LEVEL

AGARD-CP-263

AGARD-CP-263

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AD A080855

DDC FILE COPY

DDC
RECEIVED
FEB 20 1980
E

AGARD CONFERENCE PROCEEDINGS No. 263

Special Topics in HF Propagation

Best Available Copy

This document has been approved for public release and sale; its distribution is unlimited.

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

80 2 19 179

(2)

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANIZATION DU TRAITE DE L'ATLANTIQUE NORD)

D D C
FEB 20 1980
E

(9) AGARD Conference Proceedings No.263

(6) SPECIAL TOPICS IN HF PROPAGATION

Edited by

(10) V.J.Coyne
Strategic Surveillance Branch
Rome Air Development Center
Griffiss A.F.Base
New York 13441
USA

(11) Nov 79

(12) 594

(14) AGARD-CP-263

Copies of papers presented and discussions held at a Symposium of the Electromagnetic Wave Propagation Panel held in Lisbon, Portugal 28 May–1 June 1979.

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 North Atlantic Military Committee in the field of aerospace research and development;

- 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.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published November 1979

Copyright © AGARD 1979
All Rights Reserved

ISBN 92-835-0253-1

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

PREFACE

Utilization of the HF portion of the spectrum has long been recognized as a useful and economic method of achieving wide distribution of energy over long distances, whether it be for communications or other purposes. Broad area surveillance systems are being contemplated which must contend with problems ranging from short term predictions and real time frequency management to clutter backgrounds and noise and interference variations. Although satellite communication systems will replace HF in some areas it will still be retained by many of the world nations for point-to-point transfer of information and by most of the nations for ship and aircraft communications. These systems will depend heavily on accurate predictions and new technology to improve circuit availability and reliability. The utility of HF has also created the overall problems of interference and spectrum crowding and it is anticipated that by the early 1980's spectrum usage will be a major world concern. It is important, therefore, that as new technology evolves which will contribute to a better understanding of the transmission medium that it be readily made available to all potential beneficiaries.

The purpose of this symposium was to transfer information on the state-of-the-knowledge of HF propagation as it applies to communication, surveillance, and other systems contemplated or presently in use by NATO nations. The goal was to foster an exchange which will allow examination of requirements, an assessment of capability and, wherever possible, a definition of areas where additional work is required.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or special
A

PROGRAM COMMITTEE AND MEETING OFFICIALS

PROGRAM CHAIRMEN

Mr V.J.Coyne
Strategic Surveillance Branch
Rome Air Development Center
Griffiss A.F. Base
New York 13441
USA

Dr G.Lange-Hesse
Max-Planck Institut fur Aeronomie
Postfach 20
D-3411 Katlenburg-Lindau 3
Germany

MEMBERS

Capitaine de Frégate P.Halley
Ingénieur en Chef au CNET
38 -40 rue du Général Leclerc
92131 Issy-les-Moulineaux
France

Dr T.Damboldt
FI 33b
Forschungsinstitut der DBP
beim FTZ
Postfach 5000
D-6100 Darmstadt
Germany

Professor M.Cutolo
Universita di Napoli
Istituto di Fisica
Via Monteoliveto 3
80134 Napoli
Italy

Dr T.B.Jones
Physics Department
University of Leicester
Leicester LE1 7RH
UK

Dr C.Rush
NITA/ITS
Boulder
Colorado 80302
USA

Dr E.V.Thrane
NDRE
PO Box 25
N-2007 Kjeller
Norway

PANEL OFFICERS

CHAIRMAN

Dr H.J.Albrecht
Germany

DEPUTY CHAIRMAN

Dr J.Aarons
USA

PANEL EXECUTIVE

Lt. Colonel J.B.Catiller, USAF
AGARD

HOST COORDINATOR

Dr A.S Mendes
Director do Servico
Meteorologica Nacional
R.Saraiva de Carvalho 2
Lisboa 3
Portugal

EDITOR'S COMMENTS

The primary purpose of the symposium was to exchange information on the state of the knowledge of HF propagation. That purpose was successfully accomplished by the efforts of the program committee and session chairmen who selected the papers, directed their presentation and monitored the discussions.

A secondary purpose of the symposium was to identify areas where additional effort is required. That purpose was only partially satisfied, however, it should be noted that a working group on HF communications has been formed which will survey the current use of HF in NATO nations and will determine future HF communications requirements for NATO in both a primary and backup mode. Dr Jules Aarons, deputy chairman of EPP, will chair the study group which will operate for the next 30 months. It is hoped that the symposium and proceedings will be useful to their effort.

These proceedings represent a compilation of the papers presented at the symposium as well as a significant amount of the discussions following each presentation. The papers are reproduced from copies furnished by the authors. The discussions published at the end of each paper are not verbatim as spoken but rather more concise written records provided by the individuals involved. In some instances it was necessary for the editor to extract and interpret comments from the taped transcript. I hope that I have been accurate and apologize for changes in meaning that may have resulted.

The session summaries presented herein were mostly provided by the session chairmen listed on a following page and I thank them for their efforts.

I wish to acknowledge the support provided by the EPP Executive, Lt Colonel John Catiller and his secretary, Madame Tessier, in the preparation of these proceedings as well as the support provided by Mrs P.Van Dresar, RADC.

Grateful acknowledgment is extended to Dr A.S.Mendes of The Instituto Nacional de Meteorologica F Geophysica, Lisboa for his efforts toward the successful accomplishment of the symposium.

VINCENT J.COYNE

SESSION CHAIRMEN

SESSION I	OPERATIONAL CONSIDERATIONS	
	Dr B.Burgess	RAF Farnborough Hants, UK
	Dr T.B.Jones	University of Leicester Physics Department Leicester, LE1 7RH UK
SESSION II	MODELING OF IONOSPHERIC PARAMETERS	
	Dr C.Rush	NITA/ITS Boulder, Colorado, 80302 USA
SESSION III	HIGH LATITUDE IONOSPHERE EFFECTS	
	Dr G.Lange-Hesse	Max-Planck Institüt fur Aeronomie Postfach 20 D-3411, Katlenburg-Lindau 3 Germany
SESSION IV	PROPAGATION THROUGH IRREGULARITIES	
	Dr J.Aarons	AFGL/PHP Hanscom AFB, MA 01731 USA
SESSION V	REMOTE SENSING AND OCEAN SURVEILLANCE	
	Prof P.Broche	LSEET-Université de Toulon Toulon, France
SESSION VI	TUNNEL TRANSMISSION AND SURFACE WAVES	
	Mr J.R.Wait	University of Colorado Boulder, CO 80309 USA
SESSION VII	EQUIPMENT CONSIDERATIONS	
	Dr T.Damboldt	Fi 33b Forschungsinstitut der DBP beim FTZ Postfach 5000, D-6100 Darmstadt Germany

CONTENTS

Page

PREFACE iii

PROGRAM COMMITTEE AND MEETING OFFICIALS iv

EDITOR'S COMMENTS v

SESSION CHAIRMEN vi

Reference

SESSION I OPERATIONAL CONSIDERATIONS

SESSION SUMMARY 1

DES MODELES THEORIQUES DE LA COMPOSITION NEUTRALISEE ET IONISEE DE LA MESOSPHERE ET DE LA BASSE THERMOSPHERE*
par L A.Mata 1

THE ROLE OF HF IN AIR-GROUND COMMUNICATIONS AN OVERVIEW
by B.Burgess 2

HF COMMUNICATIONS TO SMALL LOW FLYING AIRCRAFT
by N.M.Maslin 3

MODERN HF COMMUNICATIONS FOR LOW FLYING AIRCRAFT
by J.F.Brune and B.V.Ricciardi 4

REAL-TIME ADAPTIVE HF FREQUENCY MANAGEMENT
by R.B.Fenwick and T.J.Woodhouse 5

ASSESSMENT OF HF COMMUNICATIONS RELIABILITY
by N.M.Maslin 6

COMPARISON OF MEASURED AND PREDICTED MUF'S AT A REMOTE LOCATION
by T Damboldt 7

AUGMENTATION OF HF PROPAGATION
by T.J.Barrett 8

TROPOSPHERIC EFFECTS ON HF PROPAGATION
by G.H.Millman 9

SESSION II – MODELING OF IONOSPHERIC PARAMETERS

SESSION SUMMARY II

TOWARD GLOBAL MONITORING OF THE IONOSPHERE IN REAL-TIME BY A BOTTOMSIDE SOUNDING NETWORK: THE GEOPHYSICAL REQUIREMENTS AND THE TECHNOLOGICAL OPPORTUNITY
by J.W.Wright and A.K.Paul 10

HYBRID RAY-MODE FORMULATION OF TROPOSPHERIC PROPAGATION
by S.H.Cho, C.G.Migliora and L.B.Felsen 11

INFLUENCE DES MODELES D'IONOSPHERE SUR LES CALCULS DE PROPAGATION DES ONDES DECAMETRIQUES
par R.Fleury et P.Gourvez 12

METHODS OF DETERMINING IONOSPHERIC STRUCTURE FROM OBLIQUE SOUNDING DATA
by R.E.DuBroff, N.Narayana Rao and K.C.Yeh 13

---

* Not available at time of printing.

