	.003
80°	.14	.10	.026	.003
70°	.14	.10	.032	.003
60°	.14	.10	.032	.003
50°	.14	.14	.032	.003
40°	.18	.18	.04	.01
30°	.22	.32	.05	.02
20°	.50	.56	.08	.03
10°	.71	.71	.20	.10
0°	.94	.94	.89	.16
-3°	1.00	1.00	1.00	1.00

*from Burnside, W. D. et al 1973
Clanton, S., 1974
Maroth, V., 1963

TABLE 6: Antenna Illumination Factor Variances (K^2)*
(For Aircraft at 30,000 ft.)
Typical Horizon Coverage Antenna

ELEVATION ANGLE	MINIMUM K^2	AVERAGE K^2	MAXIMUM K^2
90°	.10	.18	1.00
80°	.10	.14	.79
70°	.03	.14	.56
60°	.03	.14	.31
50°	.03	.14	.56
40°	.03	.18	.56
30°	.03	.22	.79
20°	.03	.50	.79
10°	.10	.71	.89
0°	.89	.94	1.00
-3°	1.00	1.00	1.00

* from Clanton, S., 1974

The reflection coefficient used in the preceding equation is a combination of the basic reflection coefficient of the surface multiplied by the specular, or diffuse, reflection coefficient. The basic reflection coefficient is dependent upon the type of reflecting surface, the polarization of the reflected wave, the elevation angle of the reflection, and frequency. For a horizontally polarized incident wave, the reflection coefficient is given by (Reed and Russell, 1964):

$$R_h = \frac{\sin\psi - \sqrt{n^2 - \cos^2\psi}}{\sin\psi + \sqrt{n^2 - \cos^2\psi}}$$

where n^2 is a complex reflectivity factor = $\varepsilon_r - j\frac{18000\,\sigma}{f_c}$

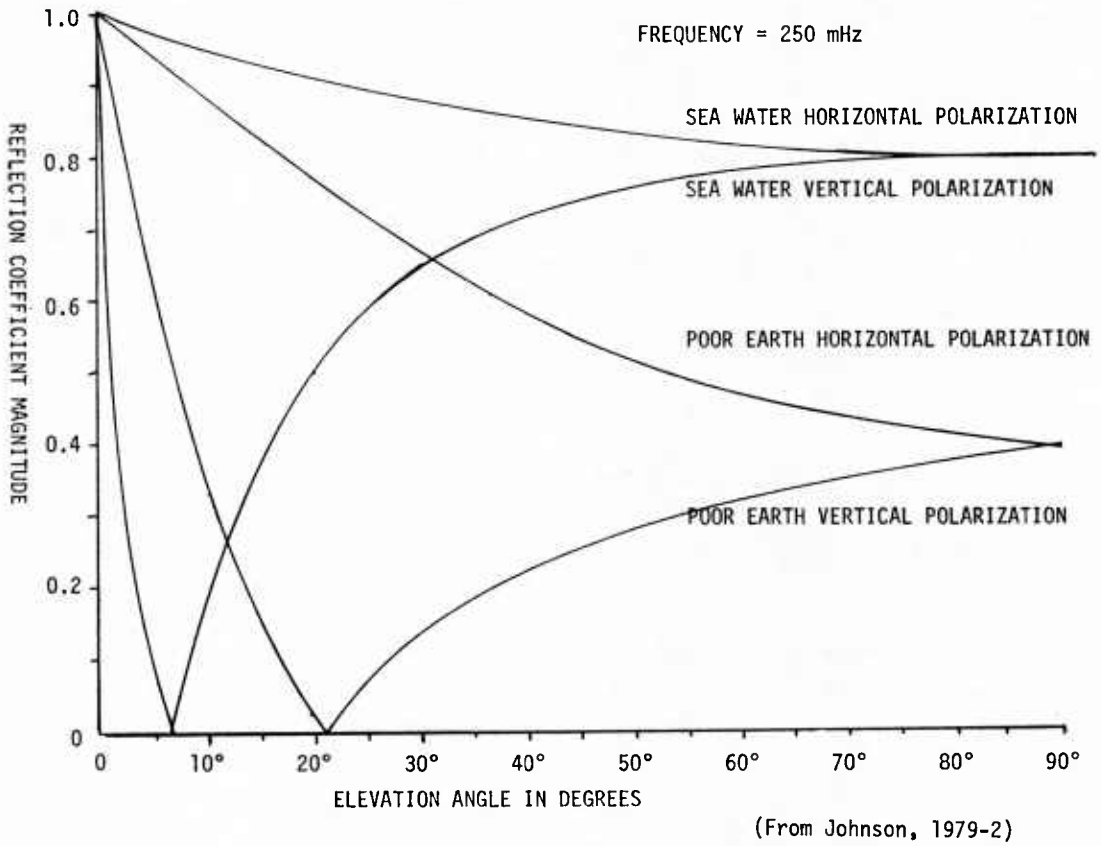
ψ = grazing angle shown in Figure 79
ε_r = relative dielectric constant
σ = conductivity in mhos/meter
f_c = carrier frequency in MHz

Neither the amplitude nor phase of the horizontal reflection coefficient is very sensitive to the physical surface constants nor frequency, Figure 77.
Typical values of the physical constants for various types surfaces are given in Table 7.

TABLE 7: Physical Properties of Various Reflecting Surfaces*

SURFACE TYPE	PERMEABILITY μ	DIELECTRIC ε_r	CONDUCTIVITY σ
Free space	1.257×10^{-6} Henry/meter	8.855×10^{-12} Farad/meter	mhos/meter
Distilled Water		78	1×10^{-5}
Fresh water		80	8×10^{-3}
Sea water		80	5
Ice (fresh water)		3	2×10^{-4}
Dry, sandy, flat coastal land (dry earth)		5	1×10^{-4}
Marshy, forested, flat land (wet earth)		30	1×10^{-4}
Farmland -- low hills		15	1×10^{-2}
Pastoral land, medium hills		13	5×10^{-3}
Rockly land, steep hills		10	2×10^{-3}
Mountainous		5	1×10^{-3}
City, residential area		5	2×10^{-3}
City, industrial area		3	1×10^{-1}

* From Kerr, D.E., 1951
Reed and Russell, 1964



(From Johnson, 1979-2)

FIGURE 77. VARIATION IN REFLECTIVITY WITH ELEVATION ANGLE

When there is water, snow, or ice over another surface such as land, the depth of penetration of the electromagnetic wave determines whether the top surface or a combination of the top and underlying surface determine the physical constants. The depth of penetration is give by (Jordan, E. C., 1955):

$$\delta = \frac{\left(\frac{\mu\varepsilon}{2}\left[\left(1 + \frac{\sigma^2}{\omega^2\,\varepsilon^2}\right)^{\frac{1}{2}} - 1\right]\right)^{-\frac{1}{2}}}{\omega}$$

where: δ = depth of penetration
μ = permeability in Henry/meter
ε = dielectric constant in Farad/meter
σ = conductivity in mhos/meter
$\omega = 2\pi f_c$ in Hertz
f_c = carrier frequency in Hertz

The depths of penetration for several surfaces at 300 MHz are:

Sea Water	0.015 meters
Fresh Water	5.93 meters
Ice	45.93 meters
Average land	2.06 meters

For a vertically polarized incident wave, the reflection coefficient is given by (Reed and Russell, 1964):

$$R_v = \frac{n^2 \sin\psi - \sqrt{n^2 - \cos^2\psi}}{n^2 \sin\psi + \sqrt{n^2 - \cos^2\psi}}$$

The amplitude of the vertical reflection coefficient goes through a minimum at a grazing angle (called the Brewster angle) which is dependent on the physical surface constants and frequency. Two examples are shown in Figure 77. The phase goes through a 180° reversal around the point of minimum amplitude.

For a circularly polarized incident wave, the reflection coefficient is given by (Chinnick, J. J., 1977):

$$R_{cs} = \frac{\sqrt{R_h^2 + R_v^2 + 2 R_h R_v \cos(\phi_h - \phi_v)}}{2}$$

when the direct and reflected waves have the same sense circular polarization and

$$R_{co} = \frac{\sqrt{R_h^2 + R_v^2 - 2 R_h R_v \cos(\phi_h - \phi_v)}}{2}$$

when the direct and reflected waves have opposite sense circular polarization.

The amplitude of the same sense circular reflective coefficient falls between the horizontal and vertical reflection coefficients. The amplitude of the opposite sense circular reflection coefficient starts high at large grazing angles like the horizontal reflection coefficient and goes to zero as the grazing angle decreases to zero.

The surface roughness will determine the ratio of specular to diffuse reflected energy. For a smooth surface the specular component will predominate, while for a rough surface, the diffuse scattered components predominates, Figure 78 (Beckmann and Spizzichino, 1963). For a relatively smooth surface the coherent scattering is limited to the first Fresnel Zone. However, for a very rough surface such as a rough sea, the scattering may come from a considerably larger area.

Surface roughness tends to be elevation angle dependent. Even the very rough sea appears smooth when viewed from a very low elevation angle. The relationship between surface roughness and elevation angle is shown in Table 8. When the grazing angle exceeds about twice the critical angle shown in Table 8, the specular components becomes insignificant and the reflection coefficient depends upon the diffuse scattering factor.

TABLE 8: Maximum Angles for Specular Reflection at Different Frequencies and Sea States*

Sea State	Description of Sea	Wave Height	Rms Height, λ_h (m)	Critical Angle, max-deg λ=0.7m	λ=0.23m	λ=0.1m	λ=0.03m
1	Smooth	0-0.3	0-0.065	45	13	6	1.8
2	Slight	0.3-1	0.065-0.21	12-45	4-13	1.8-6	0.5-1.8
3	Moderate	1-1.5	0.21-0.32	8-12	2.6-4	1.2-1.8	0.3-0.5
4	Rough	1.5-2.5	0.32-0.54	5-8	1.6-2.6	0.7-1.2	0.2-0.3
5	Very Rough	2.5-4	0.54-0.86	3-5	1-1.6	0.4-0.7	0.12-0.2
6	High	4-6	0.86-1.3	2-3	0.7-1	0.3-0.4	0.04-0.12
7	Very High	6	1.3	2	0.7	0.3	0.04

*from Beckmann and Spizzichino, 1963

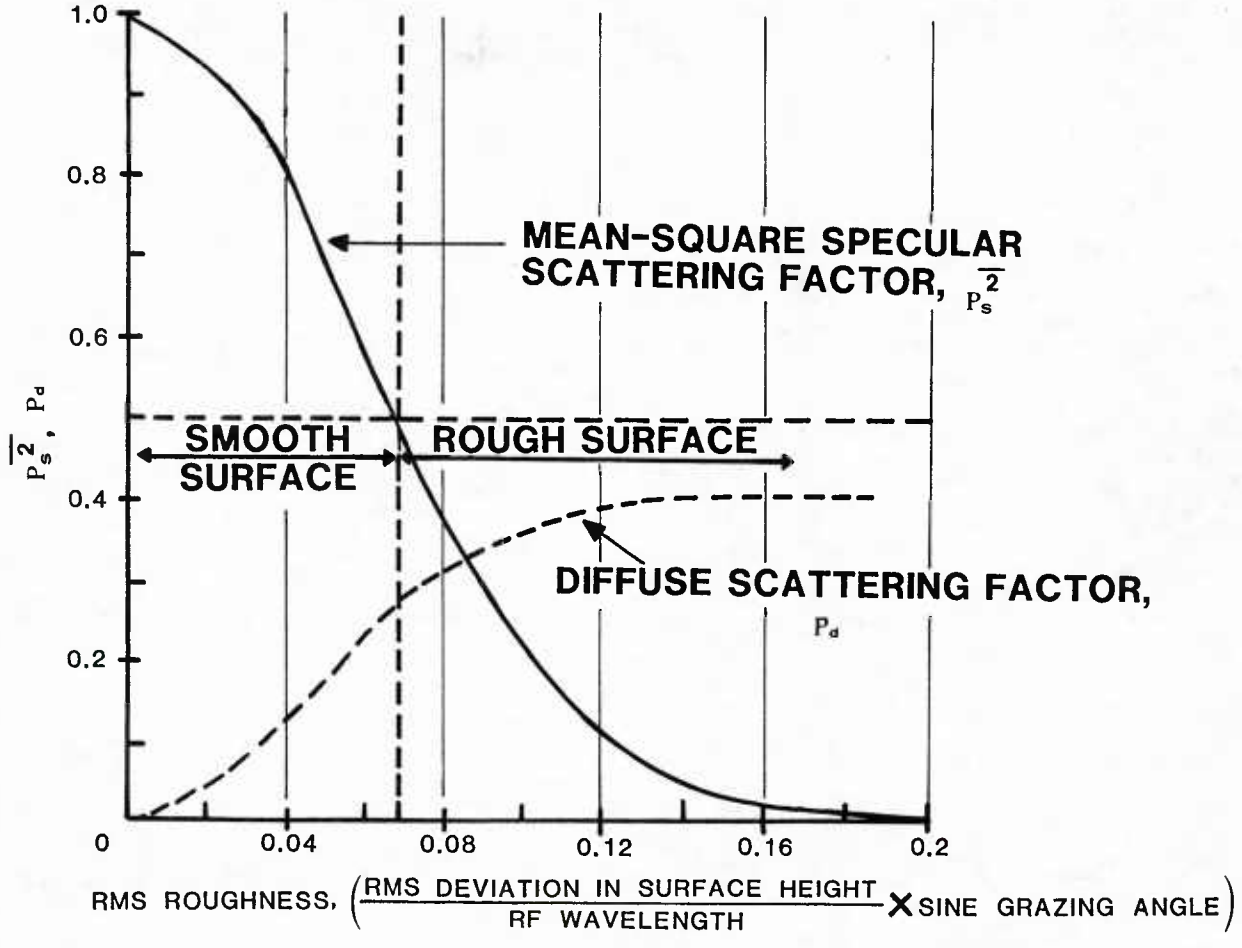


FIGURE 78. SCATTERING FACTOR VERSUS SURFACE ROUGHNESS

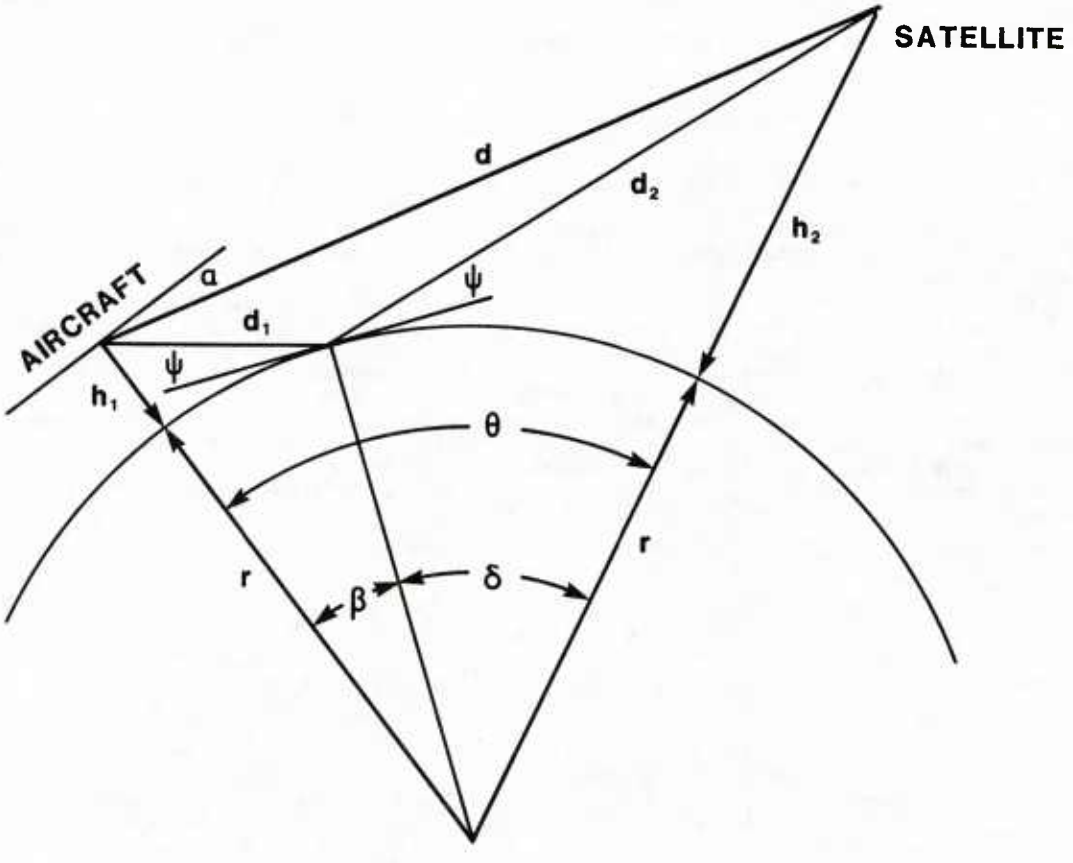


FIGURE 79. MULTIPATH GEOMETRY

The divergence factor takes into consideration the convex shape of the reflecting surface. The divergence is given by (Beckmann and Spizzichino, 1963):

$$D = \left(1 + \frac{2\,d_1 d_2}{r\,(d_1 + d_2)\,\mathrm{Sin}\,\psi}\right)^{-\frac{1}{2}} \left(1 + \frac{2 d_1 d_2}{r\,(d_1 + d_2)}\right)^{-\frac{1}{2}}$$

Refer to Figure 79 for definition of terms.