Reference

THE GEOMORPHOLOGY OF THE HF BREAKTHROUGH PHENOMENON
by J.M.Goodman 14

COUPLING BETWEEN THE NEUTRAL AND IONIZED UPPER ATMOSPHERE DURING DISTURBED CONDITIONS
by G.W.Prölss 15

BASIC FINDINGS HELPFUL FOR IONOSPHERIC PREDICTIONS
by E.Harnischmacher and K.Rawer 16

SUR LA DETERMINATION DE LA MUF CLASSIQUE
par R.Hanbaba 17

SESSION III – HIGH LATITUDE IONOSPHERE EFFECTS

SESSION SUMMARY III

CHARACTERISTICS OF THE HIGH LATITUDE IONOSPHERE PRODUCED BY AURORAL PARTICLE PRECIPITATION
by J.A.Whalen 18

PERSPECTIVE ON THE PREDICTION OF AURORAL ABSORPTION
by V.Agy 19

DIRECTION AND DOPPLER CHARACTERISTICS OF MEDIUM AND LONG PATH HF SIGNALS WITHIN THE NIGHT-TIME SUB-AURORAL REGION
by R.W.Jenkins, E.L.Hagg and L.E.Montbriand 20

RECENT ADVANCES IN HF PROPAGATION SIMULATION
by T.J.Elkins 21

SESSION IV – PROPAGATION THROUGH IRREGULARITIES

SESSION SUMMARY IV

TRANSEQUATORIAL PROPAGATION THROUGH EQUATORIAL PLASMA BUBBLES DISCRETE EVENTS
by M.L.Héron 22

THE INFLUENCE OF E-REGION AND F-REGION IRREGULARITIES ON HF PROPAGATION IN THE EQUATORIAL ZONE
by J.Röttger 23

THE PHENOMENOLOGY OF TRANSEQUATORIAL RADIO PROPAGATION UNDER SPREAD-F CONDITIONS
by J.Röttger 24

DETECTION, RANGING AND DRIFTSPEED MEASUREMENTS OF EQUATORIAL IONOSPHERIC IRREGULARITIES BY MEANS OF AIRGLOW OBSERVATIONS
by G.Lange-Hesse and H.Lauche 25

HF WAVEFRONT IRREGULARITIES OBSERVED ON A LARGE APERTURE RECEIVING ARRAY
by T.B.Jones and E.C.Thomas 26

A SPORADIC E PREDICTION TECHNIQUE
by M.A.Merkel and R.Rubio 27

SCATTER INJECTION/DUCTED MODE HF RADAR
by G.S.Sales 28

EFFETS IONOSPHERIQUES D'UNE ECLIPSE SOLAIRE DANS LES ILES DU CAP VERT
par A.S.Mendès, F.Bertin, P.Vila et J.Papet-Lépine 29

Reference

GROUND-WAVE AND SKY-WAVE SEA-STATE SENSING EXPERIMENTS IN THE UK
by E.D.R.Shearman, W.A.Sandham, E.N.Bramley and P.A.Bradley 30

SESSION V – REMOTE SENSING AND OCEAN SURVEILLANCE

SESSION SUMMARY V

ESTIMATION DU SPECTRE DIRECTIONNEL DES VAGUES PAR RADAR DECAMETRIQUE COHERENT
par P.Broche 31

DEVELOPMENT OF HF SKYWAVE RADAR FOR REMOTE SENSING APPLICATIONS
by T.W.Washburn, L.E.Sweeney Jr, J.R.Barnum and W.B.Zavoli 32

HF SKYWAVE RADAR ESTIMATES OF THE TRACK, SURFACE WIND AND WAVES OF HURRICANE ANITA
by J.W.Maresca Jr and C.T.Carlson 33

OCEAN SWELL PARAMETERS FROM NARROW-BEAM HF RADAR SEA ECHO
by B.Lipa and D.Barrick 34

SESSION VI – TUNNEL TRANSMISSION AND SURFACE WAVES

SESSION SUMMARY VI

PRINCIPLES OF HF COMMUNICATION IN TUNNELS USING OPEN TRANSMISSION LINES AND LEAKY CABLES
by J.R.Wait, D.A.Hill and D.B.Seidel 35

MODE CONVERTERS FOR HF TUNNELS TRANSMISSION
by P.Delogne 36

LEAKY COAXIAL CABLES FOR OBSTACLE DETECTION AND CONTINUOUS ACCESS GUIDED COMMUNICATIONS
by N.A.M.Mackay, J.C.Beal, D.J.Gale and J.L.Mason 37

DEFINITION OF SUBSURFACE FEATURES BY GEOPHYSICAL PROBING
by R.J.Lytle 38

RESULTATS EXPERIMENTAUX CONCERNANT LA PROPAGATION LIBRE DES ONDES UHF DANS LES TUNNELS
par P.Degauque, P.Mangez, B.Demoulin et R.Gabillard 39

EXCITATION OF THE HF SURFACE WAVE BY VERTICAL AND HORIZONTAL APERTURES
by J.R.Wait and D.A.Hill 40

EFFECTIVE USE OF NATURAL MODES IN VHF AND UHF TUNNEL PROPAGATION
by L.Deryck 41

COMPARISON OF LOOP AND DIPOLE ANTENNAS IN LEAKY FEEDER COMMUNICATION SYSTEMS
by D.A.Hill and J.R.Wait 42

MODE CONVERSION BY TUNNEL NON-UNIFORMITIES IN LEAKY FEEDER COMMUNICATION SYSTEMS
by D.B.Seidel and J.R.Wait 43

SESSION VII – EQUIPMENT CONSIDERATIONS

SESSION SUMMARY VII

Reference

A MOBILE HF IMPULSE SOURCE LOCATOR
by M.L.Héron 44

A NEW COMPUTER-CONTROLLED HIGH FREQUENCY DIRECTION-FINDING AND TRANSMITTER LOCATING SYSTEM
by R L.Tanner 45

DIGITAL ON-LINE PROCESSING AND DISPLAY OF MULTIPARAMETER HF TRANSMISSION DATA
by K.Bibl and B.W Reinisch 46

NEW TECHNOLOGY TO IMPROVE HF CIRCUIT RELIABILITY AND AVAILABILITY FOR REMOTE REGIONS
by J S.Belrose and L.R.Bode 47

ELECTRICALLY SHORT HF AERIAL SYSTEMS
by M.N.Sweeting and Q.V.Davis 48

RADIO-LINK COMPUTATIONS OPTIMIZE PATTERN SHAPING OF SHORTWAVE ANTENNAS
by A.Stark 49

APPENDIX A – LIST OF PARTICIPANTS A

SUMMARY OF SESSION I

OPERATIONAL CONSIDERATIONS

The session opened with a review of airborne communications by **Dr B. Burgess**, (UK). Following a historical introduction the current techniques employed were discussed in some detail. Attention was given to factors which affect the system performance including the ionospheric limitations. Suggestions were then presented for improving the performance Finally, future trends in the general field of airborne communications were reviewed and the advantages of techniques such as digital communications, channel evaluations etc, emphasised.

The second paper, by **Dr N. Maslin**, was presented by **Mr A. May**, (UK) and was concerned with the assessment of HF communications reliability. It was shown that circuit reliability depends on two important parameters (1) the ratio of the operating frequency to the MUF and (2) the ratio of the received to required signal-to-noise ratio It was demonstrated that when these parameters are optimized by correct frequency management appreciable improvement in system performance can be achieved.

The next paper, by **Mr Brune** and **Ricciardi** (US), was presented by **Mr Ricciardi** and dealt with the problems of communications with low flying aircraft, and with helicopters in particular The talk emphasised the difficult operational conditions which are experienced in this type of communications activity. The results of field trials were considered, problem areas identified and solutions postulated.

In **Dr Fenwick**'s paper on 'Real-Time Adaptive HF Frequency Management' the problem of optimum utilization of the HF frequency band was addressed. By use of a Chirp sounder and a spectrum-monitoring system it was experimentally shown that real-time information of propagation and channel occupancy allowed sharing of the HF spectrum with an acceptable minimum of interference. This would be of particular interest to the tactical communicator with a high-priority requirement

Dr Maslin's paper on 'HF Communications With Small Low-Flying Aircraft' gave an application of his earlier paper on HF communications reliability to this short-sample communications requirement. He showed that this type of communications gives very poor reliability if direct communications is required and arises from both propagation limitations and aircraft antenna performance. The use of long range propagation to distant HF relay stations is recommended and shows major increases in reliability of the HF link and ease in frequency management problems.

Dr Damboldt drew attention to the day-to-day variability of the ionosphere which has an appreciable influence on the performance of HF communications systems. By noting the time at which signals from a distant transmitter are first received, a measure of the MUF can be obtained. These 'transition times' allow a measure of ionospheric variability on a day to day basis.

An interesting paper on 'The Augumentation of HF Propagation' was given by **Mr Barrett**. Here emphasis was placed on the use of chemical releases in the ionosphere to create an enhanced ionized reflection region in order to allow communication at certain critical times, i.e., at times of high-altitude nuclear explosions. Evidence was given to show that communications over wide geographical areas could be achieved by the appropriate chemical releases at a small number of locations in say the NATO area.

The session closed with a paper by, **Dr Millman**, which drew the conferences attention to the influence that the troposphere can play in extending somewhat the range of HF propagation due to the refractive nature of the lower atmosphere. Applications to radar and communications system were mentioned and showed that to a first-order, tropospheric effects are negligible.

THE ROLE OF HF IN AIR-GROUND COMMUNICATIONS

AN OVERVIEW

Dr. B. Burgess
Royal Aircraft Establishment
FARNBOROUGH, Hants
UK

SUMMARY

HF radio is a prime means for beyond line-of-sight air-ground communications and will remain as such into the foreseeable future, complementing Satellite communications where this latter system is employed.

This paper will review various aspects that make up an HF communications system and the contribution that they make to the overall system performance. Consideration will be given to the needs of the modern user in the use of HF communications, emphasising the changing requirements on the HF link that for example, the use of digital communications brings. Such aspects as error rate performance, avoidance of interference, HF prediction techniques for Northern latitudes, role of channel evaluation techniques etc will be addressed.

The aim of the paper is to acquaint the research worker with the status and trends in airborne HF communications and to indicate where further work may profitably be undertaken to eventually improve sytem performance.

1. INTRODUCTION

Airborne communications goes back to almost the start of flying. Wireless had been tried out in airships and aircraft for some time before the First World War; in the UK the Royal Engineers and Marconi had carried out experiments on what is now part of the RAE airfield at Farnborough. During that war wireless was used more extensively in aircraft than in the forward areas of the Western Front. These wireless sets operated on frequencies in the MF band and used long wire antennas that were trailed from the aircraft (Ref 1). Whilst not strictly HF, this form of communications underlined the importance that the wireless/radio has played in flying from its inception. Prior to World War II, communications with aircraft was almost entirely by HF, and commonly employed hand operated morse. However it is of interest to note that as early as 1928 the first trials of voice air-ground communications were being made in the USA and in that country there was no transition from morse to voice in airborne communications (Ref 2).

It was World War II that saw the development and introduction of VHF for airborne communications (Ref 19). This was employed for the control of fighter aircraft during the Battle of Britain and the end of the war saw VHF in general use in the allied air fleets with longer term plans to go to the 225-400 MHz band. As of today the VHF band, 118-136 MHz, is used for Civil Air Traffic Control communications and the 225-400 MHz UHF band for military air-ground communications. These frequency bands are of course constrained by line-of-sight (los) propagation and any communication link to the aircraft at ranges beyond line-of-sight (blos) required the use of HF frequencies.

It is interesting to note that just after the end of World War II, before VHF Air Traffic Control communications were fully implemented, air to ground communications were conducted using frequencies between 3 and 6 MHz and ground to air communications at about 300 kHz. Using these frequencies comments such as "Why can we hear the broadcast bands loud and clear but ATC communications is terrible?" were common, and where air to ground communications ranges of about 50 km were desired, communications were often found to be very poor, whereas the ground station could receive distance stations (>1000 km) loud and clear. These problems, typical of airborne HF communications, will be returned to later in the paper.

In the early 1960s with the emergence of satellite communications, the possibility of high quality communications for blos ranges was with us. While air-ground communications via satellites has been demonstrated (Ref 3), with superior quality to HF, it is still not, in the late 1970s, evident in aircraft use except in a few special applications. It is noteworthy that the AEROSAT programme which was launched in the early 1970s for Air Traffic Control use over the North Atlantic has now been abandoned. One of the main reasons for this state is the cost of satellite communications, including its airborne installation. Thus, until the day arrives in the near future when satellite communications for aircraft is viable, HF will remain the primary means of blos communications, due mainly to its lower cost and free provision by nature of a reflecting region in the upper atmosphere, albeit an imperfect one.

Since HF communications for aircraft will remain with us into the foreseeable future, and more demanding requirements will be sought by the users, it is appropriate to review the use of HF in current airborne systems and its role in future systems; also what the problem areas are and what measures can be taken to improve the current systems.

2. CURRENT USES AND PERFORMANCE

In considering the use and performances of HF communication systems for aeronautical purposes it is convenient to categorize the usage into en-route and off-route applications. An en-route usage can be exemplified by civil aircraft flying the North Atlantic route between Europe and North America. The routes are laid down geographically and most flights are scheduled. The aircraft need to communicate over

Copyright © Controller HMSO, London 1979

long ranges by HF and this is undertaken in order to give positional information at certain times, or times of crossing certain longitudes, and also such information as estimated time-of-arrival, fuel state, airfield diversion etc. These messages from the aircraft are relatively short and usually require a short acknowledgement from the ground. If the information does not get through first time but takes say ten minutes or so, this is usually acceptable.

The Communications frequencies to be used, both primary and secondary, are provided at crew briefing and because the tracks flown are well charted, past experience helps a great deal in achieving reliable communications. Overall such communications is relatively satisfactory, but of course problems occur during ionospheric disturbances. For long transport routes in temperate and lower latitudes, high availability and reliability is usally achieved.

On the other hand when one considers off-route usage, a different picture can emerge. Take a military aircraft flying on a route that is not regularly used; it may be a large aircraft with a crew which may or may not include a specialist radio operator, or a high performance aircraft with only a one or two man crew who use the radio as a tool. In these cases it could be that the aircraft does not wish to transmit unless it is essential, but that when it does, the response of the communication system must be highly reliable and quick. The choice of the correct frequency is in this case important. As an example take an aircraft operating to the North of the UK in summer 1977 and wishing to communicate with a base in Southern England. Figure 1 indicates the frequency that should be used depending on time of day and distance from the base. One could easily require up to 8 frequencies for a typical sortie. As Wright (Ref 4) indicates communications in this environment has proved to be difficult and improvements are needed.

The above remarks have implied that speech is the normal mode of communications. This is the case in Civil aircraft where the use of the radio operator aboard airliners has been eliminated. Military aircraft however do operate other modes besides speech. Morse is still a fall back position where a crew member has this skill; also radio teletypes (RATT) operating at low speed data rates are in use in some Air Forces where security is a requirement.

Future requirements where security of speech transmissions and an ECCM capability are deemed desirable indicate a trend towards higher speed digital transmissions (up to 2.4 kb/s) and the use of advanced modulation and coding techniques in HF communications (Ref 5).

3. FACTORS GOVERNING PERFORMANCE

In order to gain an insight into the performance of aircraft communications systems it is necessary to look in some detail at such an air-ground link and highlight areas where there are deficiencies in the system and possibly indicate improvements. Consider Figure 2 which gives in block form the components of a communications system. At the outset it is important to appreciate the difference between a mobile air-ground circuit and a point-to-point or broadcast circuit. On a point-to-point circuit advantage can be designed to give azimuthal and some elevation gain or directivity. This is not possible in the military aeromobile case where a low gain antenna is usually employed in order to cover the wide geographic area where the aircraft is likely to fly. Currently adaptive antenna arrays have not been used to select the optimum beam for communications with the mobile. In the case of transmissions from the aircraft, no gain is possible from the aircraft antenna; in fact the efficiency of the aircraft antenna can be very low (vide Reference 6). On antenna gain alone some 20 dB or more is lost to the mobile compared with the fixed user.