The divergence is very sensitive to the height of the receiving antenna. The divergence factors for an antenna at 100 feet, 30,000 feet, and 70,000 feet is shown in Figure 80 (Foley, et al, 1968).

The multipath fade rate can be determined geometrically by calculating the rate of change of the path difference between the direct and the reflected component, (Bond, F.E., 1967).

$$\text{Fade rate} = \frac{d}{dt}\,.\,\frac{d(\Delta d)}{d\theta}\,.\,\frac{f_c}{c}$$

where $\theta = \beta + \delta$ = great circle angle between aircraft and satellite (See Figure 79)
$\Delta d = d_1 + d_2 - d$
f_c = carrier frequency
c = speed of light
$d = [(r + h_1)^2 + (r + h_2)^2 - 2\,(r + h_1)\,(r + h_2)\,\mathrm{Cos}\,\theta]^{\frac{1}{2}}$
$d_1 = [r^2 + (r + h_1)^2 - 2r\,(r + h_1)\,\mathrm{Cos}\,\beta]^{\frac{1}{2}}$
$d_2 = [r^2 + (r + h_2)^2 - 2r\,(r + h_2)\,\mathrm{Cos}\,\delta]^{\frac{1}{2}}$.

$$\mathrm{Cos}\,\beta = \frac{r}{r + h_1}\left(w + \left[w^2 + \left(\frac{r + h_1}{r}\right)^2 - w\left(\frac{r + h_1}{r}\right)^2 - w\right]^{\frac{1}{2}}\right)$$

where
$$w = \frac{(r + h_2)^2\,\mathrm{Sin}^2\,\delta}{(r + h_2)^2 + r^2 - 2r\,(r + h_2)\,\mathrm{Cos}\,\delta}.$$

These equations are in good agreement with the fading data presented in Section I.C. of this report.

F. Mitigation Techniques

The communications system designer should consider the propagation characteristics of the communications channel and design the system to achieve acceptable communications performance. Based on the propagation characteristics described in this report, a number of techniques can be employed to overcome anomalous propagation affects.

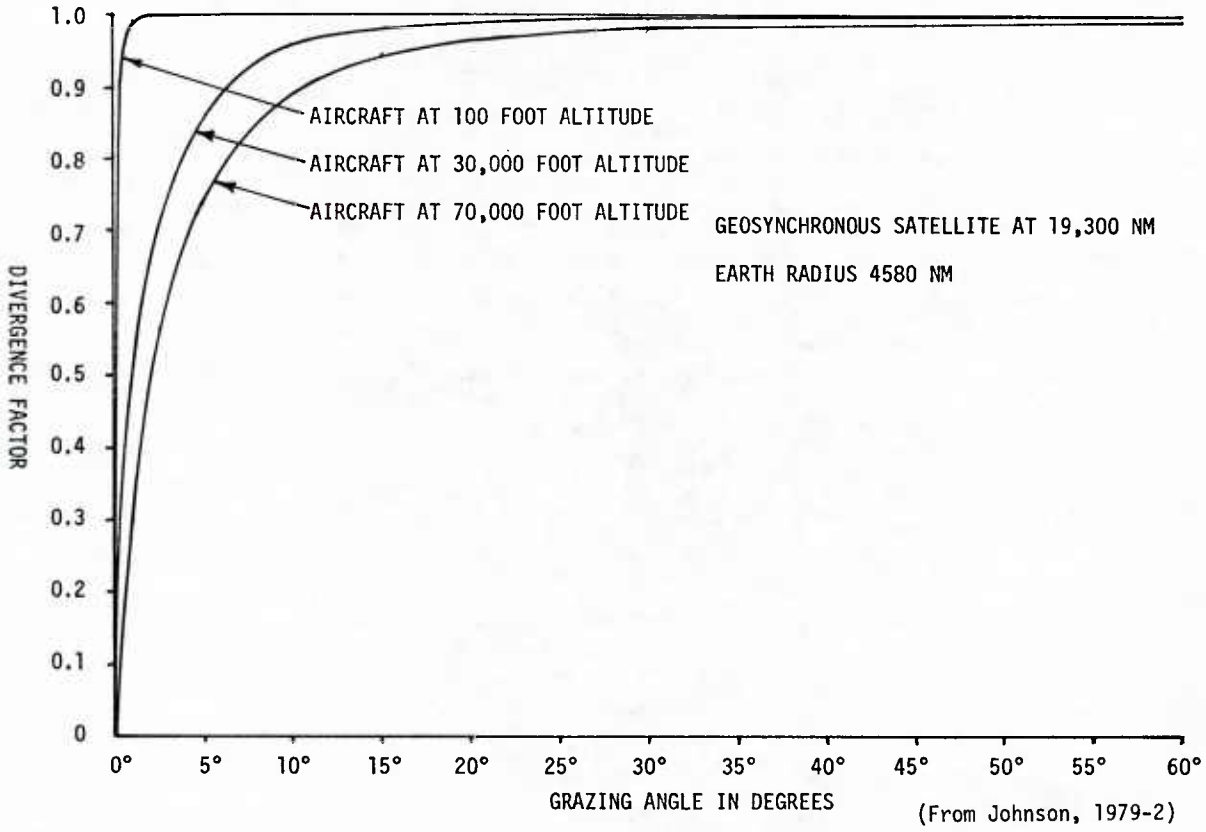
1. Scintillation Mitigation Techniques

Ionospheric scintillation fading tends to cover a broad frequency band. For this reason frequency diversity appears to offer little advantage for mitigating ionospheric scintillation fading (extracted from Johnson, 1980). At 300 MHz frequency separations of 100 MHz would be required to provide any significant diversity improvement. Measurements of space diversity made by Paulson and Hopkins (1977) showed space diversity requires antenna separations of 400 to 1000 meters to obtain good decorrelation, Figure 81. Obviously, airborne platforms cannot provide this much antenna separation. Polarization diversity offers no improvement for UHF frequencies against ionospheric scintillation.

One space diversity technique which would provide significant improvement at the cost of system complication is the use of two satellites. Spaced satellites would allow the aircraft-to-satellite paths to be spatially separated to decorrelate the fading and offer a diversity improvement. However, such a technique would complicate the system operation.

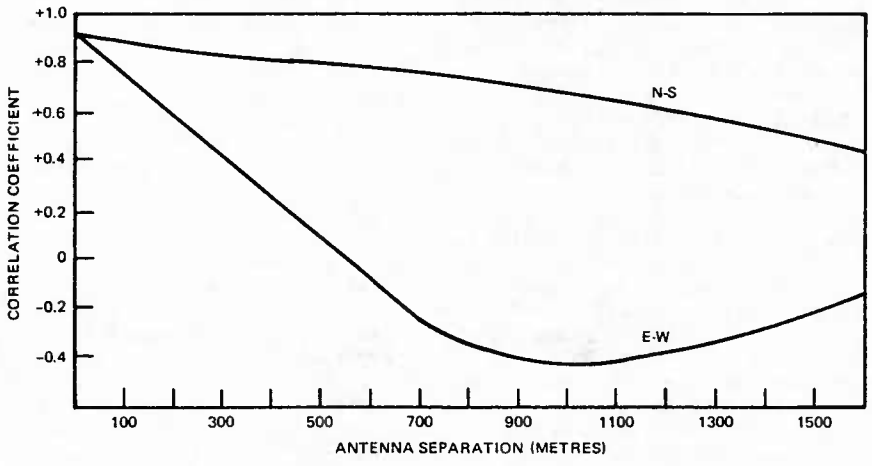
The technique which appears to offer significant improvement is the use of time diversity (Foshee, 1977). Since ionospheric scintillation fading has no significant attenuation factor associated with it, it should be possible to make use of the periods of normal or enhanced signal amplitude to carry through the periods of degraded signal amplitude. Several systems involving forward error correction coding and interleaving have been built and tested (White, 1977). These techniques use one-half rate or three-quarter rate forward error correction coding along with bit interleaving to compensate for the fade periods. Since the fading is relatively long, large bursts of errors are periodically generated. The use of bit interleaving randomizes these burst errors, providing a more uniform distribution of bit errors. The forward error correction decoder then operate on the randomly distributed errors providing an error free decoded message. Coded communications systems have been tested under both the equatorial and polar scintillation conditions and have shown that error free message copy can be accomplished even during severe scintillation periods.

Another technique which offers a pseudo space diversity approach is the use of the bottom mounted aircraft antenna in conjunction with the upper top mounted antenna. Experimental tests have shown that at relatively high elevation angles, sea multipath degrades the signal amplitude only 3 to 6 dB (Johnson, 1980). If the aircraft-to-satellite path is 40° elevation or greater, the direct path and sea reflected path propagate through different parts of the ionosphere, resulting in a space diversity. In addition, the sea multipath reflection comes from a large area encompassing the first Fresnel zone and inherently displays space diversity characteristics. Measurements made in the equatorial region during severe ionospheric scintillation showed the multipath reflected signal to display considerably less amplitude variations than the direct signal received by a top mounted antenna. Therefore, the combination of the sea multipath and direct signal using a diversity combiner or even the use of the multipath signal by itself offers a significant degree of scintillation protection. Measurements made in the equatorial region indicated that the effect of severe ionospheric equatorial scintillation can be reduced significantly using this technique.



(From Johnson, 1979-2)

FIGURE 80. DIVERGENCE FACTOR VERSUS GRAZING ANGLE



(From Paulson & Hopkins, 1977)

FIGURE 81. EQUATORIAL SCINTILLATION CORRELATION COEFFICIENT VERSUS ANTENNA SEPARATION

a) FLEETSAT Coding Example

Under the sponsorship of the Navy, MIT Lincoln Laboratory conducted a coding test to evaluate the effect of UHF ionospheric scintillation on the Navy's fleet broadcast system on the FLEETSAT Satellite (extracted from White, 1977). An interleaving scheme for the systematic and non-systematic codes was developed. Implicit in this concept was the design of an interleaver (as well as delay lines and deinterleaver) structure that had relatively simple synchronization requirements. Note that all the interleavers (and deinterleavers) associated both with the transmitter and the receiver had to operate in synchronism and hence a periodic synchronization signal had to be transmitted for this purpose. With fades of the order of 1 sec the interleaver memory requirements become large and standard RAM memories were logical choices. The multiplicity of interleavers required a design that exhibits a high degree of commonality in the RAM addressing scheme.

The design included a simple means of implementing the Viterbi decoder at coder rates 1/2, 2/3, and 3/4. Fortunately the data input rate was limited to 225 bps (for the highest code rate) and this allowed use of a simple computer type structure, the advantage being a relatively small number of medium scale integration (MSI) chips (i.e., 80 chips).

Some analog circuit additions were made to the Navy's fleet broadcast SSR-1 receiver. The normal SSR-1 configuration has a Costas phase locked loop for tracking the carrier with the data stream resulting from the I (in phase) channel output being bit synchronized and matched-filtered. They then added a parallel and identical circuit for the Q (quadrature-phase) channel. Moreover 3 bit soft decisions were being taken from both channels rather than the normal hard quantization on the I channel matched filter output alone. Thus a differentially coherent (in addition to the normal differentially encoded coherent) demodulation scheme and a soft decision capability was added to the SSR-1. Circuitry was also added to measure the received power to noise density (P_r/N_o) ratios at the last linear point (as well as P_r alone). The AGC voltage was sampled and used to place a confidence level on the data; this being done in a DPSK digital combiner network prior to the decoder input.

There were a number of important aspects of the experimental program. First and foremost it was intended as a demonstration that time diversity via interleaving and convolutional encoding - Viterbi decoding can provide significant error protection in a fading environment such as that caused by equatorial UHF scintillation. Second, it was important to demonstrate that such a scheme could easily be retrofitted wherever needed to the existing fleet broadcast network. There are, of course, different options that one might wish to evaluate and the experiment was designed to test some of the more relevant possibilities. For example using the systematic codes chosen, one could protect 1, 2 or 3 information streams at the expense of one overhead parity stream (with of course decreasing error protection). This system would require minimal change in the fleet broadcast operations. Non-systematic codes were also examined, mostly to demonstrate the added error protection that such codes give and thus provide some substance for future satellite communications planning where error correcting may be required.

The performance of the AN/SSR-1 receiver was evaluated and modifications that may improve its use in a severe fading environment were suggested. For example, the joint use of AGC information and multibit quantization of the matched filter output are obvious means of improving estimates.

In order to verify satisfactory operation of all the system interfaces, MIT Lincoln Laboratory conducted two sets of tests with the transmitter encoder and interleaver system located at the Norfolk, Virginia, Navy Communication Area Master Station (CAMS) and the receiver located at Lincoln Laboratory's Lexington, MA Facility. The October 1975 tests utilized the LES-6 satellite and the March 1976 tests used the Atlantic MARISAT. In both cases there was no discernible fading (as expected) and hence the tests were solely for the purpose of identification of interface and operational problems. It should be noted that aside from the problem of nonalignment of data channels (due to the transmitter's TDM 1150 behavior) the complete system worked satisfactorily.

In June of 1976 the transmitter encoder-interleaver was installed at CAMS EASTPAC Honolulu and the receiver at CAMS WESTPAC Guam. These tests use the Pacific MARISAT satellite. The test systems were operational from August 30 through January 1, 1977. Daily operations were conducted throughout the fall and early winter during the hours 2300 to 0500 GMT which corresponds to 1900-0100 Guam standard time. This is the period in which maximum scintillation had previously been observed. It was expected that scintillation would occur only on the downlink to Guam, since Hawaii is not in the equatorial scintillation zone. The transmitter encoder was also moved to Guam and tests run where both the uplink and downlink are subject to scintillation. Strong or weak scintillation was observed on 30 out of 123 days of observation. The tests proved successful in that typically during scintillation periods a raw channel symbol error rate of 3×10^{-3} (for $E_b/N_o \sim 21$ dB) would, after decoding, result in an information bit error rate of essentially zero.

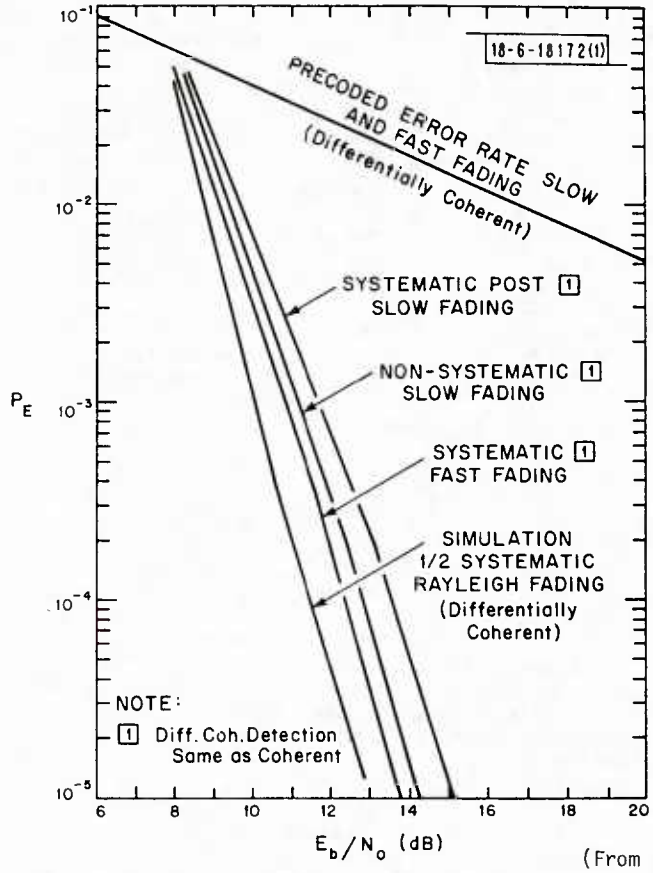
The satellite used in the comprehensive field tests was the Pacific MARISAT Satellite (Gapfiller). The test broadcast was sent over one of the 25 KHz (narrowband A) channels. The average energy per information bit to noise ratio were measured as:

$$E_b/N_o = \frac{P_r}{N_o R} = 21.4 \text{ dB}$$

As an example of the improvement provided by the coding scheme, consider the case of a R = 1/2 systematic code with coherent processing. The energy per information bit to noise ratio could then be 24.4 dB. From Figure 82 it can be estimated that the system margin for a post error rate of 10^{-5} is about 10.5 dB (using the experimental fast fading curve as being representative of the system response to the actual scintillation). At a energy per chip to noise ratio of 21.4 dB, the predecoded error rate is approximately 4×10^{-3}, which is generally considered unacceptable. Without coding, a SNR of about 40 dB would be required for 10^{-4} bit error rate. Thus in the presence of fading the coding system requires significantly less SNR to provide a reliable link.

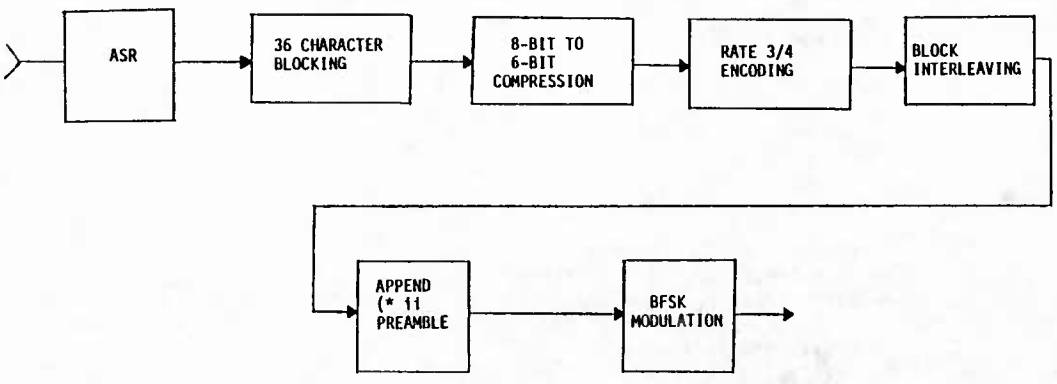
b) AFSAT Coding Example

The overriding requirement guiding the selection of encoding/interleaving schemes for this example was compatibility with existing AFSATCOM I modes of operation of existing hardware (extracted from Bernal, 1979). On one hand, this prescribed certain features as modulation system and channel symbol rate, and constrained such parameters as message length. On the other hand, it provided strong motivation to make maximum use of existing facilities within the modem, with minimum complexity in any alterations which might be needed.



(From White, 1977)

FIGURE 82. ERROR RATE FOR R=1/2 DPSK ENCODED RAYLEIGH FADING, K=7



(From Bernal, 1979)

FIGURE 83. CODED MODEM TRANSMIT PROCESS, RATE 3/4 CODE

The approach taken was to consider initially two baseline convolutionally-coded/interleaved techniques, one which used an LF89 decoder, and the other an LF1Ø11 decoder. These baseline systems use linear feedback decoding and incorporate symbol interleaving over the prevailing coded message length.

This mitigation technique is described here in terms of its actual implementation in existing hardware for purposes of demonstrating performance improvement in the scintillation environment. It employs a rate 3/4 convolutional code, decoded by an LF89 decoder, together with a block de-interleaver designed to separate bursts of channel symbol errors before decoding as much as possible for the block length of 36 characters. This LF89 code is able to correct all single and double error patterns within a span of 36 channel symbols. In addition to channel coding, the implemented scheme performs data compression by converting the 8-bit characters from the ASR (Automatic Send/Receive device, i.e., the teletype) into 6-bit characters before channel coding. The net result is that the channel throughput is unchanged from the uncoded system i.e., a nominal rate of 75/8 = 9.375 characters per second.

A block diagram of the transmit process is shown in Figure 83. The modem accepts ASR inputs in the usual way, in response to the transmit enable signal from the ASR. The 8-bit ASCII characters are grouped into 36 character blocks for transmission. If less than 36 characters are supplied by the ASR, the block is filled with SP (space) characters.

The 8-bit ASCII characters are then compressed to 6-bit ASCII. This was done by deleting bit b8, the parity bit, and b6. The compressed 36 character buffer is then convolutionally encoded using a rate 3/4 code with tail biting; that is, the 216 information bits are encoded as a ring to avoid use of tails containing superfluous symbols unrelated to information. The resulting coded message consists of 288 channel symbols. These symbols are then effectively block interleaved as they are transmitted. In this process we send first symbol 36, then 72, the 1Ø8, etc. to 288, followed by 35, 71, 1Ø7,...287, etc., until the entire coded message block is sent, Table 9.

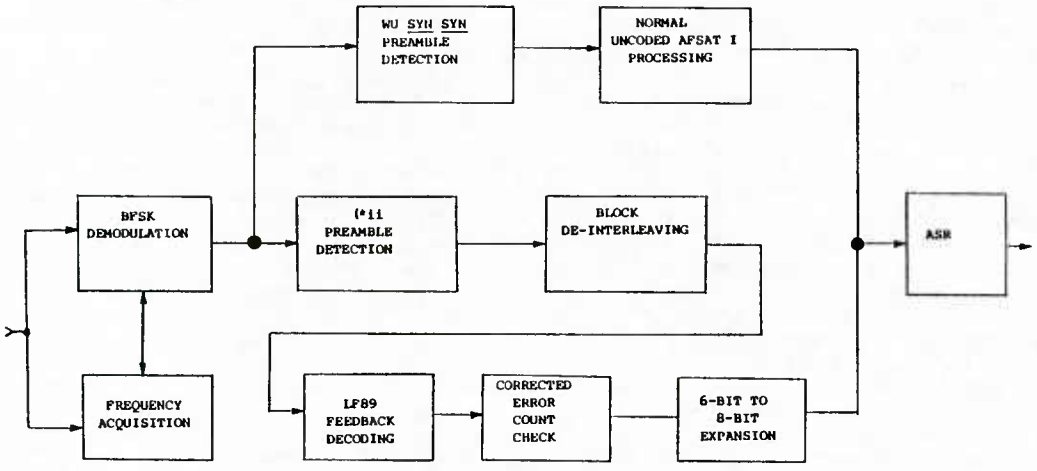
TABLE 9 Interleaver Sequence

	↓ Insert						
1	37	73	109	145	181	217	253
2	38	74	110	146	182	218	254
3	39	75	111	147	183	219	255
4	40	76	112	148	184	220	256
5	41	77	113	149	185	221	257
6	42	78	114	150	186	222	258
7	43	79	115	151	187	223	259
8	44	80	116	152	188	224	260
9	45	81	117	153	189	225	261
10	46	82	118	154	190	226	262
11	47	83	119	155	191	227	263
12	48	84	120	156	192	228	264
13	49	85	121	157	193	229	265
14	50	86	122	158	194	230	266
15	51	87	123	159	195	231	267
16	52	88	124	160	196	232	268
17	53	89	125	161	197	233	269
18	54	90	126	162	198	234	270
19	55	91	127	163	199	235	271
20	56	92	128	164	200	236	272
21	57	93	129	165	201	237	273
22	58	94	130	166	202	238	274
23	59	95	131	167	203	239	275
24	60	96	132	168	204	240	276
25	61	97	133	169	205	241	277
26	62	98	134	170	206	242	278
27	63	99	135	171	207	243	279
28	64	100	136	172	208	244	280
29	65	101	137	173	209	245	281
30	66	102	138	174	210	246	282
31	67	103	139	175	211	247	283
32	68	104	140	176	212	248	284
33	69	105	141	177	213	249	285
34	70	106	142	178	214	250	286
35	71	107	143	179	215	251	287
Transmit ← 36	72	108	144	180	216	252	288

A preamble consisting of the 8-bit ASCII characters (*ii is transmitted immediately prior to the transmission of the 288 symbol message buffer just described. Note that this special coded mode preamble consists of the standard AFSAT preamble, WU SYN SYN with the mark and space interchanged, and that it is neither coded nor interleaved.

The total of 32Ø symbols are transmitted using the standard AFSAT BFSK modulation with 2.5 KHz deviation and 75 baud transmission rate. Transmit frequencies are chosen as in standard AFSAT I modulation. For the purpose of simplifying the implementation, the modem waits until it has finished sending the current block before inputting the next block from the ASR. Thus, as implemented, the average throughput is reduced to approximately 4.69 characters/second when sending a long message; in general, however, the throughput rate of this technique could be maintained at twice this rate, unchanged from the unencoded case.

A block diagram of the receive process is shown in Figure 84. The received signal is processed through the analog circuitry and the BFSK demodulator portion of the firmware in the usual way. The demodulator outputs are then processed parallel by two preamble detectors, one detector the the uncoded mode WU SYN SYN preamble, and one for the (*ii coded mode preamble. The WU SYN SYN detector uses the normal error threshold of two symbol errors; if this preamble is detected, the subsequent message is received and printed in the normal way.



(From Bernal, 1979)

FIGURE 84. CODED MODEM RECEIVE PROCESS, LF89 DECODER

In the implemented version, an error tolerance of three is allowed for the coded mode preamble to improve detection probability in the scintillation environment. The error tolerance in the coded mode can be changed by altering the front panel ROM. If the coded preamble is detected, the modem then attempts to de-interleave and decode the following 288 channel symbols.

The de-interleaving process separates adjacent received symbols by 36 symbols. This permits a single burst error event to span up to 2(288/36) = 16 successive symbols before a decoding error is possible, given the error correcting capability of the code. Thus, fades up to 16/75 (= 0.213) second in duration can be accommodated, assuming that only one such fade occurs per message.

To guard against false alarms due to the relaxed requirements on preamble detection, an error count check is provided that disables output of a message if more than a pre-specified number of errors are corrected in a block. The tolerance on this check is also alterable through change of front panel ROM. To allow data collection over the widest possible range of conditions during demonstration test, this tolerance was set to permit the output of every decoded message.

A successfully decoded block is then expanded from 6-bit ASCII to 8-bit ASCII. This is accomplished by setting b6 equal to the complement of b7, and then setting b8 to yield odd parity over b1 through b8. The resulting message is then delivered to the ASR.

As an alternative to the implemented system, another linear feedback decoded mitigation technique will be described. The chief difference here is the use of a rate ½ convolutional code, decoded with an LF1Ø11 decoder. The primary motivation for investigating the use of this more powerful code, which can correct all single, double, and triple error patterns within a span of 22 channel symbols, is to compare its performance with that of the LF89 system.

The proposed transmit process is shown in block diagram form in Figure 85. The modem accepts ASR characters in response to the ASR transmit enable signal. In this case, up to 39 8-bit characters may be included in a message, and if less than this amount are supplied by the ASR, the block is filled with SP characters.

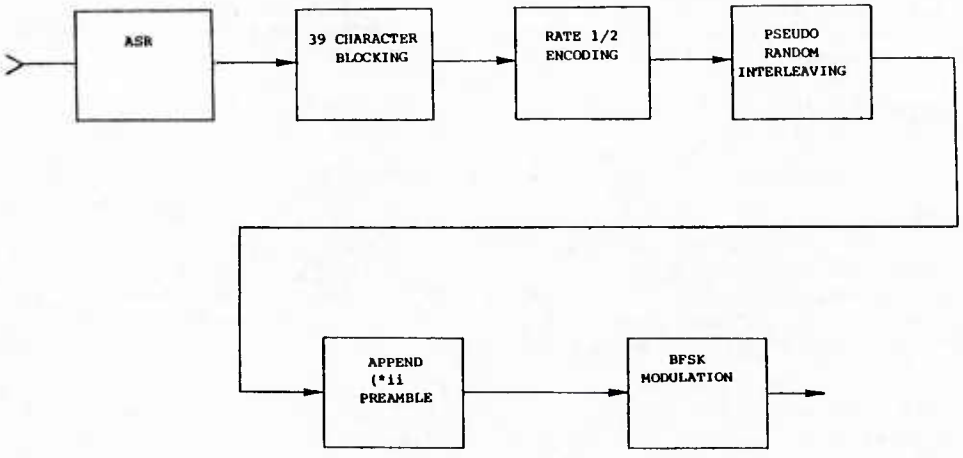
The 39-bit character buffer is then convolutionally encoded using a rate ½ code again with tail biting. The coded message therefore consists of 624 channel symbols. These channel symbols are interleaved as they are transmitted in accordance with a suitable pseudo-random sequence. A preamble consisting of the 8-bit ASCII characters (*ii is transmitted immediately prior to the transmission of the 624 symbol message buffer just described. This preamble, which is neither coded nor interleaved, once again consists of the standard AFSAT I preamble WU SYN SYN with mark and space interchanged.

The total of 656 symbols are transmitted using standard AFSAT BFSK modulation with 2.5 KHz deviation and 75 baud transmission rate. Note that bit rate in this case is approximately ½ that of the 3/4 rate coded technique, and decoding delay is twice as long. Alternatively, a message coded in this way requires twice the transmission time of an uncoded AFSAT I message.

The block diagram of the receive process is shown in Figure 86, and is quite similar to that of the 3/4 rate decoding technique. The differences are the use of pseudo-random de-interleaving and LF1Ø11 decoding, and no requirement for expansion of character size. Demodulation occurs as usual, and the outputs are processed in parallel by the two preamble detectors as previously described. Preamble error and decoding correction tolerances may be set and varied as in the case of the implemented system.

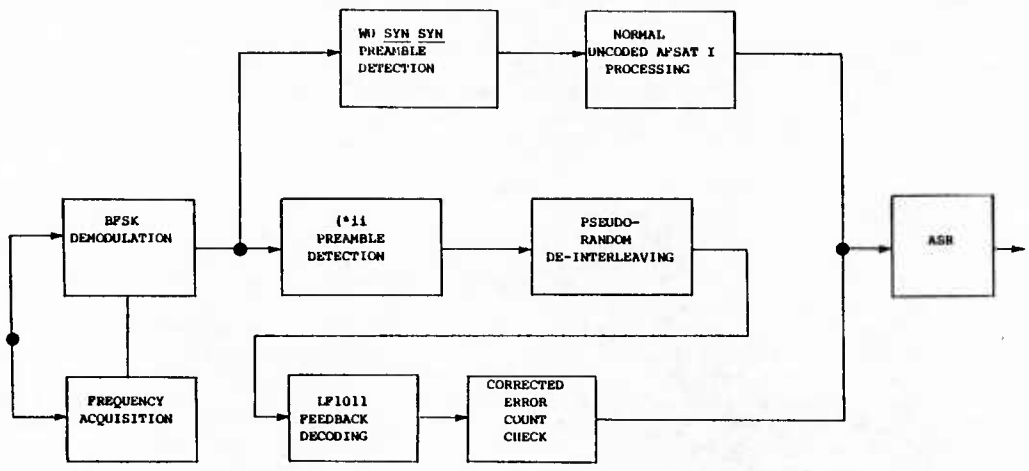
To evaluate the field performance of the coded system, the fade resistant concept was implemented in the dual UHF modem (AN/MD-1024) with a modification of one card in the modem. A 3/4 rate convolutional code with a constraint length of 8 was implemented (Bernal, 1979). A linear feedback decoder with look-ahead length 9 (an LF89 decoder) was employed for error correction. A 288 symbol block interleaver was used to interleave each 36-character message. That message is preceded in transmission by a 4-character uncoded preamble.

The fade resistant dual UHF modem checks each 40-character receive message to see if it has the normal AFSAT I 4-character preamble (WU SYN SYN) or the fade resistant preamble (inverted WU SYN SYN). If the modem correctly recognizes 29 of the 32 inverted preamble bits, or 30 of the 32 uninverted bits, it proceeds to process respectively a coded or uncoded message. If the bit error rate in the received preamble is higher than this, the preamble will be missed.



(From Bernal, 1979)

FIGURE 85. CODED MODEM TRANSMIT PROCESS, RATE 1/2 CODE



(From Bernal, 1979)

FIGURE 86. CODED MODEM RECEIVE PROCESS, LF1011 DECODER

The fade resistant dual UHF modem has been implemented in a manner which will operate well through fast fades, but not through long, slow fades. The fade rate can be measured in fades per minute or in the time it takes the autocorrelation function to decrease to 0.37 of its original value (τ_0).

The modified dual UHF modem was installed in a USAF SATCOM testbed aircraft and a series of test flights flown in the arctic region in 1979 to measure the modem performance during auroral ionospheric scintillation conditions.

The test sequence involved a UHF transmission from a facility in San Diego, California through a polar satellite, down to the test aircraft in the arctic. A four-second test message was sent in a sequence of fifteen minutes of coded transmission followed by fifteen minutes of uncoded transmission.

The test messages received on the aircraft were printed on a teletype and the message error rate determined. A four-second message was considered to be in error if one or more bits were in error. In order to vary the received E_b/N_o, a calibrated variable attenuator was added to the receiver between the UHF antenna and the UHF preamplifier. The attenuator allowed the received signal level to be varied. The receiver noise level was determined by the UHF preamplifier and did not vary as attenuation was added. The received Message Error Rate (MER) was plotted against E_b/N_o for the coded and uncoded messages.

The modem was operated airborne through the polar satellite during the polar flight test. Tests were conducted under conditions of Additive White Gaussian Noise, (AWGN), multipath fading, and scintillation fading.

By varying the attenuator in the receive path, the E_b/N_o was adjusted to give a message error rate curve against AWGN. The dual UHF modem in an uncoded AFSAT I mode gave 100% message copy at an E_b/N_o greater than 12 dB as shown in Figure 87. The coded AFSAT I mode gave 100% message copy at an E_b/N_o greater than 9 dB.

While flying over the Bering Sea, severe multipath fading was encountered. The fading had a peak-to-null depth of 10 to 15 dB and a period of approximately 0.8 seconds (72 fades per minute). The message error rate performance of the coded and uncoded modulation during the multipath fading is shown in Figure 88. The coding provides approximately a 10 dB improvement in performance.

During the test series, the aircraft encountered scintillation fading from just north of Goose Bay, Labrador to Thule, Greenland. Rapidly deep fading from the polar satellite, Figure 89, occurred for over 5 hours. The performance of the coded and uncoded AFSAT modulation is shown in Figure 90. During the fading no more than 30% of the uncoded messages were received error free. The coded messages were received error free until approximately 6 dB of attenuation was added to the downlink path. The 30% error free message copy occurred with approximately 12 dB of attenuation.

c) Multipath Bounce Example

Another technique which offers a pseudo space diversity approach is the use of the bottom mounted aircraft antenna in conjunction with the upper top mounted antenna, Figure 91 (extracted from Johnson et al, 1980). Experimental tests have shown that at relatively high elevation angles sea multipath degrades the signal amplitude only 3 to 6 dB (Johnson 1980). If the aircraft-to-satellite path is 40° elevation or greater, the direct path and sea reflected path propagate through different parts of the ionosphere, resulting in a space diversity. In addition, the sea multipath reflection area is a large area encompassing the first Fresnel zone and inherently displays a space diversity characteristic.

The performance of a UHF satellite link was measured from a top antenna (direct path) and from a bottom antenna (reflected path) under the condition of no scintillation fading. The results of the test are shown in Figure 92 which indicates the bottom antenna performance is approximately 7 dB worse than the top antenna performance. Plots of the top and bottom antenna received signal cumulative distribution function under similar conditions show that the 50% occurrence points are 7 dB apart in signal level. The received signal from the top and bottom antennas with no scintillation fading is shown in Figure 93. The signal received on the top and bottom antennas under severe ionospheric scintillation conditions is shown in Figure 94.