Turning to the transmitter, fixed and broadcasting services use powers often in excess of 10 kW and usually in the hundreds of kW range. Aircraft power and weight restrictions limit the power output of airborne HF transceivers to a maximum of 1 kW, thus giving a deficit of the order of 20 dB on transmitter power. These factors alone give some indication of why the early users queried why their communications were so poor compared with the broadcast bands.

Haslin (Ref 7) has considered these factors in detail in assessing HF blos airborne communications reliability and the reader is referred to this paper for a consideration of the relative importance of the system parameters.

Two environmental factors that are significantly different for the aeromobile user and have to be taken into account under certain conditions are (1) the aircraft cockpit noise level and (2) the level and nature of interference in the aeromobile bands. For high performance aircraft the acoustic noise that is generated in the cockpit is extremely high and even with flying helmets the noise occlusion is not sufficient to enable a good dynamic range for the audio output to give high quality speech intelligibility.

Regarding the radio interference in the aeromobile bands it has been found (Ref 8) that this is very often of a narrow band nature, indicating that, for example, an FSK radio teletype signal, which uses two tones in the 3 kHz bandwidth, can be strongly interfered with, but if these tones were moved to another spot in the 3 kHz audio band, no interference would be observed. This fact gives a possibility of improvements to this type of transmission.

If we look to the possible use of 2.4 kb/s transmissions of data over HF links, the ionosphere has proved to be a limiting factor in the design of an efficient modem (modulation/demodulator). Due mainly to multipath effects a serial bit rate of the order of 100 b/s is an upper limit to avoid excessive error due to time dispersion. Thus in a 3 kHz bandwidth multiple tones have to be used in order to obtain bit rates of the order of 2400. Frequency dispersion in the ionosphere also puts a bound on the minimum spacing of the tones. Vincent et al (Ref 9) show that under these time and frequency dispersion conditions, increasing the signal/noise ratio while improving the error rate performance up to a certain level, does not give improvements at higher levels, the dispersion effect often limiting the error rate to relatively high values.

Over the past 20 years a number of modem designs have been tested to see if they will successfully combat ionospheric dispersion and variability. Of attempts, such as Kathryn, Andeft, Adapticom and Kineplex (Ref 10), only Kineplex has entered service on point-to-point circuits. Recent approaches using coding (Refs 11 and 12) as well as complex modulation schemes look promising, but experimental validation of these techniques coupled with an economic solution to the cost of these complex modems will be needed before they are likely to be used.

If such techniques as those above are used to try and combat ionospheric effects care has to be taken in the airborne environment to also design the modems to take account of the doppler effect introduced by the motion of the aircraft. For small high performance aircraft, doppler shifts of the order of tens of Hertz can easily be achieved and the tone filters used in multitone modems have to either accomodate this shift of have compensating circuits designed in order not to suffer loss of performance.

4. IMPROVEMENTS IN HF SYSTEMS PERFORMANCE

HF Communications with aircraft flying on off-route schedules is difficult for a number of reasons as outlined above. This is the case even though it has been assumed that prediction techniques have been used to select the optimum frequency of transmission and also that sufficient equipment is available to utilise the range of frequencies that are predicted for optimum use. A case can be made out for improving prediction techniques by limiting the geographic extent of the mapping of ionospheric parameters that are used in these prediction programmes. For example higher accuracy could accrue for operations in NW Europe by limiting the mapping to that area instead of using world-wide maps. In the context of Northern Europe prediction techniques have been lacking in covering northern latitudes where auroral activity is evident. Recent work by Bradley and Fapiano (Ref 13) have addressed this problem but more work is needed to achieve better prediction methods on a world-wide basis (Ref 14). What is recommended here is a family of procedures that would cater for needs between the highest accuracy using large computers and those only employing hand held calculators with a consequent reduction in accuracy.

For shorter ranges of communication (50-500 km) in the case of small high performance aircraft, where the aircraft antenna efficiency plays a major role in the system performance, Maslin (Ref 15) has shown that good frequency management, coupled with the use of (1) directive antennas on the ground and (2) a number of geographically separated remote receiving stations, are vital to the provision of satisfactory communications.

Work in the UK over the past several years (Ref 3) has shown that a dominant cause for degradation to HF communications in the aeromobile band has been narrow band interference from other legitmate users of the allocated frequencies. This is particularly the case for two tone FSK low speed data transmissions (75 b/s). Interference statistics show that conditions at either end of the link are uncorrelated and hence if one is to choose the best frequency from an interference point of view then information on the interference characteristics at the receiving end of the link needs to be provided to the transmitting terminal. This can be achieved by monitoring the interference statistics at the reception point on the frequencies allocated for communications and transmitting this on a broadcast basis to all users. The user can then choose the propagating frequencies that indicates the best signal to interference level that would be received at the reception point. Sloggett (Ref 16) has shown that the error rate performance of a 75 b/s transmission can be very significantly improved by (1) choosing the tone pair in the 3 kHz baseband giving the least interference; (2) using directional antennas and (3) employing error detection and correction techniques. The percentage of error free messages on a 2000 km link in Northern Europe was improved from 20-40% to 80-95% using these techniques. A real time channel estimation (RTCE) technique such as described above shows promise of optimising the performance of HF circuits for mobile applications and ensuring that their reliability is acceptably high. RTCE system of the type described above have also been proposed by Darnell (Ref 17) and Canadian agencies (Ref 18).

As mentioned earlier in the paper the aircraft antenna can be a source of major system degradation. Particular attention must be paid to its design to ensure that an optimum performance is achieved. For small aircraft this may well be the limiting factor in a system design and could well preclude the use of HF for certain operational uses, eg in helicopters.

5. FUTURE TRENDS

In both Civil and Military applications there has been a move to pass information in digital form. Sensors in aircraft have been evolving so that their output is in digital form and information passed to ground terminals is, these days, most conveniently handled in digital formats. Information sent through HF communications links will be no exception. Concepts of data links for passing information from civil airlines for Air Traffic Control purposes have been discussed over the past 10 years or so. In military circles secure voice requirements indicates digital radio transmissions. At HF this implies data rates of 2.4 kb/s. Higher bit rates would give better speech quality from the speech synthesis aspect, but channel bandwidth considerations indicate that around 2.4 kb/s data rates are the highest that can be tolerated in 3 kHz channel. As discussed earlier the design of 2.4 kb/s modems to achieve satisfactory transmissions of the bit rate over HF channels has not yet been solved satisfactorily and the airborne environment will put extra constraints on the modems performance.

Ince (Ref 5) in his paper has indicated that the military will pay increasing attention to ECCM aspects of communications. HF will need to be protected and complex modulation schemes involving frequency hopping and spread spectrum techniques will need to be addressed. Consideration of the propagation medium to determine whether these forms of wide bandwidth modulation techniques are transmitted with fidelity will have to be studied. For example is there a limit to the bandwidth that can be used and does the medium support one type of modulation better than any other? These techniques will tend to be implemented digitally and hence time and frequency dispersion effects on error rates become of prime importance.

A question that is raised when considering predictions is: can one predict the error rate performance of systems under different ionospheric conditions? Is this the right question to ask or should one rely on simulating the medium and the system to measure its performance?

REFERENCES

1. DICKSON, D. A., 1979, "Wireless in World War I Aircraft", Journal of the Royal Signals Institution, XIV, 42.

2. WHITE, F. C., 1973, "Air-Ground Communications: History and Expectations", IEEE Trans Comm 21, 398-407.

3. BARNES, G. W., HIRST, D., JAMES, D. J., 1971, "Chirp Modulation Systems in Aeronautical Satellites", AGARD Conference Proceedings No 87, Paper 30.

4. WRIGHT, P. J., 1968, "RAF Coastal Command Maritime Air-Ground Communications to and from Auroral Region", Ionospheric Radio Communications (Plenum Press), 260-262.

5. INCE, A. N., 1978, "FM and NATO Communications", Signal, p66-73.

6. PAVEY, N. A. D., 1973, "Radiation Characteristics of HF Notch Aerials Installed in Small Aircraft", AGARD Conference Proceedings No 139.