The error-rate performance of signals received on the top and bottom antennas are shown in Figure 95. Note that the performance of the signal received on the top antenna is degraded approximately 7-9 dB from the unfaded performance. The performance of the signal received on the bottom antenna is degraded 0-2 dB from its performance under nonfading conditions. The results show that under severe scintillation fading the reflected signal received on the bottom antenna performs as well or better than the signal received on the top antenna.

2. Multipath Mitigation Techniques

The satellite communications system designer has to weight the cost and complexity of a variety of mitigation techniques to minimize the effects of multipath fading on the path from an aircraft-to-satellite (extracted from Johnson, 1979-2). Some of the techniques that can be considered include the use of space, frequency, time, or antenna illumination diversity.

The fact that the reflection coefficient is dependent upon antenna or signal polarization leads to the obvious conclusion that different signal polarizations will experience different multipath fading depths. Experimental observations have confirmed the predicted results. Measurements of the signal received from a satellite utilizing a linear antenna polarization showed that the multipath fading depth increased sharply as the elevation angle decreased, Figure 96. Multipath fade depths of greater than 20 dB were experienced at very low elevation angles (Jorden, 1969). However, satellites which employ circular polarization provide considerable protection against low angle multipath fading. Extensive measurements are summarized in Figure 96 showing that the multipath fading tends to peak up at a moderately low elevation angle with the fade depth seldom exceeding 10 or 12 dB (Johnson, 1974). As the elevation angle decreased, the fade depth decreased to approximately 5 dB.

If a directive aircraft antenna can be used to minimize the illumination of the reflecting surface, the multipath fade depth can be reduced significantly. The optimum choice of antenna pattern shape and antenna location can emphasize the gain in the direction of the satellite while minimizing the gain in the direction of the horizon. As discussed in Section I.C.1, antenna directivity is the largest multipath fade factor under the system designer's control.

Due to the link geometry, the multipath fading is very frequency selective. As a result, frequency diversity techniques work extremely well to overcome the multipath fading. The results of a triple frequency diversity scheme are shown in Figure 97 (Johnson, 1979-2). The maximum frequency separation from Frequency 1 to Frequency 6 is 300 KHz. The elevation angle at which this multipath occurred was approximately 8°. From Figure 97 it can be seen that when Frequency 1 is in a fade, Frequency 6 is at its

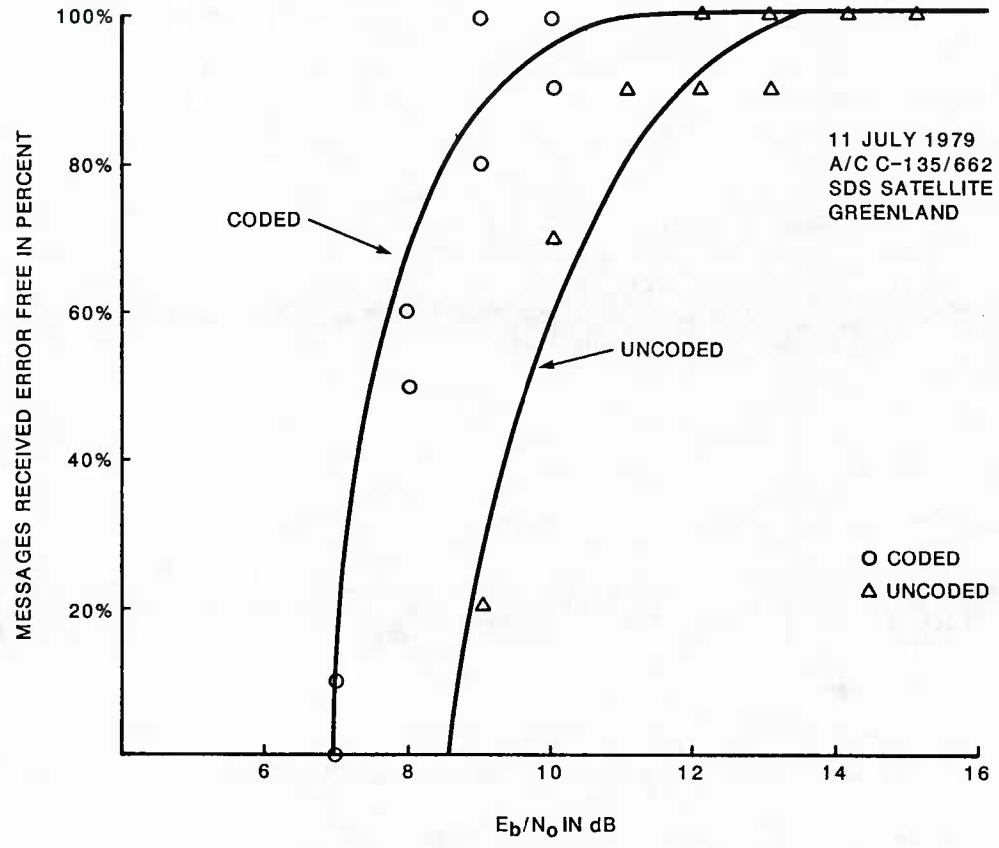


FIGURE 87. MESSAGE ERROR RATE PERFORMANCE AGAINST ADDITIVE WHITE GAUSSIAN NOISE

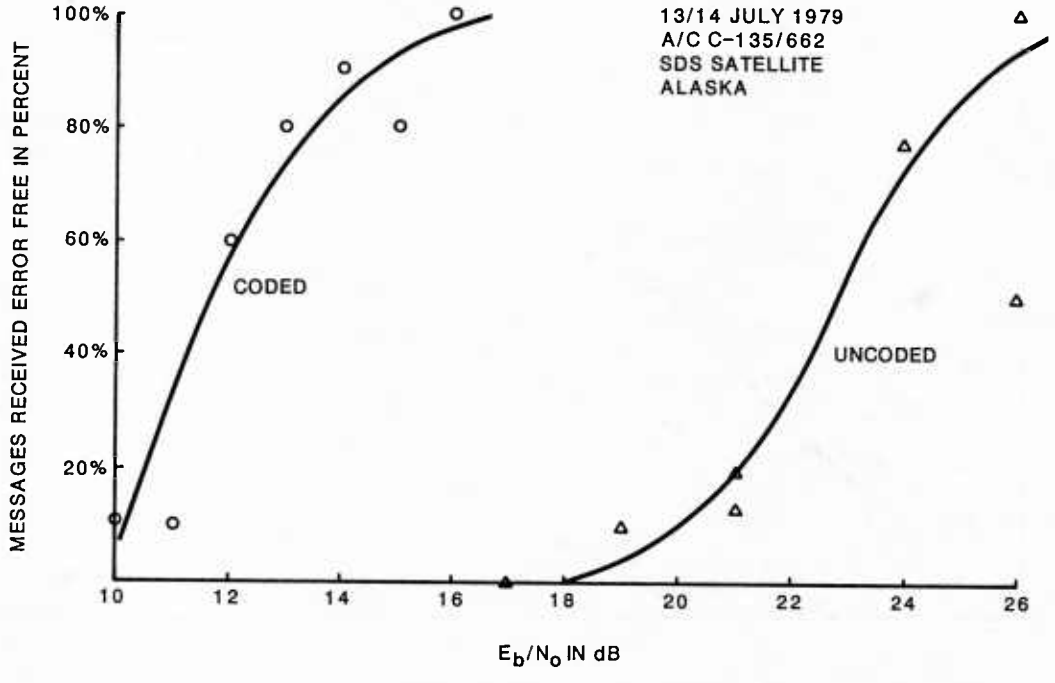


FIGURE 88. MESSAGE ERROR RATE PERFORMANCE AGAINST MULTIPATH FADING

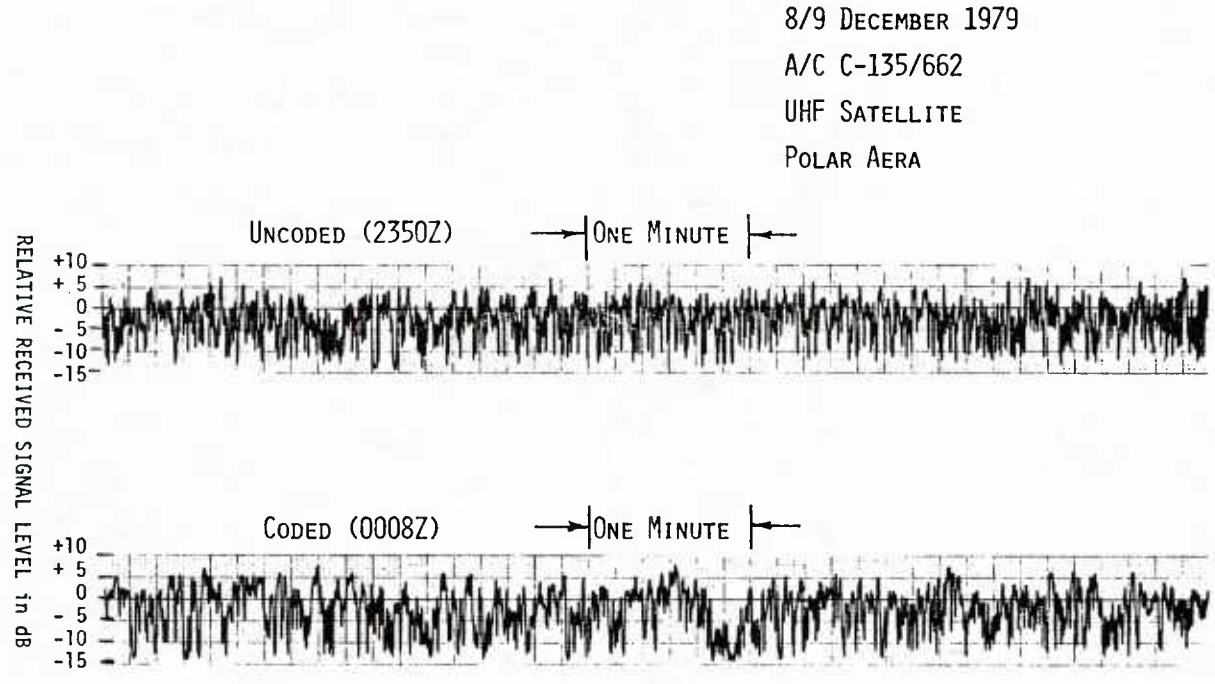


FIGURE 89. UHF RECEIVED SIGNAL LEVEL DURING IONOSPHERIC SCINTILLATION FADING

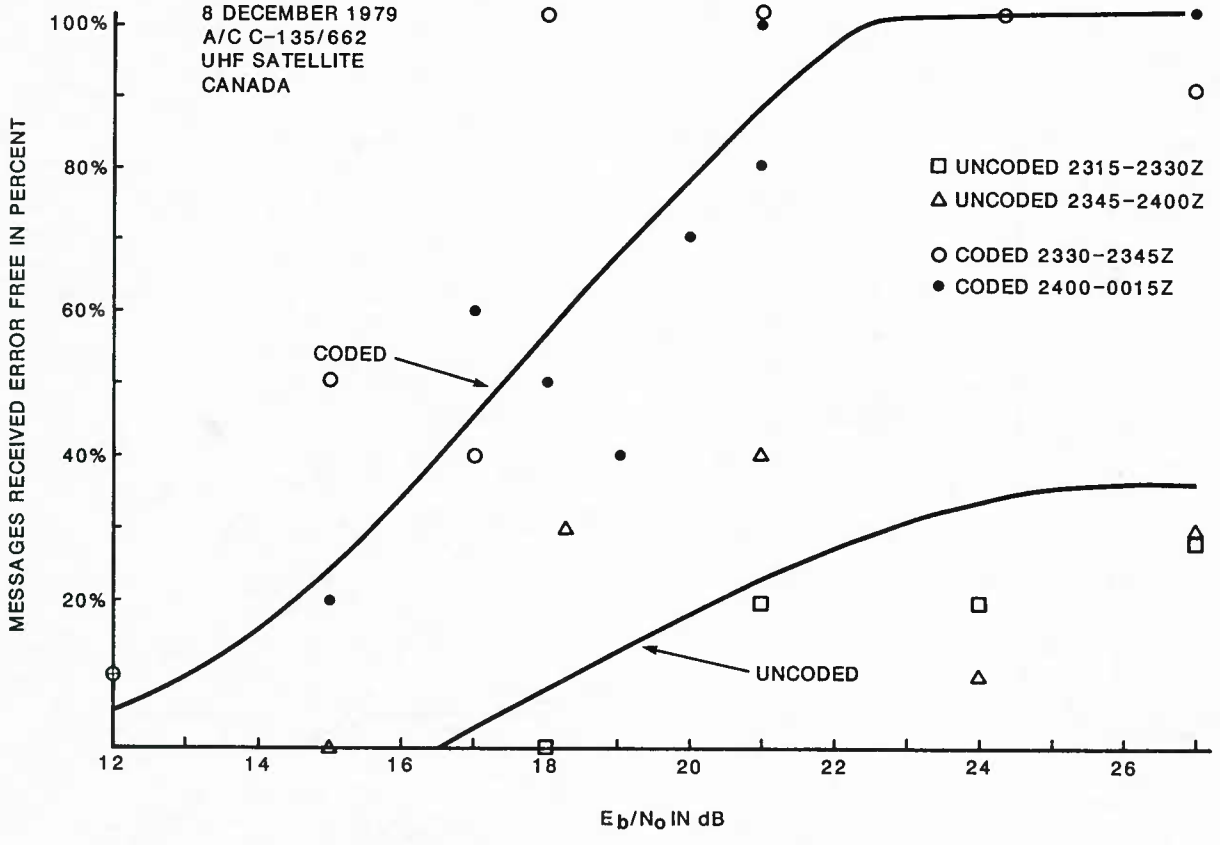
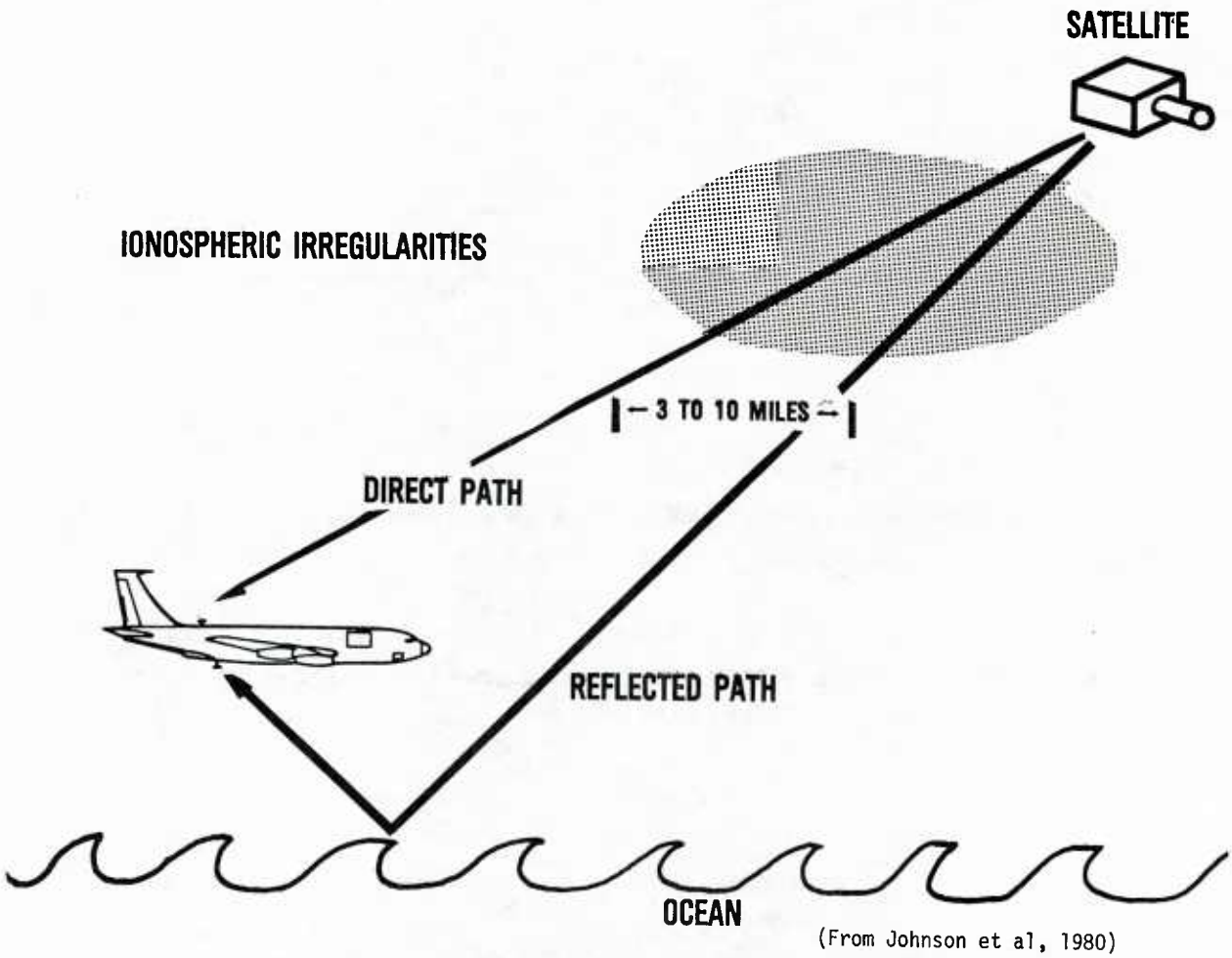
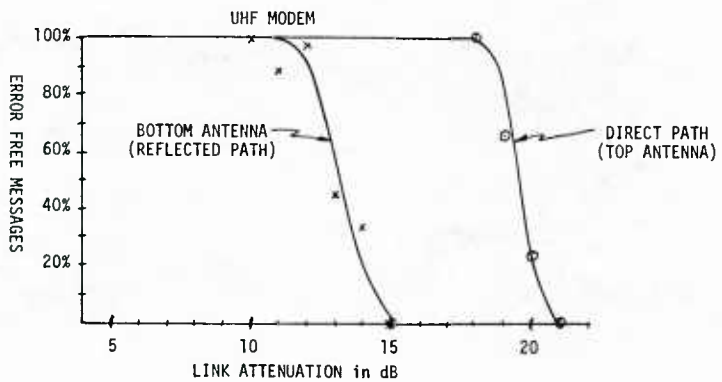


FIGURE 90. MESSAGE ERROR RATE PERFORMANCE AGAINST IONOSPHERIC SCINTILLATION FADING



(From Johnson et al, 1980)

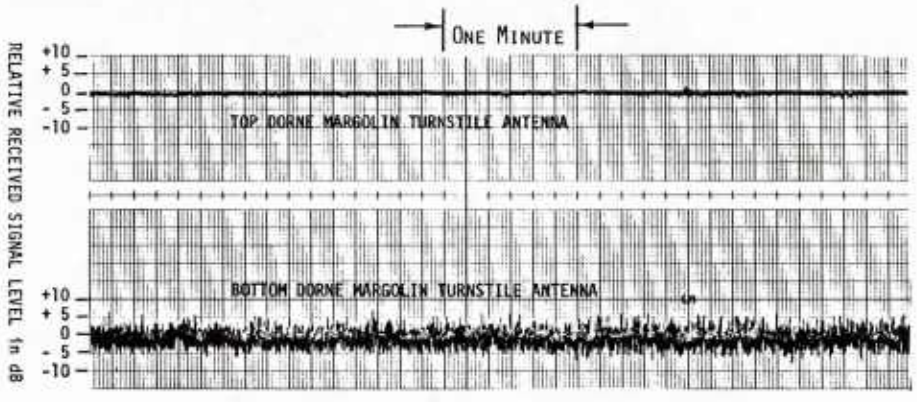
FIGURE 91. GEOMETRY OF FADING MITIGATION TECHNIQUE



(From Johnson et al, 1980)

FIGURE 92. PERFORMANCE OF AFSATCOM II MODULATION WITH EARTH SURFACE REFLECTED SIGNAL

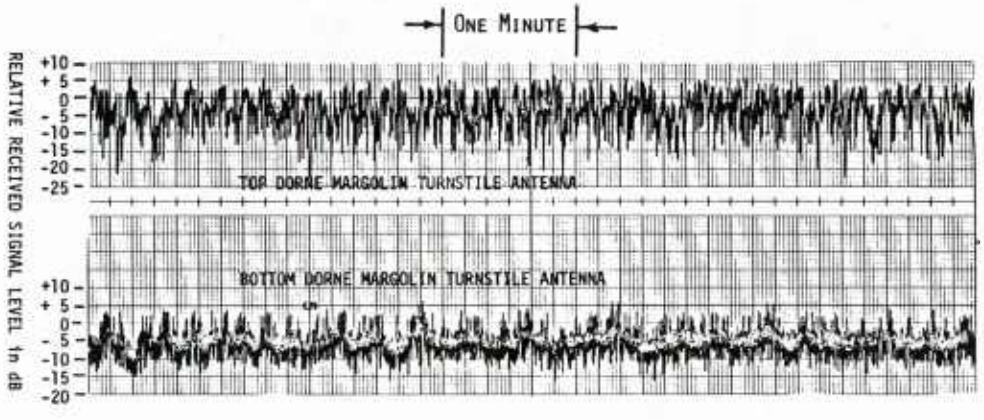
18 March 1979 (0230Z)
Aircraft C-135/662
UHF Satellite (250 MHz)
Peru-Elevation Angle 46°



(From Johnson et al, 1980)

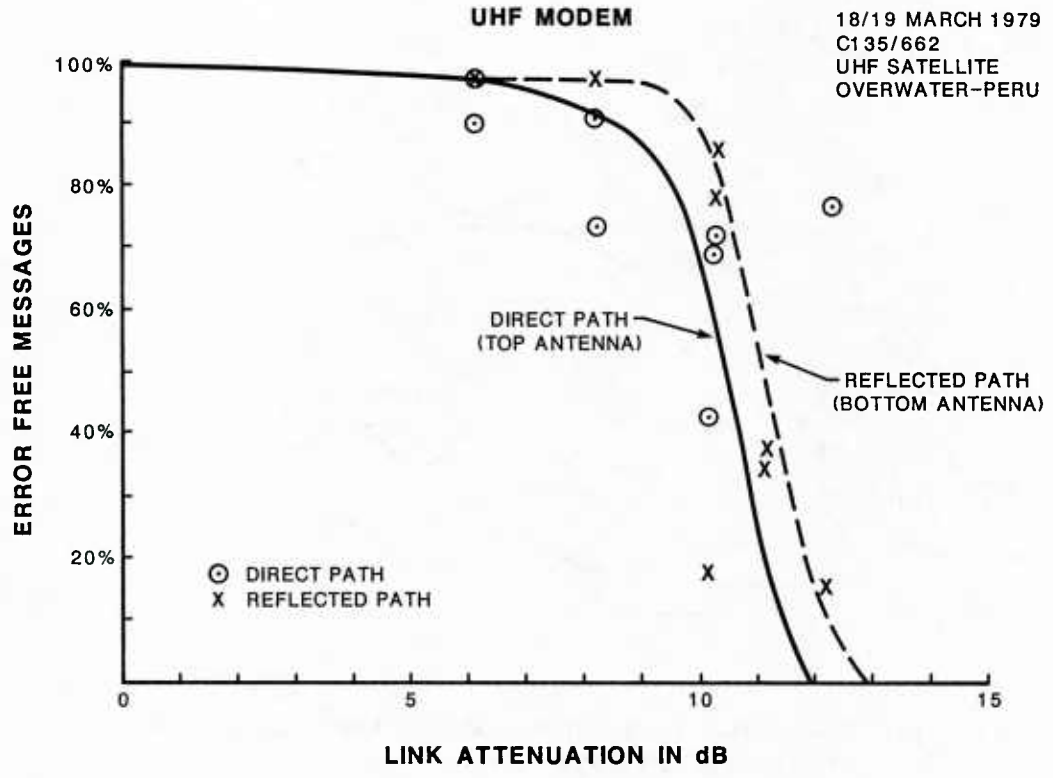
FIGURE 93. RECEIVED SIGNAL LEVEL FROM TOP AND BOTTOM ANTENNA DURING QUIET PERIOD

19 March 1979 (0220Z)
Aircraft C-135/662
UHF Satellite (250 MHz)
Peru- Elevation Angle 48°



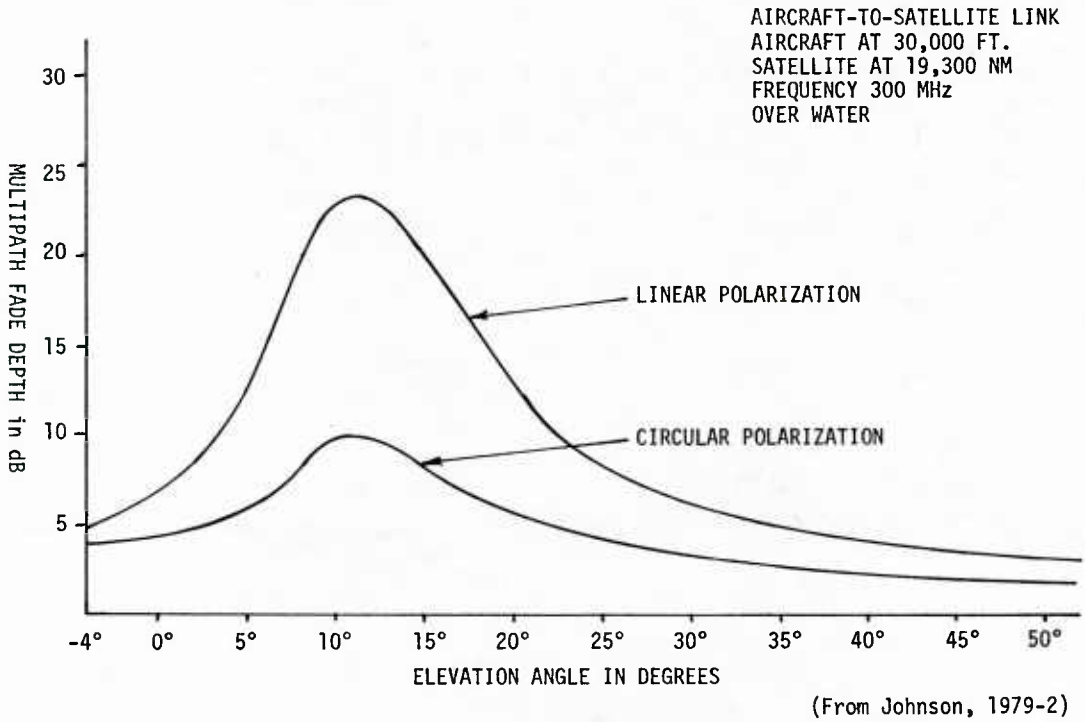
(From Johnson et al, 1980)

FIGURE 94. RECEIVED SIGNAL LEVEL FROM TOP AND BOTTOM ANTENNA DURING SCINTILLATION FADING



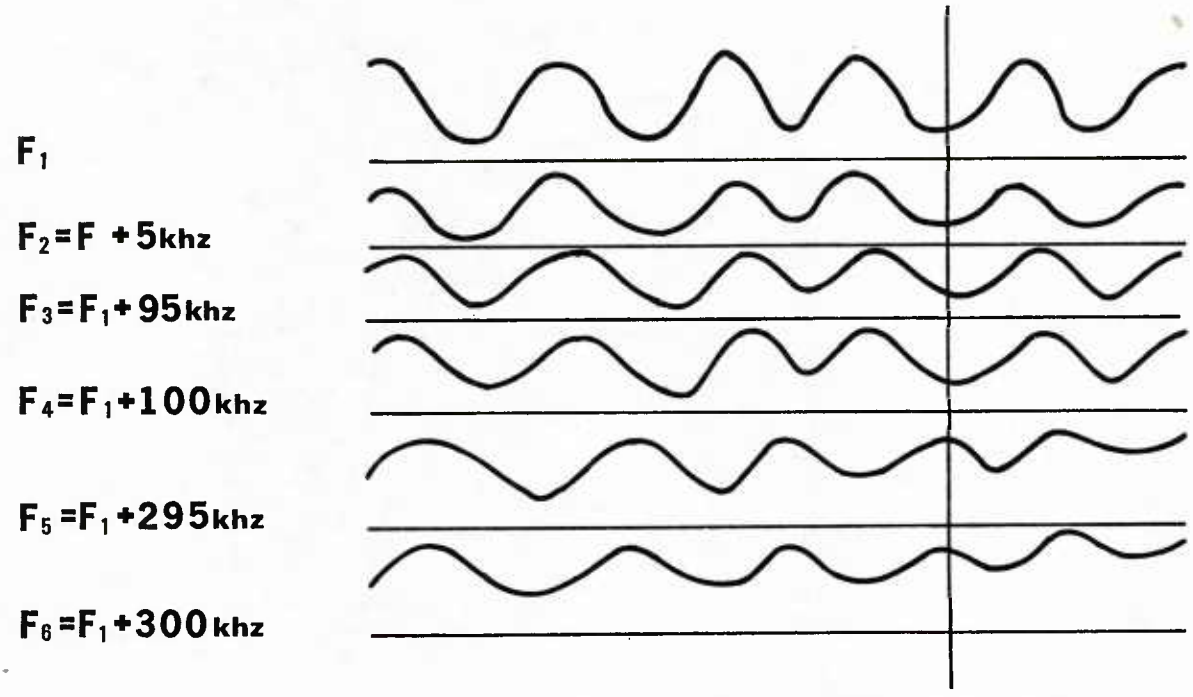
(From Johnson et al, 1980)

FIGURE 95. PERFORMANCE OF AFSATCOM II MODULATION DURING SCINTILLATION FADING



(From Johnson, 1979-2)

FIGURE 96. TYPICAL MULTIPATH FADING DEPTH FOR LINEAR AND CIRCULAR ANTENNA POLARIZATION



(From Johnson, 1979-2)

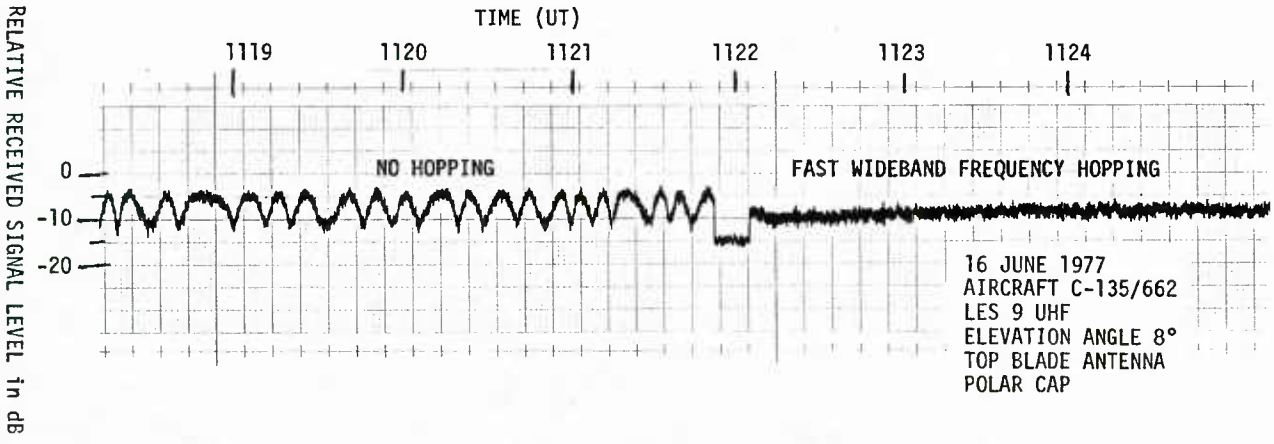
FIGURE 97. FREQUENCY DIVERSITY OF MULTIPATH FADING SIGNAL RECEIVED OVER GREENLAND

multipath peak. The utilization of a diversity combiner can virtually eliminate the effect of this type of multipath fading.

Another example of frequency diversity is rapid frequency hopping of a signal where the hopping rate is greater than the data rate (Johnson, 1979-2). Figure 98 shows the received signal level of an unhopped and a fast-hopped signal. Combining the multiple frequency hopped chips into a single data bit significantly reduces the effect of the multipath fading period. The apparent drop in received signal level during the frequency hopping reception is due to the instrumentation. In fact, the frequency hopping system provided 5 dB improved margin over the non-hopping system.

Space, or antenna diversity, can significantly reduce the multipath fading if the antennas are spaced far enough apart to decorrelate the fading. Vertical spacing of approximately 5 to 10 wavelengths is usually sufficient to decorrelate the fading. Horizontal spacing of 50 to 100 wavelengths is needed to provide similar decorrelation.

Time diversity techniques improve multipath fading performance in the same way they improve scintillation fading performance (see Section II.F.1). The use of error correction coding can provide the needed time diversity. Usually interleaving of data bits is required to randomize the burst errors that multipath fading produce since the error correction coder/decoders can only effectively correct random error patterns.



(From Johnson, 1979-2)

FIGURE 98. MULTIPATH FADING PROTECTION OF FAST FREQUENCY HOPPING SIGNAL

III. Future Trends In Satellite Communications

A. Spacecraft Trends

The major trend in spacecraft is the move from the crowded lower frequency bands to the higher frequency bands which offer wider bandwidths and the possibility of reducing mutual interference by narrow beam antennas. A large number of satellites are planned for the 20 and 30 GHz band as well as the 11 to 15 GHz region. Military satellites in the 20 and 44 GHz band, such as MILSTAR, are also planned.

Satellite transmit power is expected to increase as research continues on high-power solid-state devices. The satellite antennas are moving towards electronically steerable arrays that can rapidly hop a spot beam around the area of coverage. This approach allows frequency reuse in several spot beam areas.

Most future military satellites will contain satellite-to-satellite crosslinks to allow coupling among a constellation of satellites and in some cases coupling to other constellations of satellites.

B. Earth Terminal Trends

Trends in the earth terminals (ground, airborne, and shipborne) include the use of electronically phased arrays to provide a flush-mounted high-gain antenna alternative. Small improvements are expected in the receiver noise temperatures. Low noise preamplifiers already provide gains at noise temperatures that are just a few dB above absolute zero. The frequency of earth terminals is increasing to match the move to higher spacecraft frequencies. This increase in frequency reduces most propagation effects except for atmospheric attenuation.

Another trend is toward to the use of Demand Assignment Multiple Access networks and the use of time division or code division modulation in place of the heavily used frequency division multiple access techniques. Most future earth terminals are expected to utilize coded and interleaved modulation to overcome propagation anomalies and to minimize interference and jamming.