7. MASLIN, N. M., 1979, "Assessment of HF Communications Reliability", AGARD Conference Proceedings. (To be published).

8. GOTT, G. F., STANIFORTH, M. J. D., 1978, "Characteristics of interfering Signals in Aeronautical HF Voice Channels", Proc IEE 125, 1208-1212.

9. VINCENT, W. R., DALY, R. F., SIFFORD, B. M., 1968, "Modelling Communications Systems", Ionospheric Radio Communications, 321-340 (Plenum Press).

10. GOLDBERG, B., 1966, "300 kHz-30 MHz MF/HF", IEEE Trans Comm Tech 14, 767-784.

11. CHASE, D., 1973 "A Combined Coding and Modulation Approach for Communications over Dispersive Channels", IEEE Trans Comms, 21, 159-174.

12. BELLO, P. A., 1978, "A Review of Signal Processing for Scatter Communications", AGARD Conference Proceedings No 244, Paper 27.

13. CCIR, 1978, "Second CCIR Computer Based Interim Method for Estimating Sky-wave Field Strength and Transmission Loss at Frequencies between 2 and 30 MHz", Draft Supplement to Report 252/2, Submission to XIV Plenary Assembly, Kyoto.

14. BRADLEY, P. A., 1978, "Development in Techniques for Predicting HF Sky-wave Field Strengths", AGARD Conference Proceedings No 238, Paper 10.

15. MASLIN, N. M., 1979 "HF Communications to Small Low Flying Aircraft", AGARD Conference Proceedings. (To be published).

16. SLOGGETT, D. R., 1979 "Improving the Reliability of HF Data Transmissions", IEE Colloquium Digest 1979/48 (Recent Advances in HF Communications Systems and Techniques) pp 74-78.

17. DARNELL, M., 1975, "Channel Estimation Techniques for HF Communications", AGARD Conference Proceedings No 173, Paper 16.

18. STEVENS, E. E., 1968, "The CHEC Sounding System", Ionospheric Radio Communications (Plenum Press) 359-369.

19. GATES, B. G., 1947, "Aeronautical Communications", Journal IEE 94, Pt III A, No 11, p78-80.

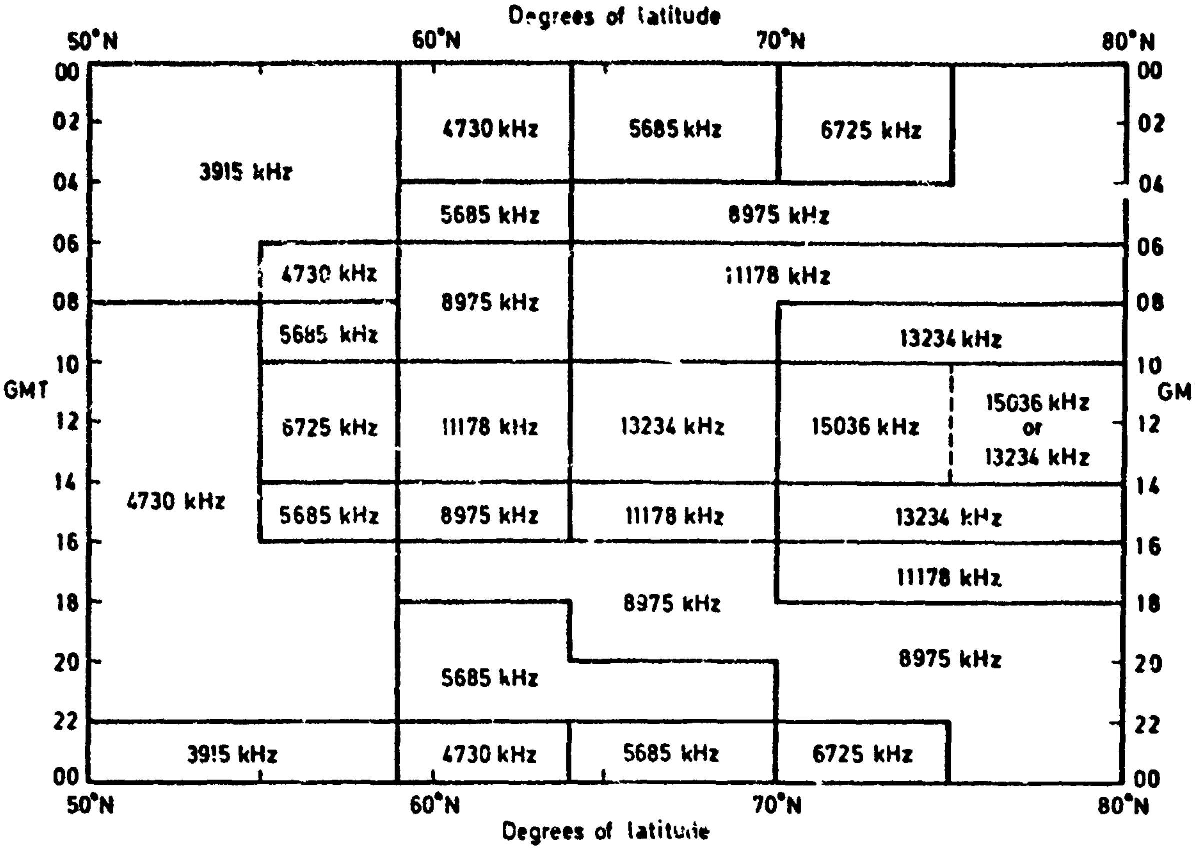


Fig.1 HF frequencies for communication with Cove experimental radio station during May, June, July, August 1977 over paths 57°N to 80°N

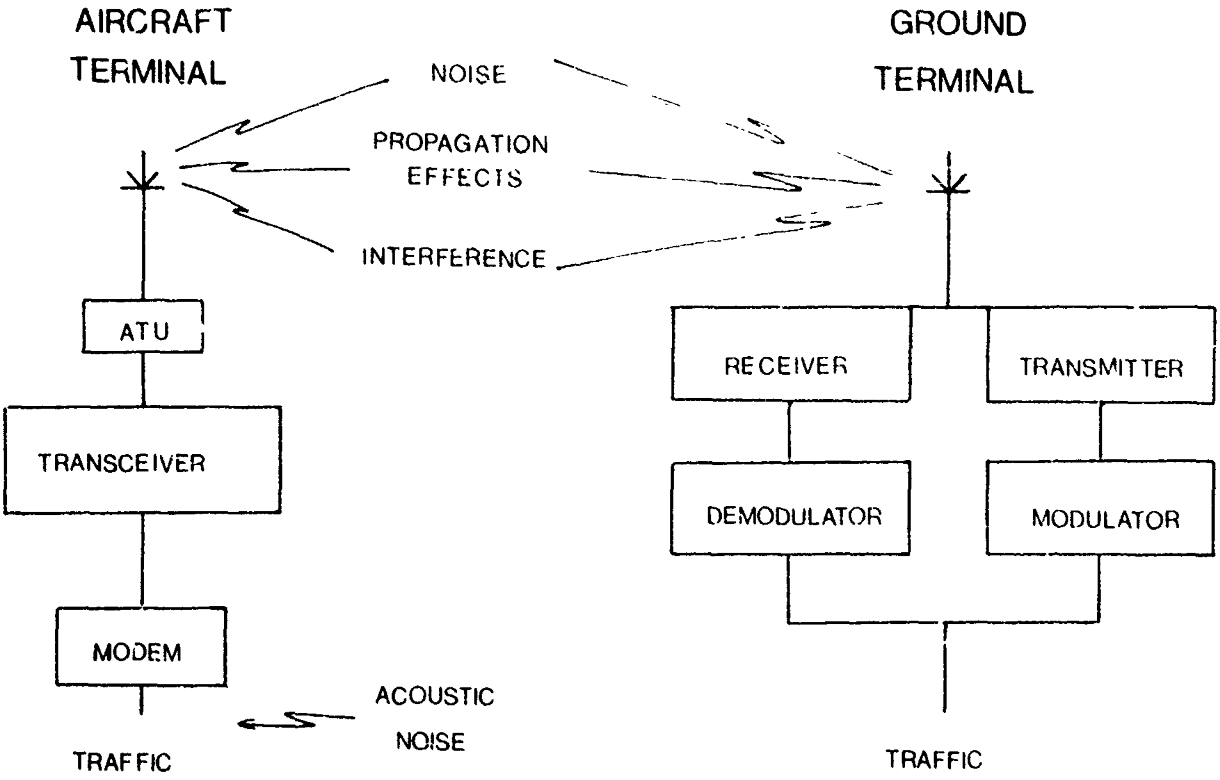

AIRCRAFT
TERMINAL
GROUND
TERMINAL
NOISE
PROPAGATION
EFFECTS
INTERFERENCE
ATU
TRANSCEIVER
MODEM
RECEIVER
TRANSMITTER
DEMODULATOR
MODULATOR
ACOUSTIC
NOISE
TRAFFIC
TRAFFIC

Fig.2 Air-ground communications link

DISCUSSION

E. Lampert, Ge

It is appreciated that there is an operational necessity of providing ECM-resistant links in HF However, considering the spectral occupancy of the band and restricted RF-bandwidth because of the necessity of turning from a theoretical anti-jamming margin of about 0 to 6dB nothing will be left. Could you therefore comment on whether there is a real chance of getting ECCM equipment for HF in the near future?