IV. Need For More Propagation Data

Information on atmospheric attenuation and multipath fading is rather well developed and probably only required for special system applications. Additional information is required on ionospheric scintillation phenomena. Worldwide data collection is still necessary to refine the time and frequency of occurrence for the various geographical regions of the world. The theoretical analysis explaining the mechanism that triggers ionospheric scintillation is still in its infancy. Such information would allow much better modeling and provide for a prediction system. Especially lacking are leading indicators that could give communicators a warning as to when, where, and how severe the disruption from ionospheric scintillation would be.

REFERENCES

Aarons, J., J.P. Mullen, and H.E. Whitney, "The Scintillation Boundary," J. Geophys. Res, Vol 74, pp 884-889, 1969.

Aarons, J., "Equatorial Scintillation - a Review," IEEE Trans, Ant. Prop. Vol AP-25 P729, 1977.

Aarons, J., P. Mullen, H.E. Whitney, E.M. MacKenzie, (1980-1), "The Dynamics of Equatorial Irregularity Patch Formation, Motion and Decay," J. Geophys Res, Vol 85, pp 139-149, 1980.

Aarons, J., E. MacKenzie, K. Bhavanani, (1980-2), "High Latitude Analytical Formulas for Scintillation Levels," Radio Science, Vol 15, pp 115-127, 1980.

Aarons, J., "Global Morphology of Ionospheric Scintillations," Proc of IEEE, V 70, pp 360-378, April 1982.

Beckmann, P. and Spizzichino, A., The Scattering of Electromagnetic Waves from Rough Surfaces, MacMillan Co., New York, 1963.

Bem, D.J., "Propagation Aspects in Planning of Telecommunications Services," J. Telecomm 46 (XI), 1979.

Bernal, R., "Coding Techniques to Reduce Ionospheric Scintillation Effects on the AFSATCOM Dual Modem," AFAL-TR-79-1218, Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio, 27 Dec 1979.

Bond, F. E., "Precise Results for Differential Delay and Fading Rate For Aircraft/Satellite Link, Test Report, Aerospace Corp.; Los Angeles, CA; 30 Oct 1967.

Booker, H.G., J. Ratcliffe, and D.H. Shinn, "Diffraction From An Irregularity Screen With Applications To Ionospheric Problems," Phil. Trans. Roy. Soc. Ser. A, Vol 242, p 579, 1950.

Briggs, B.H., and I.A. Parkin, "On The Variation of Radio Star And Satellite Scintillation With Zenith Angle," J. Atmos. Terr. Phys., Vol 25, pp 339-365, 1963.

Buchau, J., Private Communication, Air Force Geophysics Laboratory, Hanscom AFB, MA, November 1983

Burnside, W. D. and Marhefka, R. J., Yu, C. L., "Roll-Plane Analysis of On-Aircraft Antennas, IEEE Transaction on Antennas and Propagation, Vol. AP-21, #6, pp. 780-786; Nov 1973.

Crane, R.K., "Propagation Phenomena Affecting Satellite Communications Systems Operating in The cm and mm Wavelength Bands," Proc IEEE, Vol 59, No. 2, February 1971.

Chinnick, J. H., "A Report On Low Angle Ocean Scatter Measurements," Communication Research Centre Report; Ottawa, Canada, 29 Aug 1977.

Clanton, Steve, "AFSCS Antenna Patterns on B52-H," Boeing Test Report T-3-1687; Renton, Wash.; 25 April 1974.

DasGupta, A., S. Basu, S. Basu, J. Aarons, J.A. Klobuchar, and A. Bushby, "VHF Amplitude Scintillations and Associated Electron Content Depletions As Observed At Arequipa, Peru," J. Atmos & Terr. Phy, Vol. 45, No. 1, pp 15-26, 1983.

DCA, "Satellite Communications Reference Data Handbook," Defense Communications Agency, Code 483, AD 746165, Washington, D.C., July 1972.

Foley, T.K., Gaumond, B. J. and Sestak, E., "Experimental L-Band SST Satellite Communication/Surveillance Terminal Study," Boeing Report D6-60105-2, Vol. III; Renton, Wash.; Nov 1968.

Foshee, J.J., "A Modulation/Encoding Scheme For Use In A Fading Environment," NAECON Symposium, Dayton, Ohio, May 1977.

Franklin, Sidney B., "A Brief Summary of Multipath Effects That Will Tend to Limit RPV-to-Relay Range," AFAL Tech Memo; WPAFB, Ohio, 6 Nov 1973.

Frazier, J.A. Jr, and K.G. Caster, "Viterbi Versus Sequential Decoding in a Scintillated Satellite Communications Channel," NAECON Symposium, Dayton, Ohio, May 1983.

Fremouw, E.J., and J.F. Bates, "Worldwide Behavior Of Average VHF-UHF Scintillation," Radio Sci, Vol 6, pp 863-869, 1971.

Fremouw, E.J., and C.L. Rino, "Modeling Of Transionospheric Radio Propagation," RADC-TR-76-35, Rome Air Development Center, Griffiss AFB, NY, 1976.

Fremouw, E.J., C.L. Rino, A.R. Hessing, and V.E. Hatfield, "A Transionospheric Communication Channel Model," Quart Tech Rep 7, SRI Int., Menlo Park, CA, 1977.

Fremouw, E.J., and C.L. Rino, "A Signal-statistical And Morphological Model Of Ionospheric Scintillation," Proc. AGARD Conf, Ottawa, Ont. Canada, 1978.

Fremouw, E.J., (1980-1), "Geometrical Control Of The Ratio Of Intensity And Phase Scintillation Indices," J. Atmos, Terr. Phys., Vol 42, pp 775-782, 1980.

Fremouw, E.J., (1980-2), "Computer Modelling of Auroral-Zone Ionospheric Scintillation," Progress Rpt No. 4, 5, & 7, Contract DNA001-79-C-0372, Physical Dynamics Inc. Bellevue, WA, 7 Nov 1980.

Fremouw, E.J., and J.M. Lansinger, "Recent High Latitude Improvement In A Computer Based Scintillation Model," Proc. Ionospheric Effects Symp, Alexandria, VA, April 1981.

Goodman, J.M., "Environmental Constraints in Earth-Space Propagation," AGARD Conference Proceedings #284, London UK, 12 May 1980.

Haerendel, G. and A. Valenzuela, Coloured Bubbles Natal Brazil, Max Planck Inst for Physics and Astro Physics, Report 38, Munchen West Germany, September 1982.

Heelis, R.A., W.B. Hanson, J.L. Burch, "Ion Convection Velocity Reversals in the Dayside Cleft," Journal of Geophysics Research, Vol. 81, No. 22, pp. 3803, 1 August 1976.

Heppner, J.P., "Empirical Model of High Latitude Electric Fields," Journal of Geophysics Research, Vol. 82, No. 7, p. 1115, 1 March 1977.

Hocutt, A. E., "Results Of Recent Equatorial Scintillation Testing," NAECON Symposium, Volume 1 pp 543-48, Dayton OH, 17 May 1983.

Ippolito, L.J., "Millimeter Wave Space Communications," Lecture Notes, George Washington Univ, NY NY, Course 501, 1975.

Johnson, A.L., "Flight Test Results of Propagation Experiments through Inhomogenous Media," NATO AGARD Conference Proceedings #191, Boston, MA, January 1974.

Johnson, A.L., (1979-1), "The Effect of Ionospheric Scintillation Fading on Aircraft-to-Satellite Communications", AFAL-TR-78-171, February 1979.

Johnson, A.L., (1979-2), "Airborne Measurements of Electromagnetic Wave Reflections from Land and Sea Water," AGARD Conference Proceedings #269, Spatind, Norway, Paper #11; Sept. 1979.

Johnson, A.L. "Propagation Anomalies Affecting Airborne Satellite Communications, "AGARD Conference Proceedings #284, London UK, May 1980.

Johnson, A.L., R.C. Beach, W.O. Fischbach, R.N. Wright, W.T. Hunt, and R.C. Foster, "South American Flight Test Report," AFWAL-TR-80-1079, Air Force Wright Aeronautical Laboratories, Wright-Patterson AFB, Ohio, October 1980.

Johnson, A., J. Aarons, J. Buchau, J. Mullen, E. Weber, and H. Whitney, "Occurrence of F-Layer Irregularities in the Polar Cap," IES Symposium, Alexandria, VA, 15 April 1981.

Johnson, A.L. and A.M. Hocutt, "Scintillation Results From Ionospheric Modification Experiment, " Radio Science, Vol 19, No 3, pp 741-48, May-June 1984.

Johnson, A.L., "Influence of Ionospheric Irregular Shape and Velocity on the Design of Airborne Satellite Communications Systems," AGARD Conference Proceedings #382, Fairbanks, Alaska, June 1985.

Johnson, A.L. "Cross-Correlation Between Widely Spaced Equatorial Receivers," URSI/COSPAR Beacon Satellite Symposium, Oulu Finland, June 1986.

Johnson, F.S., Satellite Environment Handbook, Stanford University Press, Stanford CA, 1961.

Jordan, Edward C., Electromagnetic Waves and Radiating Systems, Prentice-Hall: New York, 1955.

Jorden, Kenneth L., Jr., "Multipath Characteristics in a Satellite-Aircraft Link at 230 MHz," MIT Lincoln Laboratory Report MS2605, Lexington, Mass., September 1969.

Kerr, D. E., Propagation of Short Radio Waves, MIT Radiation Laboratory Series, Vol. 13, McGraw-Hill Book Publishers, Inc., New York, 1951.

Koster, J.R., "Ionospheric Research Using Satellites - Equatorial Scintillation," Scientific Report #1; University of Ghana, Accra, Ghana; September 1971.

Maroth, V., "Model Radiation Pattern Study AT-1076A Antenna on the KC135 Aircraft," Engineering Report 1610.1, Dorne & Margolin; Westbury, New York; 2 April 1963.

Mass, J., Houminer Z., "Ionospheric Research Using Satellites," AFCRL-TR-69-0317, Air Force Cambridge Research Laboratory, L.G. Hanscom Field, Mass, 28 February 1969.

McDaniel, D.R., editor, "Project SECEDE Final Program Review," Vol II, SRI 2-5462, October 1972.

Millman, G.H., "A Survey of Tropospheric, Ionospheric, and Extra Terrestrial Effects on Radio Propagation Between The Earth and Space Vehicles," NATO/AGARD Conference Proc, Maidenhead England, 1967.

Nichols, B.E., "UHF Fading From a Synchronous Satellite Observed at Kwajalein," ESD TR-74-139, MIT Lincoln Lab, Lexington, MA, 22 March 1974.

Ossakow, S.L., "A Review of Recent Results of Spread F Theory," Naval Research Laboratory; Memo Rpt 2909; Washington D.C.; 9 January 1979.

Paulson, M.R., RUF Hopkins, "Spatial Distribution Characteristics of Equatorial Scintillation," Naval Ocean System Center Tech Rpt 113; San Diego, CA; 2 May 1977.

Paulson, M.R., "Scintillation of Satellite Signals at Guam For 2 Elevation Angles And 2 Frequencies," Proc IES, Alexandria, VA, 1981.

Prettie, C.W., A.L. Johnson, J.M. Marshall, T.A. Grizinski, R.L. Swanson, "Project STRESS Satellite Communication Test Results," AFAL Tech Rpt 77-158, Avionics Laboratory, WPAFB OH, July 1977.

Rastogi, R.G., "Seasonal and Solar Cycle Variations of Equatorial Spread-F in the American Zone," J Atmos Terr Phys, Vol 42, PP593-97, 1980.

Reed, Henry R. and Russell, Carl M., Ultra High Frequency Propagation, Boston Technical Publishers, Inc., Lexington, Mass., 1964.

Rino, C.L., R.C. Livingston and S.J. Matthews, "Evidence for Sheetlike Auroral Ionospheric Irregularities," Geophys Res Ltr 5; 1039-1042, 1978.

Rino, C.L., "A Power Law Phase Screen Model for Ionospheric Scintillation," Radio Science, Vol 14, No. 6, pp 1147-55, November 1979.

Rothmuller, I., "CPM/P Scintillation Model, "Interim Report #2, Linkabit Corp., San Diego CA, 18 May 1983.

Sagalyn, R.C., M. Smiddy, and M. Ahmed, "High-latitude Irregularities In The Topside Ionosphere Based On ISIS-thermal Probe Data," J. Geophys. Res., Vol 79, P 4252, 1974.

Singleton, D.G., "Dependence Of Satellite Scintillation On Zenith Angle And Azimuth," J. Atmos, Terr. Phys. Vol 32, pp 789-803, 1970.

Singleton, D.G., "Predicting Transionospheric Propagation Conditions," Electronics Research Laboratory Tech Rpt 0049; Salisbury, South Australia, Nov 1978.

Slack, F.F., "The Ringing Irregularity In Ionospheric Scintillation," AFCRL-68-0263, Air Force Cambridge Research Laboratories, L.G. Hanscom Field, Mass., May 1968.

Spilker, J.J., Digital Communications by Satellite, Prentice-Hall, Englewood Cliffs, NJ, 1977.

Tsunoda, R., "Occurrence of Equatorial Scintillation Fading," URSI/COSPAR Beacon Satellite Symposium, New Delhi, India, February 1983.

Vestine, E.H., L. Laporte, I. Lange, and W.E. Scott, "The Geomagnetic Field, Its Description and Analysis," Publication No. 580, Carnegie Institute of Washington, Washington DC, 1947.

Weber, E.J., J. Buchau, J.G. Moore, "Airborne All-Sky Imaging of Equatorial Airglow," Air Force Geophysics Laboratory TR-78-0224; Hanscom AFB, MU - 21 Sept 1978.

Whalen, J., "Auroral Oval Plotter and Nomograph for Determining Corrected Geomagnetic Local Time, Latitude, and Longitude for High Latitudes in the Northern Hemisphere," AFCRL-TR-70-0422, Hanscom AFB, MA, 27 July 1970.

White, D.P., "A Time Diversity Coding Experiment for a UHF/VHF Satellite Channel with Scintillation & Equipment Description," ESD-TR-77-217, Electronic Systems Division, Hanscom AFB, MD, 1 September 1977.

PROBLEMS OF SPECTRUM MANAGEMENT AND SYSTEM COMPATIBILITY

BY

T.K. FITZSIMONS

ALLIED RADIO FREQUENCY AGENCY,
IMS, NATO HQ, BRUSSELS

I. INTRODUCTION

The problems of radio frequency spectrum management stem basically from requirements that are increasing more rapidly than our ability to exploit certain parts of the spectrum. The expanding demand is caused partly by the transition from analogue to digital transmission methods but of course also to the increase in communications needs and uses. Much of the increase in demand is for allotments and assignments in frequency bands already heavily exploited. New frequency bands are periodically added to the table of allocations by the ITU (International Telecommunication Union) and the rate at which this has happened in the recent past is itself an indication in the almost explosive growth in the uses of the radio frequency spectrum. Table 1 shows the way in which the allocated spectrum space has increased.

TABLE 1

The Allocated Radio Frequency Spectrum

Year	Allocated Spectrum
1927	10 kHz - 30 MHz
1932	- 60 MHz
1938	- 200 MHz
1947	- 10.5 GHz
1959	- 40 GHz
1971	- 275 GHz
1979	- 400 GHz

The ITU had seen that there was a need to agree to some simple allocation arrangements as early as 1906. Since the first formal allocation table in 1927 the increase has been very rapid. It is interesting to note that at the end of the second World War, the allocated spectrum for world-wide use was of the same order as that presently used for some individual systems or at some busy communications cosites. Arguably there is sufficient spectrum allocated to support the operation of C-E systems for a long time to come but of course propagation and technology problems preclude the use of much of the r.f. spectrum by systems having particular operational and bandwidth requirements. Some of the problems, both of technology and propagation will be overcome but it is likely that military C-E systems will, for some years to come, make limited use of bands above say 20 GHz. It can be dangerous to prophesy though since past experience has shown that as the frequency bands were opened up for use, problems were overcome, systems were adapted or invented to exploit the new parts of the spectrum, and that those applications then expanded to fill the spectrum space available.

2. MILITARY RADIO COMMUNICATIONS

The first radio communication systems were extremely efficient in their use of the spectrum. Morse and amplitude modulated carrier systems are still widely used today; Morse of course being particularly efficient. With analogue systems such as amplitude modulated voice transmission the transmitted bandwidth can be approximately equal to the information bandwidth. Even before the advent of digital systems, problems of sharing frequency bands between unlike systems surfaced with the introduction of band spreading systems based on frequency modulation. These systems have within themselves the advantages that go with the trade-off between bandwidth and signal to noise input at the receiver but given the simple frequency management tools of the recent past and uncertainties of the order of some tens of dBs regarding wanted and unwanted field strengths, particularly in dynamic environments, assigning frequencies to systems having different bandwidths, power flux densities and receiver sensitivities has had its problems. These problems have been compounded in recent years by the introduction of digital systems where not only is there usually bandwidth spreading during the modulation process but an expansion of the information (or baseband) bandwidth at the analogue to digital coding stage. Typical coding standards are 16 kB/s for military voice channels and 64 kb/s for civil voice channels. Although the coded baseband can be filtered and then be used to modulate a carrier so as to be contained in a bandwidth corresponding to about 1 Hz/bit, the cost in terms of power and receiver complexity are such that systems with ratios of about 2 Hz/bit are more common. Thus the r.f. transmission

bandwidth for a voice channel may be well in excess of 25 kHz. Again the benefits of these systems themselves are well known but it can mean that the frequency manager may have to establish the conditions for compatibility when there are several types of radio system in the same band even when the operation being performed is the same for all of the systems. An example of this can be found in the military frequency band, 225-400 MHz, where equipments and systems supporting tactical air/ground/air communications can be any of the following:

analogue voice, double-sideband amplitude modulation

digitalised voice, frequency modulation

digitalised voice, fm or fsk, frequency hopping

digitalised data, FM or fsk, fixed frequency or frequency hopping.

All of these methods are presently in use for communications between pilots and ground controllers, and the first three may be alternative means to perform exactly the same function - a ground controller talking to a pilot and vice versa.

3. FREQUENCY MANAGEMENT RESPONSIBILITIES AND ALLOCATIONS

It may be useful to know where the responsibility lies for solving these problems when they are caused by the deployment of military systems, particularly in the NATO environment where the forces of several nations are involved. The International Telecommunication Union (ITU) is a collection of member states represented at policy making meetings by civil administrations, normally the Ministry of Posts and Telecommunications. At world or regional conferences of the ITU, decisions are made regarding how the various parts of the radio frequency spectrum are to be used. NATO is not a recognised body at the ITU and cannot appear. For sure, there is a military influence to a lesser or greater degree in the national positions taken by the representatives. In the case of the NATO nations this influence may take the form of a coordinated position agreed by all the NATO nations' representatives. This coordinated NATO position will have been developed during joint civil/military meetings held under the auspices of the Allied Radio Frequency Agency. Although all the NATO nations recognise the need of their own forces, and of guest forces to have frequency spectrum allocated to them in order to support military C-E activities, the large demands for other uses such as broadcasting, transport, business and private communications, mean that the amount of spectrum space allocated for military use is very limited. Table 2 shows all of the bands within which there are allocations for military use for all C-E purposes.

TABLE 2

OVERVIEW OF MILITARY USE OF FREQUENCY SPECTRUM

Band	Application
9 KHz-1605.5 KHz	Maritime mobile surface and sub-surface tactical broadcast, radio navigation for aeronautical and maritime forces
1605.5 KHz-30 MHz	Tactical communications nets, command and control nets, A/G/A communications
30-108 MHz	Tactical, ground mobile, frequency modulated low power voice radio
108-225 MHz	Aeronautical Mobile (OR)
225-400 MHz	Radio Relay - A/G/A - Tactical Satellites
400-960 MHz	Fixed and Mobile (tactical radio relay)
960-1215 MHz	Aeronautical Radionavigation (WORLDWIDE)
1215-2700 MHz	Tactical Radio Relay - Radiolocation
2.7-3.6 GHz	Radiolocation - Fixed and Mobile
3.6-5.0 GHz	Radio Relay (4.4-5.0 GHz military band in most NATO NATIONS)
5-8.5 GHz	Fixed Satellites (7/8 GHz band)
8.5-15.23 GHz	Fixed Satellites (11/12 GHz band) - RPVs
15.23-17.7 GHz	Fixed, Mobile, Radiolocation
17.7-31.0 GHz	Fixed Satellites (20/21 GHz band, 30/31 GHz band), Fixed Mobile Radiolocation

31.0-51.4 GHz	Fixed Satellites (39/40 GHz band, 50/51 GHz band), Fixed Mobile, Radiolocation
51.4-105 GHz	PLANNED MILITARY USAGE UNDER REVIEW

Most of these bands are used by the military on a shared basis and some may be available only in wartime. Only a very few of these bands are exclusively military bands. One notable one is the band 225-400 MHz which is designated as a military band (with the exception of the ILS glidepath sub-band) in NATO Europe and is managed by the Allied Radio Frequency Agency on behalf of the nations. So far there has been little exploitation of the military bands above 10 GHz while there is considerable congestion in bands below 1 GHz.

The relatively small availability of frequencies, to the military, below 1 GHz, together with the special requirements of high mobility, physical and electronic concealment often means that there is no real choice of frequency band and the military frequency management community are forced to accept the consequences of a high density of assignments to systems that use the spectrum in different ways and have different performance level requirements.

With the exception of one or two bands such as the 225-400 MHz band, the use of frequency bands may not be exactly the same in all nations. In general though it can be stated that there is less than 170 MHz of total spectrum suitable for use by tactical, mobile systems. This is based on the assumption that tactical, rapid deployment or mobile systems must operate in frequency bands where diffraction can occur and it is not necessary to establish line of sight conditions. This spectrum space is shared between Fixed and Mobile systems, Radiolocation and Satellite systems.

To obtain assignments or allotments within the sub-bands listed in Table 2 application must be made to the host nation, i.e. where the system is to be used, unless the 225-400 MHz band is to be used. Another exception applies to satellite systems for which there is a notification system laid down by the ITU. Accommodating systems in the 225-400 MHz band provides examples of most of the problems encountered by frequency managers at the allocation, allotment and assignment stages when dealing with:

- an oversubscription of requirements
- a variety of systems
- different transmission methods within systems
- differing operational constraints, and
- differing communication demand patterns.

4. THE MILITARY BAND 225-400 MHz

The problems encountered in the band are good examples of those encountered where bands must be shared by different types of operation by systems using different modulation and using the spectrum in very different ways. The band is designated by the ITU for used by Fixed and Mobile services and by Satellite services and NATO nations have agreed to its use by only military systems in NATO Europe. Within the broad categories of Fixed and Mobile, a very diverse set of systems have developed as shown in Table 3. These have to use the band by being restricted to certain allotments or by sharing the same frequencies - usually on a co-equal basis.

TABLE 3

PRINCIPAL SYSTEMS USING THE BAND 225-400 MHz in NATO Europe

System	Modulation	Bandwidth	Circuit Utilisation	Circuit Length
A/G/A	Analogue Digital	6 KHz 25-50 kHz	±1%	1 km-400 km
NAVY	Analogue Digital	6 KHz 25-50 kHz	1-10%	1 km-50 km (Ship-Ship) 1 km-200 km (Ship-Air)
RADIO RELAY	Analogue Digital	500 KHz 1 MHz 1.5 MHz	Continuous	20-40 km
SATELLITE	Digital	5 KHz-500 kHz	Intermittent	(Between Geostationary orbit and aircraft and ships)

The total spectrum requirements of these systems exceed the spectrum available. Solutions are achieved by some sharing between different systems but mostly by the application of well-developed frequency assignment methods to accommodate each of the systems within a separate allotment of frequencies or sub-bands.

Air-Ground-Air Systems

Within NATO Europe, almost all air/ground/air (A/G/A) communications use frequencies in the 225-400 MHz band. ARFA makes assignment plans in which over 7000 A/G/A requests are assigned at any one time. These assignments are for tactical operations, air traffic control and various other functions. The circuits are analogue (clear voice, with a nominal bandwidth of about 6 KHz) and digital systems (data and coded voice of 16 kb/s). For the most part the digital systems contain the r.f. bandwidth within 25 KHz and although in the minority at present it is likely that within a few years most A/G/A transmissions will be digital. Many of the analogue equipments are based on old system designs and reflect a time when there were fewer pressures on spectrum space and the equipment tuned in increments of 50 and 100 KHz. Using double-sideband amplitude modulation the analogue systems have always been efficient in their occupancy of the band but in the past the receivers were not designed with a correspondingly narrow selectivity perhaps partly because it was not considered necessary, partly to accommodate transmitters with low tuning accuracy and stability, and partly due to the use of offset carrier systems (en-route air traffic control) with large offsets. Figure 1 shows the receiver selectivity characteristics of some equipments presently in service. It can be seen that there has been a considerable improvement in reducing receiver selectivity since the introduction of 50 KHz and then 25 KHz equipments. The improvement has not been pro-rata with the tuning interval; for example the rejection of a 50 KHz adjacent channel by a 50 KHz equipment is not as good as for a 100 KHz tunable equipment where the first adjacent channel was (originally at least) 100 KHz away. The introduction of digital systems which have to be assigned with frequencies from the same air-ground-air resources has lead to rather unequal adjacent channel rejection characteristics. This increases the separation distance required between adjacent channel assignments. There is no compensating improvement in packing on co-channel assignments since although a wideband transmission can be partly rejected by a narrow band receiver, the decision for co-channelling will probably be based on the narrow band transmission entering the wideband receiver.

Difficulties in managing the air-ground portion of the 225-400 MHz band have been compounded recently by the introduction of frequency hopping systems. With conventional fixed frequency systems, some of the problems of non-compatibility are solved by the frequency assignment process. This is more difficult when frequency hopping or direct sequence spread spectrum systems share frequency assets with fixed frequency systems. Frequency managers are facing up to this problem by trying to establish the conditions under which these systems can operate in a mixed environment. Variations in parameters such as power, antenna gain, receiver sensitivity, which can be taken into account in the calculations for making compatible assignments to fixed frequency systems, even highly mobile ones, are difficult to accommodate when one of the systems is frequency hopping as well. Although the 225-400 MHz band is the principal band for A/G/A operations, it has to support several other operations, some in the same environment as the aircraft and their ground stations.

Radio Relay Systems

These are mostly tactical (rapidly transportable) systems but fixed systems are eligible to use the band as well.

The systems are digital trunk-bearer systems in the form of area networks connecting computer switching centres and giving access to large groups of subscribers. The trunk bit rates are normally 512 kb/s or 1024 kb/s giving rise to occupied spectra of around 0.8-1.5 MHz. The radio relay transmitters have outputs of 20-40 watts so that the average spectral power sensitivity is much lower than is the case for air-ground-air communications. Almost a third of the available spectrum in the 225-400 MHz band is allocated to radio relay either exclusively or on a shared basis. Even so, the demand is so high that even within a Corps area it may be necessary to re-use the frequencies assignable to radio relay.

Navy Communications

The naval use of the 225-400 MHz band is for ship-ship, ship to shore and ship to aircraft operations. The use may be at sea, in harbours, in support of an amphibious operation or for communicating between ships and aircraft flying over land. The types of radio link are similar to that use by the air forces and there is a mixture of analogue and digital circuits and a trend to frequency-hopping systems.

Tactical Satellite

Some subbands within the 225-400 MHz range are allocated for tactical satellite systems. Both low orbit and geosynchronous satellite systems are in operation. The ground stations are fixed and mobile. The mobile ground stations can be on ships, in aircraft or land vehicles. Digital systems are the most comon with TACSAT systems; they provide channels to carry secure voice and low to medium speed data. Typical spectral power densities are -25 dB to -45 dBW/Hz in the space to earth direction and -15 to -25 dBW/Hz in the earth to space direction.

Other systems can also operate in the 225-400 MHz band but the main ones have been described. Problems arise from the very heavy subscription for the limited spectrum resources and from the mixture of systems and modulation methods which are often not very compatible with each other.

5. THE PROBLEMS OF SHARING FREQUENCIES BETWEEN DIFFERENT SERVICES

The problems of sharing limited spectrum resources within the tactical A/G/A system have been examined. The problems become more numerous and more intractable when different systems have to use the same allotment of frequencies. Even in the civil arena and when dealing with fixed systems there are difficulties in making compatible frequency assignments. In the tactical environment the high mobility of individual units and the complete changes of scenario make the problem complex but at the same time give the opportunity, periodically, to completely recast the assignment and effect solutions. In a simplified form, the assigning of frequencies to particular locations and establishing interference-free conditions out to a certain distance, can be regarded as analagous to packing a large number of spheres (or cylinders) of different sizes into a finite volume formed by two-dimensional area and frequency (Figure 2). The sizes of the spheres will be functions of several parameters such as power, bandwidth, receiver threshold etc. and therefore also a function of the type of system (whether analogue or digital) and the modulation characteristics. There may be a single definitive solution to this exercise which results in a minimum occupied volume or there may be a limited number of solutions to packing in these spheres which meet a performance criterion - the obvious one being that all the spheres are contained within the volume. If there is space remaining within the 3-dimensional area/frequency volume, there is always the possibility of accommodating new assignments but this may require that one goes through a process of unpacking and repacking i.e. all the users of that part of the spectrum are re-assigned in order to accommodate new ones. This is an opportunity which presents itself in military tactical systems but it is not a practical possibility when dealing with fixed systems, such as broadcasting or PTT trunk systems especially where they form part of the nations' infrastructure and a system shut down to re-set all radio frequencies may be unacceptable.

The possiblity to re-assign tactical frequencies at relatively short intervals can of course not be divorced from the need to do so for certain systems such as combat net radio or tactical radio relay. Indeed this is cental to the management of these systems. It is however another reason why the same frequency resources may not be shared with systems which are unable to respond to rapid changes, e.g. en-route military air traffic control. Another reason for not being able to use the frequency spectrum as a single set of resources without allocating exclusive parts of it to different systems is that the problems of data inflow and outflow, appropriate levels and mechanisms of management, computational tools (in terms of hardware, engineering criteria and assignment algorithms) are presently too great. Establishing the ground rules for example for sharing frequencies between radio relay and A/G/A is difficult given the differences in certain of the parameters which strongly influence the assignment decision:

	A/G/A	Radio Relay
Location	Ground-Fixed, known Aircraft-Mobile (Mach 1-2)	Stationary, not fixed
Transmitter Power	Ground: 300W-1 KW Aircraft 20W	20-40W
Radiated Bandwidth	6 KHz (analogue) $\pm$25 KHz (digital)	0.5, 1.0, 1.5 MHz (digital)
Antenna Beamwidth	Omnidirectional	$\pm 60^{\circ}$

In addition to these differences, the A/G/A receiver threshold for analogue voice though established at 10 dB S/(N + I) by standard tests, may be as low as 3-5 dB for expert listeners like trained pilots. The radio relay thresholds may also be nominally of the order of 10 dB for a bit error rate of 10^{-2} but the real threshold will depend on the type of interference, its temporal characteristics, whether the transmission is digitalised voice or data, and in the latter case on the use and type of error correction. Finally, of course, the threshold for communication acceptability may be different to that for bit-symchronisation and for crypto synchronisation.

Given then the task of assigning frequencies to fairly large systems in the same band the problem of achieving an optimum or near optimum solution, which may be necessary when resources are limited is one of solving a large number of equations with not too many variables but many inter-relationships.

The problem has been likened to that of the travelling salesman or colouring maps where finding or approaching the optimum solution may require an excessive amount of time.

This examination of the 225-400 MHz military band illustrates some of the problems facing frequency managers. The increase in existing services together with new application of C-E systems will require that new approaches to frequency management and equipment design are taken.

6. HOW TO IMPROVE THE SITUATION

Assuming that it is not possible to reduce the numbers of equipments and systems and that the numbers of systems and C-E applications will in fact continue to increase, some improvements and changes are necessary if frequency management especially of military tactical systems is to continue to be feasible. It must be assumed also that the other main causes of the problem: the increasing use of band spreading techniques of different types and the continued existence of analogue systems will continue. The problem can be tackled from both the frequency management and the system design aspects.

Frequency Management

On the question of frequency management, there are possibilities for improvement at every level. At the strategic level, when future systems are still at the feasibility stage, it is established practice to assess candidate frequency bands and express preferences or agreement. The assessment is normally done on the basis of matching some anticipated nominal system characteristics to broad descriptions of an existing e.m. environment in a particular frequency band. It is exceptional to perform detailed computational studies at this stage to determine the effect on existing systems in different candidate frequency bands. The reasons are obvious: at the feasibility stage many system and equipment parameters are not known and will be determined later by operational and wanted performance considerations. However, short of embarking on a long set of detailed EMC analyses, it is forseeable that in future computer-based tools will exist for assessing frequency bands so as to indicate, perhaps in parametric fashion, what is the remaining capacity and roughly what would be the most favoured spectrum characteristics for new systems.

The next stage in defining the new system design should be to address the problem of Electromagnetic Compatibility. Power levels, modulation methods and characteristics, filter designs, antenna characteristics, should be decided not only on the wanted performance and cost but also on the unwanted performances.

Once the system is fielded, then the frequency allotment and frequency assignment processes can help. Computers have been used for many years now to make frequency assignments; they are just about to be introduced at the battlefield level. A need is emerging for agreement on new management methods which recognise the need to be able to exchange frequency data in the battlefield. One can imagine for example expert systems being used to offer solutions to commanders and national authorities for providing frequency allotments to contingency forces. Data into and out from these decision-making centres would be in standardised format and be transmitted to destinations over the switched area networks. At the appropriate level these resources would be used in very simple assignment systems using perhaps algorithms previously developed specifically for a battlefield environment where data flow will be limited but quick decisions needed.

Also on the frequency assignment side, there is room for improvement even if making assignments centrally to permanent systems. It is possible to establish a lower limit to the number of frequencies needed to assign to a given system. The present algorithms used to choose the assigned frequencies do not approach this limit closely when dealing with large systems especially when the frequency constraints are almost equally divided between cosite and intersite. The Allied Radio Frequency Agency is working on this problem; if successful, the outcome may be applicable to a variety of situations.

System and Equipment Design

It is obvious that there is much that can be done by the system designers to ease the problems of the frequency manager though it must be admitted that the interests of the frequency manager and designer are sometimes in conflict especially when good spectrum characteristics cost money or lower the performance. Apart from participation in the studies described in the preceding section, the equipment designer can help in some obvious ways.

Band Occupancy

Acknowledging that band spreading systems such as binary f.m. have to be spread a certain amount to achieve the benefits of this type of system the filtering and the modulation characteristics should nevertheless be chosen with the e.m. environment in mind. Ideally the main transmitter spectrum should be rectangular as should the receiver response. In practical terms having steep-sided spectra and receiver responses will require more complex filtering and must be matched by comparable stability. Consideration will no doubt be given in the future to the use of narrow band digitised voice systems or perhaps the use of speech synthesised from coded information which occupies little more than the basic information bandwidth.