Author's Reply

As mentioned in the paper, frequency-agile systems give some ECCM capability and there will be constraints on the system imposed by vehicular and medium characteristics; for example, in aircraft the HF antenna can be a high-Q device with a limited bandwidth, for transmission purposes The technology for giving an ECCM capability is with us today, the question that needs to be addressed is, as in normal HF communications, what is its "reliability"?

HF COMMUNICATIONS TO SMALL LOW FLYING AIRCRAFT

by

N. M. Maslin

Radio and Navigation Department
Royal Aircraft Establishment
Farnborough, Hampshire, England

SUMMARY

HF (2-30 MHz) radio communication is a principal means of beyond line of sight communication to aircraft. There are particularly serious problems for the small aircraft. Many factors degrade the overall received signal-to-noise ratio both at the ground and in the aircraft. To achieve satisfactory results, careful consideration should be given not only to the terminal radio equipment but also to the long term geographic planning and to the management of the frequencies to be used over mobile HF links.

The worst HF communication problems occur for a short range air-ground sky-wave link at night which requires frequencies at the low end of the HF band. Working over a longer range link increases the 'optimum working frequency', thus avoiding the poor antenna efficiencies and generally reducing external noise levels.

It is shown that good frequency management, ground antenna directivity and the use of a number of geographically separated remote receiving stations are vital in providing satisfactory communications reliability to the small aircraft.

1 INTRODUCTION

HF (2-30 MHz) radio communication is an important, and in some cases the only means of beyond line of sight communication to aircraft. Diurnal ionospheric variations necessitate the use of a number of frequencies over a 24 hour period and different propagation modes are utilized according to the range.

There are particularly serious problems for a small aircraft, whose physical size is much less than the wavelength in the lower part of the HF band. Constraints are placed upon the aircraft antenna configuration so that its efficiency may be degraded; radiation patterns are obtained that may not be suited to the propagation mode. The transmitter power is limited (to about 400 W peak envelope power) and serious excess noise, both acoustical and electrical, may be present in the aircraft. High levels of external man-made noise and interference may also be experienced in the receive mode. The aircraft height may give rise to additional multipath propagation mechanisms, whilst its speed may cause Doppler frequency shifts. All these factors degrade the overall received signal-to-noise ratio both at the ground and in the aircraft. To achieve satisfactory results, careful consideration should be given not only to the terminal radio equipment but also to the long-term geographic planning and to the management of the frequencies to be used over mobile HF links.

This paper attempts to quantify the problem of air-ground and ground-air communications reliability for the small aircraft, to pinpoint sources of performance degradation, and to suggest appropriate solutions to obtain the best available performance for such a communications link. The most difficult HF communications problems occur for a short range air-ground sky-wave link at night which requires frequencies at the low end of the HF band. Working over a longer range link increases the 'optimum working frequency', thus avoiding the poor antenna efficiencies and generally reducing external noise levels. It is shown that good frequency management, ground antenna directivity and the use of remote receiving stations are vital in providing satisfactory communications reliability to the small aircraft.

Although effects such as multipath features, time and frequency dispersion, change in apparent aircraft antenna performance due to ground reflection, uncertainties in the aircraft antenna polarisation characteristics etc have not been specifically addressed, it is believed that the fundamental principles of communication to small aircraft have been established.

2 THE PROBLEM

It is common experience that at the shorter ranges (50-500 km) HF communications can be unreliable to small aircraft. These difficulties arise as a result of one or more of the following:

(a) For small aircraft the antenna efficiencies are poor at the low end of the HF band. This fact coupled with the limited available transmitter power means that the effective radiated power is of the order of a few watts or less at these frequencies.

(b) For these shorter ranges, the sky-wave mode with frequencies in the range 2-6 MHz must often be used. It is at these frequencies that the small aircraft has poor antenna efficiencies. The problem is particularly acute at night when frequencies at the low end of the 2-6 MHz range must be used and where aircraft antenna efficiency is very poor.

(c) The ground station antenna system often has inadequate high angle ($>45^{\circ}$) coverage. Moreover the range of sky-wave angles of elevation to be covered, together with the wide azimuthal coverage required, inhibits the use of good directive antennas on the ground.

Copyright ©, Controller HMSO, London 1979

(d) The rapid variation of optimum working frequency with range between 100-250 km imposes an aircrew workload which is, at best, unacceptable and at worst, unachievable in a high performance aircraft.

(e) Interference is high at night in the 2-6 MHz band as propagation conditions restrict the use of higher frequencies and users crowd into the low end of the HF band.

(f) Propagation paths in the auroral regions can be subject to extreme attenuation with consequent communication blackout.

(g) Electrical and acoustical noise generated by the aircraft can be very troublesome, particularly when the aircraft operates at low altitude.

3 PARAMETERS OF THE HF LINK

In order to study some of these problems it is necessary to calculate the received signal-to-noise ratio, S/N . This can be expressed as

$$S/N = P_t + G_t - L - V + D - (kT) - F_a - B \qquad (1)$$

where the terms are defined as follows (logarithmic units)

P_t	transmitter power (dBW)
(kT)	thermal noise power density (-204 dBW per Hz at 300 K)
L	path loss (dB)
V	polarisation mismatch loss (dB)
G_t	transmitting antenna absolute gain with respect to an isotropic antenna (dB)
D	receiving antenna directivity factor against far field noise (dB)
B	receiver bandwidth (dB Hz)
F_a	effective antenna noise power factor (dB) which results from external noise power available from a loss free antenna. The receiver performance is taken to be limited by external noise.

The signal-to-noise density ratio (dB) (*ie* the signal-to-noise ratio in a 1 Hz bandwidth) is then

$$S/N_0 = S/N + B \qquad (2)$$

For the air-to-ground link, the available power P_t from the aircraft transmitter is assumed to be 400 W pep. The HF prediction program (CCIR, 1974a; HAYDON, G.W. *et al*, 1976) 'Bluedeck' was used to calculate the path loss L over various sky-wave paths and F_a was assumed to be governed solely by atmospheric noise. The polarisation mismatch loss factor was taken to be constant at 3 dB; the aircraft antenna efficiency and hence G_t is a function of the carrier wave frequency and was assumed to be that typical for a small aircraft (PAVEY, N.A.D., 1973). For example the efficiencies at 2, 5 MHz are typically 0.01% and 1% respectively.

The assumption that man-made noise and interference are below the levels of atmospheric noise ensures that the results given here are the optimum obtainable. In some of the following examples difficulty in communicating is experienced even under ideal conditions. Hence when man-made noise, interference and aircraft noise problems are present, satisfactory communication will be even more difficult to achieve.

The calculations are made for median signal-to-noise (S/N) ratios. The CCIR (CCIR, 1974b) recommends S/N ratios of 15 dB for marginal and 33 dB for good commercial quality of communications with HF SSB voice in a 3 kHz bandwidth. This corresponds to a signal-to-noise density ratio, S/N_0 , of 50 dB for marginal and 68 dB for good commercial quality. We shall work throughout in terms of signal-to-noise density ratios so that the results presented here may be applied to other modulation modes.

To illustrate the results obtained, a number of widely geographically separated locations of ground stations have been considered; these are given in Table 1 and are designated W, X, Y, Z. Although the ranges to stations Y and Z are similar they are assumed to be well separated in latitude.

Table 1

Parameters relevant to the ground stations

Ground station	Range (km)	Classification
W	280	Short range
X	1200	Medium range
Y	2500	Long range
Z	2200	Long range

4 AIR-GROUND LINK PREDICTIONS

4.1 Form of the predictions

The predicted median signal-to-noise ratios for three sky-wave circuits are given, assuming that atmospheric noise is the predominant noise source at the receiver site. The density of shading in Figs 1 to 3 shows increasing values of received signal-to-noise density (S/N_0) for a given frequency at a specified hour of the day. Because S/N_0 is a median value, the achievement of a given criterion gives a communications reliability (MASLIN, N.M., 1978) of approximately 50%. Thus cross hatched squares in these examples indicate S/N_0 values of between 55 dB and 60 dB. They show that 50% reliability or greater would be produced for a 20 dB signal-to-noise criterion. This is equivalent to SSB voice in a 3 kHz bandwidth with $S/N_0 = 55$ dB. Where no shading is shown in the diagram S/N_0 is less than 45 dB; SSB voice transmissions would at best be barely acceptable under these conditions and communications reliability would be unacceptable.

The highest available frequency for adequate communication at a given hour is taken to be the FOT (optimum traffic frequency) defined here as the sky-wave frequency available for 90% of the time. The lowest usable frequency (LUF) depends upon the required S/N_0 criterion, and thus an improvement in the level of received signal or a reduction in the noise level will lower the LUF. The FOT, however, is independent of S/N_0 criteria, being only a function (MASLIN, N.M., 1978) of the prevailing ionospheric conditions, which in turn depend upon hour of the day, season of the year and sunspot number.

4.2 Short range communication

Fig 1 shows the predicted S/N_0 ratios for communication with ground station W during 1976, a low sunspot (SSN) number year (~15). It should be noted that communication is more difficult in summer months than in winter, partly as a result of increased atmospheric noise levels at the ground station. Between 800 hours and 0600 hours (ie mainly dusk-night-dawn) the S/N_0 ratio does not reach 55 dB for any aircraft position at any time of year.

4.3 Medium and long range communication

Consider now, ranges of communication greater than 1000 km. Figs 2 and 3 show the expected S/N_0 ratios at receiving stations X and Y for aircraft transmissions in the low SSN year. The following should be noted:

(i) Between 2000 hours and 0400 hours the S/N_0 ratio does not reach 55 dB at station X for any time of year.

(ii) For the long range link (to station Y), the received S/N_0 ratio at the FOT tends to be greater at night.

(iii) Over the long range link S/N_0 values for the daytime tend to be poor.

4.4 More than one remote station

The value of working over a longer range link is shown schematically for a conventional ground station in Figs 4 to 6. Fig 4 shows the maximum S/N_0 obtainable for any frequency for a given hour of the day by a small aircraft transmitting in both high (100) and low (15) sunspot number years. The shaded areas show reception at station W and the letters X, Y, Z designate the remote receiving stations of Table 1. A 55 dB S/N_0 criterion can be achieved for a large percentage of the day (see Fig 5) by using these longer range links. If ground station W only were to be used for reception, a 55 dB criterion would be achieved for only a small percentage of the day in winter and not at all in the summer months.

A particularly valuable feature of using more than one remote station is that a fairly constant (to within ~10 dB) S/N_0 can be achieved throughout a 24 hour period provided that different ground stations are utilized at different times of day. Moreover if the remote stations are at various ranges and on different bearings from the aircraft, careful planning could virtually eliminate the need for frequency changing.

Fig 6 demonstrates the consistency of frequency for a given range and period of the day. In the winter months the frequency used is 7.5 MHz and in summer 9 or 10.5 MHz depending upon solar activity. The medium range link (~1200 km) to station X occupies the daylight hours and the longer ranges to stations Y and Z cover the difficult dawn-dusk period and night-time conditions between them, as a result of their geographical separation. This example of a constant frequency can be compared with Fig 4 which shows the effects of using the optimum frequencies. It can be seen that only a few dB are lost by carefully choosing the constant frequency and utilizing all three remote stations. This situation would help to relieve aircrew work load of constantly changing frequencies. Advantages of the longer range links are summarised in Table 2.

4.5 Further improvements in signal-to-noise

The result of introducing a 10 dB ground directivity is effectively to reduce the 55 dB criterion to 45 dB. Once again the advantage of a remote station is apparent, since the 45 dB criterion is achieved for only a small fraction of the time at station W during the summer months (Fig 5).

An improvement in aircraft antenna efficiency at the low end of the band (2-6 MHz) can enhance the S/N_0 ratio received at station W, although it still does not offer the same advantages as working over the longer range links. Flight trials (SLOGGETT, D.R., 1975) performed by RAE, Farnborough have shown that the low effective radiated power from a small Dominie aircraft at the low end of the HF band, due to poor antenna efficiency, is a major contributory factor to the unsatisfactory air-ground communications at night.

4.6 HF communications reliability

This is defined as the fraction of days that successful communication may be expected at a given hour within the month at a specific operating frequency. A detailed discussion of HF reliability is given in a companion paper (MASLIN, N.M, 1979). The result of Figs 4 and 5 is summarised in Fig 7 in terms of the reliability at a given time of day. The advantages of the suggested improvements are self-evident.

Table 2

Factors connected with using remote ground receiving stations

Requirements for remote station working		Consequences of using remote stations	
A	Only frequencies above ∿7 MHz are required	1	Poor aircraft antenna efficiencies are avoided
		2	Interference is reduced at night since lower frequencies are employed by other users
		3	Ground antenna size is reduced
B	Antennas require only narrow azimuthal beamwidth and low angle coverage	1	Good directivity should be obtainable; thus noise and interference can be discriminated against
		2	Antenna steering techniques are not necessary
C	Aircraft should be at least 1000 km from remote stations	1	Frequency changing with time of day is not necessary as often as for shorter range links
		2	Changes in aircraft position will not necessitate frequency changing
		3	For aircraft in Northern waters reflection of signals in auroral regions can be avoided
D	Two or three remote stations ideally required, with adequate geographical separation	1	Whole 24 hour period can be covered by using ranges of ∿1200 km by day, ∿2500 km by night
		2	Frequency changing can be effectively eliminated by careful choice of ground station operation
		3	Pilot work-load is reduced by using a single frequency

5 GROUND-AIR LINK PREDICTIONS

5.1 Aircraft generated noise

Probably the most important feature of the ground-air link is the noise environment of the aircraft. All aircraft systems which use electrical energy are likely to generate unwanted EM energy, and this may couple into the aircraft radio systems and degrade their performance. Conversely, almost every aircraft radio transmitter generates intense EM fields which may affect other avionic systems, including installed radio systems operating at the same time. Coupling mechanisms between an interference source and the rest of the avionic installation may be complicated; interference levels are affected by factors such as design, practice and workmanship of the avionic installation, imperfect shielding of braided coaxial cables and the RF attenuation offered by the aircraft skin.

At HF it is common experience that the noise received by the antenna increases as the aircraft systems are progressively switched on, but little quantitative information seems to exist. Accordingly the following approach has been adopted to obtain an order-of-magnitude estimate of broadband interference fields in the vicinity of an aircraft HF antenna.

Specification BS 3G 100 lays (BSI, 1973) down maximum permissible radiated interference limits for equipment, expressed in dB relative to $1\ \mu Vm^{-1}$. These limits have been recalculated as dB relative to antenna thermal noise power kT_oB for narrowband and broadband interference. The resulting values have been taken to represent the EM field levels existing inside the aircraft from all of the installed systems and the EM fields immediately outside the aircraft skin have been estimated by putting the hull attenuation equal to 30 dB.

It is clear that in a study of ground-air communications reliability more detailed information is required about levels of aircraft electrical noise so that more accurate estimates, than those possible at present, may be made of the noise environment in the aircraft.

An aircraft in flight can become electrically charged by the influence of electric fields in the atmosphere, by the impact upon its fuselage of both charged and neutral particles and by ionization in its engine. The currents which flow during the charging and subsequent discharging processes radiate and contribute a component, called precipitation static, to the noise power received by an antenna on the aircraft. In unfavourable circumstances and if no steps are taken to reduce it, the intensity of precipitation static noise can exceed the intensity of the noise from any other sources in the HF band (GRANVILLE-GEORGE, D.A., SMITH, B.G., 1975). If dischargers are fitted to the aircraft, however, the antenna noise figures can be reduced by 40 dB to 60 dB. Under these circumstances the radiated interference, reduced by the appropriate value of hull attenuation, is the dominant noise source. This is shown as curve H in Fig 8, and can be compared with other sources of noise in the HF band.