Adjacent Channel Characteristics

In determining the shape of the receiver selectivity curve, account must be taken of the need to make assignments on the adjacent channels. Figure 3 shows two volumes within which the same frequency can be used. The minimum separation between them can be calculated using the appropriate propagation equation and the desired to undesired signal (D/u). An equal number of assignments can be made on the adjacent channel if the adjacent channel volumes can interleave with the co-channel ones. To do this the minimum adjacent channel separation distance D_{AC} should not exceed:

$(D_{cc} - D_w)/2$

where D_{cc} is the co-channel distance

and D_w is the distance between a receiver and the wanted transmitter.

If D_{cc} is very much larger than D_w then D_{AC} approaches $D_{cc}/2$ and this requires only a very modest adjacent channel rejection performance by the receiver. In a free-space loss environment the adjacent channel rejection would need to be only about 6 dB down on the co-channel. As the co-channelling distance decreases relative to the wanted distance, the required adjacent channel rejection increases rapidly (Figure 4). This occurs either because D_w is inherently large or because the assigner is forced to decrease D_{cc} and accept a finite probability of co-channel interference.

Spurious Output and Responses

Present day equipments already have sufficiently low spurious outputs and spurious receiver responses that for the most part they are not deciding factors in making frequency assignments. However transmitters may have a "noise floor" which can, when the transmitters are deployed in large numbers, produce a cumulative degradation to receivers especially when co-sited with receivers of the systems. This cumulative effect can be observed with direct-sequence (pseudo noise) spread spectrum transmitters and here the extent to which the spectrum is spread should again be determined at least partly by the existing environment and not just the wanted anti-jam margin.

Choice of Frequency Band

Particularly for tactical systems there is probably a "best" frequency band for a radio communication system. This "best" band is probably already over-subscribed and consideration should be given to using higher frequency bands which may yet be under-utilised and where the use of directive antennas can minimise at least the inter-site problems. Technical and propagation problems may arise especially for mobile or transportable systems. Sooner or later, necessity will provide the driving force to overcome these difficulties.

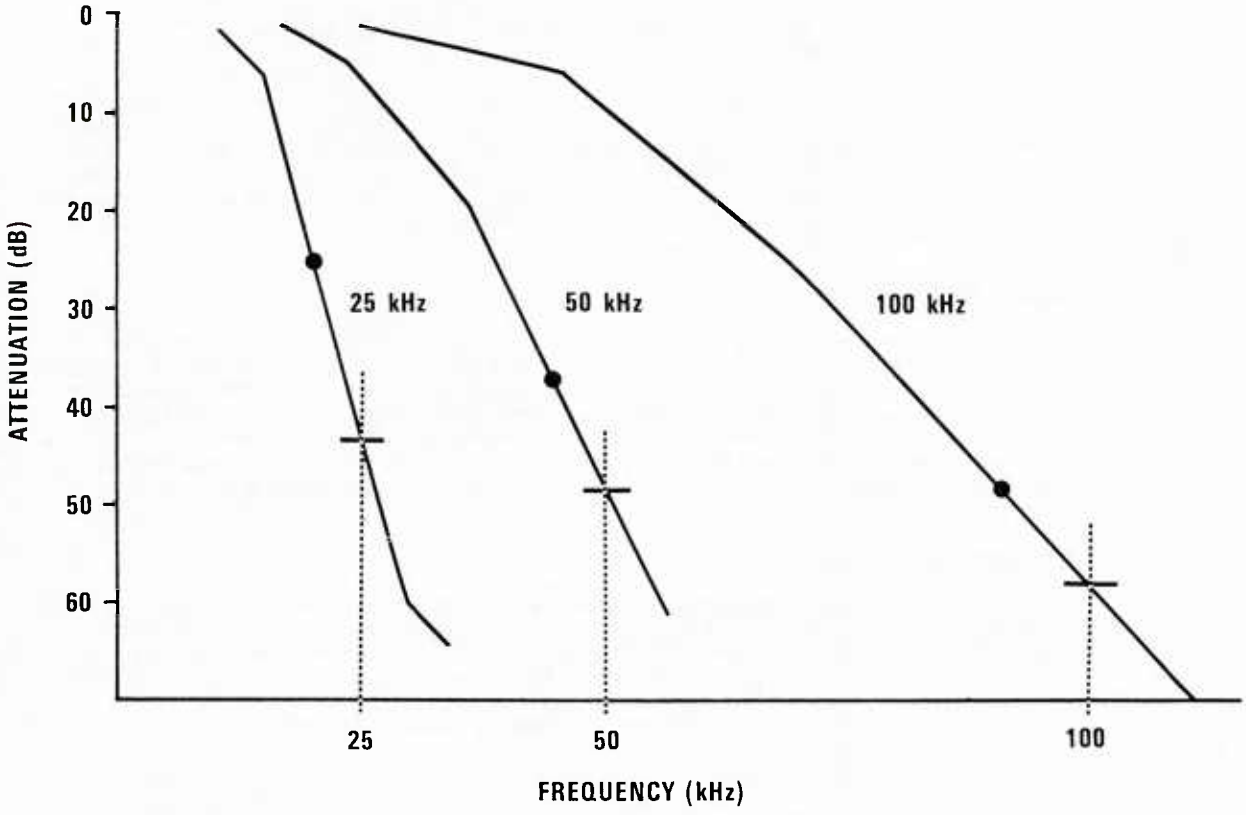


FIGURE 1 - IF. FILTER CHARACTERISTICS - A/G/A EQUIPMENTS

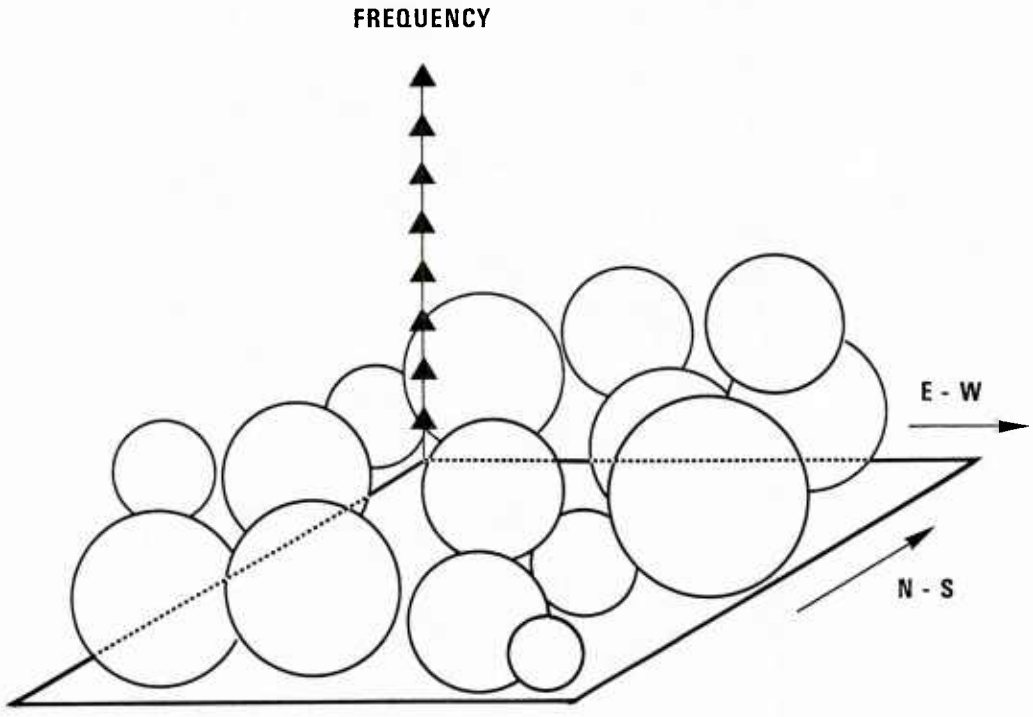
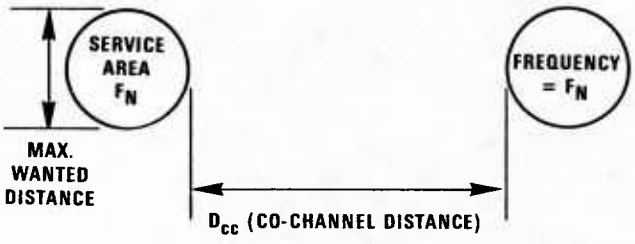
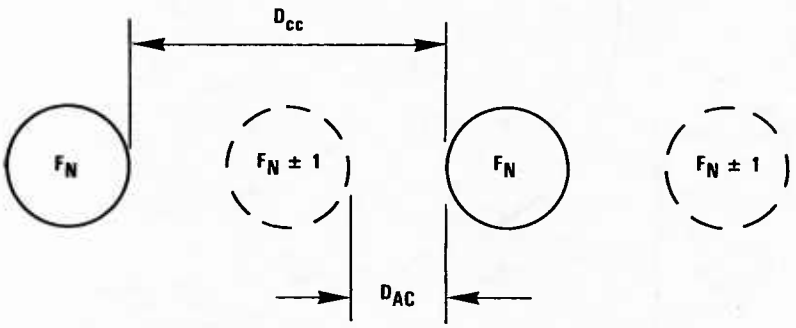


FIGURE 2 - OCCUPANCY IN THE AREA/FREQUENCY VOLUME



USE OF CO-CHANNELS



ADJACENT CHANNEL SEPARATION

$$D_{AC} = \frac{(D_{cc} - D_W)}{2}$$

INTERLEAVING ADJACENT CHANNELS

FIGURE 3 - CONDITIONS FOR ADJACENT CHANNEL USE

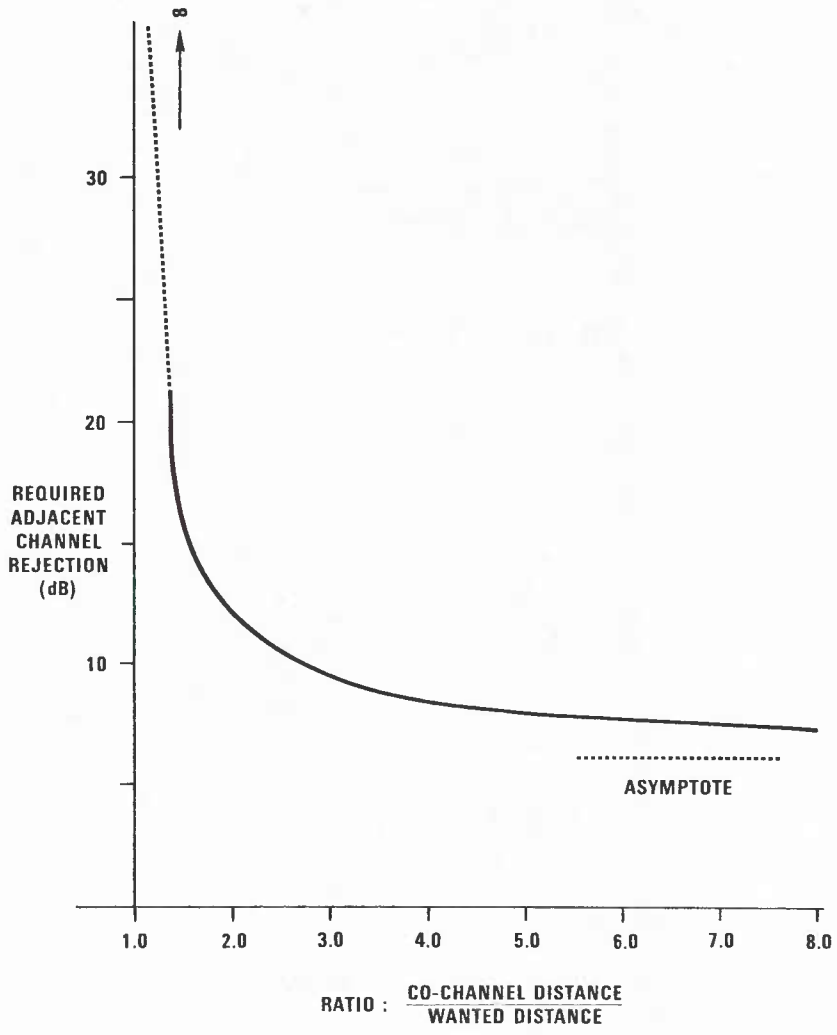


FIGURE 4 - ADJACENT CHANNEL REJECTION

AGARD Report No.744
Advisory Group for Aerospace Research and Development, NATO
SPECIAL COURSE ON INTERACTION OF PROPAGATION AND DIGITAL TRANSMISSION TECHNIQUES
Published October 1986
318 pages

This Special Course has three main objectives: 1. To provide an opportunity for engineers and scientists to learn more of propagation mechanisms and communications system design in frequency bands other than those with which they may be closely associated, thus enabling a "cross-fertilisation" of ideas to take place; 2. To attempt to identify, for the various classes of systems, the

P.T.O

AGARD-R-744

Electromagnetic wave transmission
Propagation
Data transmission
Digital techniques

AGARD Report No.744
Advisory Group for Aerospace Research and Development, NATO
SPECIAL COURSE ON INTERACTION OF PROPAGATION AND DIGITAL TRANSMISSION TECHNIQUES
Published October 1986
318 pages

This Special Course has three main objectives: 1. To provide an opportunity for engineers and scientists to learn more of propagation mechanisms and communications system design in frequency bands other than those with which they may be closely associated, thus enabling a "cross-fertilisation" of ideas to take place; 2. To attempt to identify, for the various classes of systems, the

P.T.O

AGARD-R-744

Electromagnetic wave transmission
Propagation
Data transmission
Digital techniques

AGARD Report No.744
Advisory Group for Aerospace Research and Development, NATO
SPECIAL COURSE ON INTERACTION OF PROPAGATION AND DIGITAL TRANSMISSION TECHNIQUES
Published October 1986
318 pages

This Special Course has three main objectives: 1. To provide an opportunity for engineers and scientists to learn more of propagation mechanisms and communications system design in frequency bands other than those with which they may be closely associated, thus enabling a "cross-fertilisation" of ideas to take place; 2. To attempt to identify, for the various classes of systems, the

P.T.O

AGARD-R-744

Electromagnetic wave transmission
Propagation
Data transmission
Digital techniques

AGARD Report No.744
Advisory Group for Aerospace Research and Development, NATO
SPECIAL COURSE ON INTERACTION OF PROPAGATION AND DIGITAL TRANSMISSION TECHNIQUES
Published October 1986
318 pages

This Special Course has three main objectives: 1. To provide an opportunity for engineers and scientists to learn more of propagation mechanisms and communications system design in frequency bands other than those with which they may be closely associated, thus enabling a "cross-fertilisation" of ideas to take place; 2. To attempt to identify, for the various classes of systems, the

P.T.O

AGARD-R-744

Electromagnetic wave transmission
Propagation
Data transmission
Digital techniques

propagation measurements/data still required to enable digital communication system design to be made more precise and effective; and 3. To identify trends in the requirements and design techniques for future digital communication systems.

This Special Course, sponsored by the Electromagnetic Wave Propagation Panel of AGARD, has been implemented by the Consultant and Exchange Programme of AGARD, and was presented at Jevnaker, Norway, 13—14 October 1986, at Copenhagen, Denmark, 16—17 October 1986 and in Lisbon, Portugal, 20—21 October 1986.

ISBN 92-835-1539-0

propagation measurements/data still required to enable digital communication system design to be made more precise and effective; and 3. To identify trends in the requirements and design techniques for future digital communication systems.

This Special Course, sponsored by the Electromagnetic Wave Propagation Panel of AGARD, has been implemented by the Consultant and Exchange Programme of AGARD, and was presented at Jevnaker, Norway, 13—14 October 1986, at Copenhagen, Denmark, 16—17 October 1986 and in Lisbon, Portugal, 20—21 October 1986.

ISBN 92-835-1539-0

propagation measurements/data still required to enable digital communication system design to be made more precise and effective; and 3. To identify trends in the requirements and design techniques for future digital communication systems.

This Special Course, sponsored by the Electromagnetic Wave Propagation Panel of AGARD, has been implemented by the Consultant and Exchange Programme of AGARD, and was presented at Jevnaker, Norway, 13—14 October 1986, at Copenhagen, Denmark, 16—17 October 1986 and in Lisbon, Portugal, 20—21 October 1986.

ISBN 92-835-1539-0

propagation measurements/data still required to enable digital communication system design to be made more precise and effective; and 3. To identify trends in the requirements and design techniques for future digital communication systems.

This Special Course, sponsored by the Electromagnetic Wave Propagation Panel of AGARD, has been implemented by the Consultant and Exchange Programme of AGARD, and was presented at Jevnaker, Norway, 13—14 October 1986, at Copenhagen, Denmark, 16—17 October 1986 and in Lisbon, Portugal, 20—21 October 1986.

ISBN 92-835-1539-0

AGARD

NATO OTAN

7 rue Ancelle · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone (1)47.38.57.00 · Telex 610 176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres.Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Defence Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Board
Ved Idraetsparken 4
2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92320 Châtillon

GERMANY
Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
D-7514 Eggenstein-Leopoldshafen

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens

ICELAND
Director of Aviation
c/o Flugrad
Reyjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3 Piazzale Adenauer
00144 Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
2600 AC Delft

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Portuguese National Coordinator to AGARD
Gabinete de Estudos e Programas
CLAFA
Base de Alfragide
Alfragide
2700 Amadora

TURKEY
Millî Savunma Başkanlığı (MSB)
ARGE Daire Başkanlığı (ARGE)
Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Research Center
M/S 180
Hampton, Virginia 23665

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche or Photocopy

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Branch
NASA Headquarters (NIT-40)
Washington D.C. 20546, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

ISBN 92-835-1539-0