5.2 Propagation aspects

Since the poor aircraft antenna efficiencies are relatively unimportant in the receive mode (see Appendix) it is ionospheric absorption that is the most important frequency dependent parameter in the link.

This, in turn, implies that frequency management is relatively more important than for the air-ground situation. It is assumed that the transmitting antenna efficiencies for the ground station are not frequency dependent.

5.3 Effective radiated power

For the levels of noise in the aircraft assumed here, a similar calculation to those performed for the air-ground link shows that a 1 kW effective radiated power from the ground transmitter is not adequate; 10 kW (40dBW) would be more suitable. Communication would still be difficult during the daytime, however, particularly during the summer months. Under these circumstances 50 dBW would be more appropriate. For the shorter range links the available frequency band is severely limited and there may be considerable problems with interfering stations, particularly at the lower end of the HF band. For the longer range links the available frequency band is wider but the ionospheric absorption tends to be stronger for a given frequency due to the increased propagation path length.

5.4 Communications reliability

Some of the results for the ground-air link can be summarised in Fig 9 which shows the effect of increasing the radiated power from the transmitter at station X . It is clear that 1 kW (30 dBW) does not provide adequate communications reliability since at no time does the reliability figure reach even 50%. An appropriate 'operationally acceptable' figure would be closer to 80%. Under these circumstances 50 dBW of radiated power is generally required. This implies some directional gain on the part of the ground transmitting antenna.

6 OTHER ASPECTS OF REMOTE TERMINAL OPERATION

6.1 Beamwidths of ground station antennas

Table 3

Requirements for ground station antenna coverage
(angles in degrees)

	Ground station		
	W	X	Y
Minimum elevation angle	10	0	0
Maximum elevation angle	60	25	17
Elevation beamwidth	50	25	17
Azimuthal beamwidth	30	20	10
Gain (dBi)	13	18	23

Table 3 shows the azimuthal and elevational angles and beamwidths required to accommodate the predominant propagation modes, using different remote ground stations for a typical flight profile. The achievable gain in Table 3 is taken to be

$$G_t = 30000/W_H W_V \qquad (3)$$

where W_H and W_V are the azimuthal and elevational beamwidths. Table 3 indicates the gain G_t which could theoretically be achieved by concentrating all the available transmitter power into the appropriate antenna beamwidths.

Thus, for example, the required beamwidth coverage from station Y is narrower than from station W , and an extra 10 dB gain should be possible over that achievable from station W . This should more than offset the extra ionospheric absorption which the longer range link experiences.

6.2 Operational frequencies

Ideally the optimum traffic frequency (FOT) needs to be continually changed because of the varying range of the aircraft. The rate of change is a function of the distance from the aircraft to the ground station and is dependent upon the time of day. This is illustrated in Fig 10 which shows the percentage change in FOT required for communication to the aircraft flying a typical mission.

The advantage of the longer range link is clear. It can avoid changing the FOT over, say, a two hour flight. For ground station W , however, the frequency would have to be continually changed over a wide range of frequencies.

6.3 Auroral effects

Consider an aircraft within the auroral region, which extends down to about 60°N in Europe. If station Y is used as the ground station then the reflection area of a one hop mode might be around 50°N but for the link to station X would be near 62°N. An advantage may therefore be obtained by working over the long link, thus avoiding the troublesome auroral region, (LIED, F., 1967).

7 CONCLUSIONS

The worst HF communication problems for a small aircraft occur over a short range sky-wave link at night which requires frequencies at the low end of the band, where transmitting antenna efficiencies are very poor. The only satisfactory solution is to work to ground stations at longer ranges to permit the use of sufficiently high operating frequencies. Good antennas are required at the ground stations since the improved signal-to-noise ratios that can result from the use of directional antenna can substantially increase the overall link reliability. By using good frequency management and by working to a remote ground station, a 50% reliability figure is achievable under average propagation conditions. Good ground antenna directivity can substantially improve upon this figure.

We have seen that the use of longer range (>1000 km) links offers the following advantages:

(a) Higher frequencies in the HF band can be used since the MUF is greater over the longer link. The resulting increase in aircraft antenna efficiency is considerably more than the increased propagation loss due to increased path length.

(b) The high angle coverage requirements of the ground antenna system are reduced. Moreover the range of elevation and azimuth angles to be covered is reduced, making directive antenna systems feasible. Avoidance of the usage of the lower HF band (<6 MHz) reduces antenna dimensions and cost.

(c) Interference is reduced at night since lower frequencies are employed by other users.

(d) The optimum working frequency as a function of variation in position of the aircraft changes slowly. The aircrew workload of frequency changing becomes acceptable and achievable.

(e) By using two or more ground stations which are widely separated geographically, frequency changing can often be eliminated by careful choice of ground station operation. The chance of communications blackout due to anomalous propagation conditions is considerably reduced by employing this extreme spatial diversity scheme.

(f) For aircraft in auroral regions, reflection of signals in regions where auroral absorption occurs can be avoided by using longer links. The possibility of communications failure is reduced.

In the ground-air situation improvements that can be made to the link are generally limited to the ground terminal because of cost and design considerations on the aircraft. Poor aircraft antenna efficiencies do not usually degrade the signal-to-noise ratio on reception, although it is important to receive an adequate signal level at the receiver. The primary constraint on the ground-air link is the noise at the aircraft terminal. This can be divided into two catagories: electrical and acoustical noise.

The level of electrical noise degrades the received signal-to-noise ratio at the aircraft antenna receiving terminals whilst the level of acoustical noise at the crewman's ear determines whether the signal from his headphones (with an adequate signal-to-noise ratio) is intelligible. It is important, therefore in a study of the ground-air link to be able to adequately assess the aircraft noise levels.

With a knowledge of these noise levels the required output power from the ground station in the aircraft direction can be estimated. The transmitted power is more readily controllable than in the air-ground situation. The directivity and absolute gain of the transmitted signal can be controlled reasonably effectively and it is vital to make use of these features if adequate signal-to-noise ratios are to be received at the aircraft.

Remote ground station working is not as essential for providing adequate communication as in the air-ground link. Ionospheric absorption is the most important frequency dependent factor in the ground-air direction whereas the dependence of the aircraft antenna efficiency on frequency is more important in the air-ground direction. A remote ground station link does, however, increase the available frequency range of operation and decrease the required range of elevation angles at the transmitter. The remote link may be useful if there is strong interference on some frequencies. It is usual for the aircraft to transmit and receive on the same frequency. Thus, since the air-ground situation requires the longer link for satisfactory operation the same link and hence the same operating frequency will have to be used for the ground-air link.

REFERENCES

BRITISH STANDARD INSTITUTION, 1973, "General requirements for equipment for use in aircraft (BS3G100)". Part 4, section 2 "Electromagnetic interference at radio and audio frequencies", BSI, London.

CCIR, 1974a, Report 252-2, "Interim method for estimating sky-wave field strength and transmission loss at frequencies between the approximate limits of 2 and 30 MHz", ITU, Geneva.

CCIR, 1974b, Recommendation 339-3, "Bandwidths, signal-to-noise ratios and fading allowances in complete systems", ITU, Geneva.

GRANVILLE-GEORGE, D.A. and SMITH, B.G., 1975, "An information study of noise in aircraft communication systems between 2 and 400 MHz", Smith Associates Technical Report TP-123.

HAYDON, G.W., LEFTIN, M. and ROSICH, R., 1976, "Predicting the performance of high frequency sky-wave telecommunication systems". (The use of the HFMUFES4 Program.) OT Report 76-102.

LIED, F. (Editor), 1967, "High frequency radio communications, with emphasis on polar problems", (p.112), AGARDOGRAPH 104, Technivision, Maidenhead, England.

MASLIN, N.M., 1978, "Assessing the circuit reliability of an HF sky-wave air-ground link". The Radio and Electronic Engineer, 48, 493-503.

MASLIN, N.M., 1979, "Assessment of HF communications reliability", Paper 40, AGARD Symposium on 'Special Topics in HF Propagation', Lisbon.

PAVEY, N.A.D., 1973, "Radiation characteristics of HF notch aerials installed in small aircraft". AGARD Conference Proceedings CP-139.

SLOGGETT, D.R., 1975, "Ground measurements of HF transmissions from Dominie aircraft XS 738 flying at low level": Part IV, RAE Technical Report (unpublished) TR 75003.

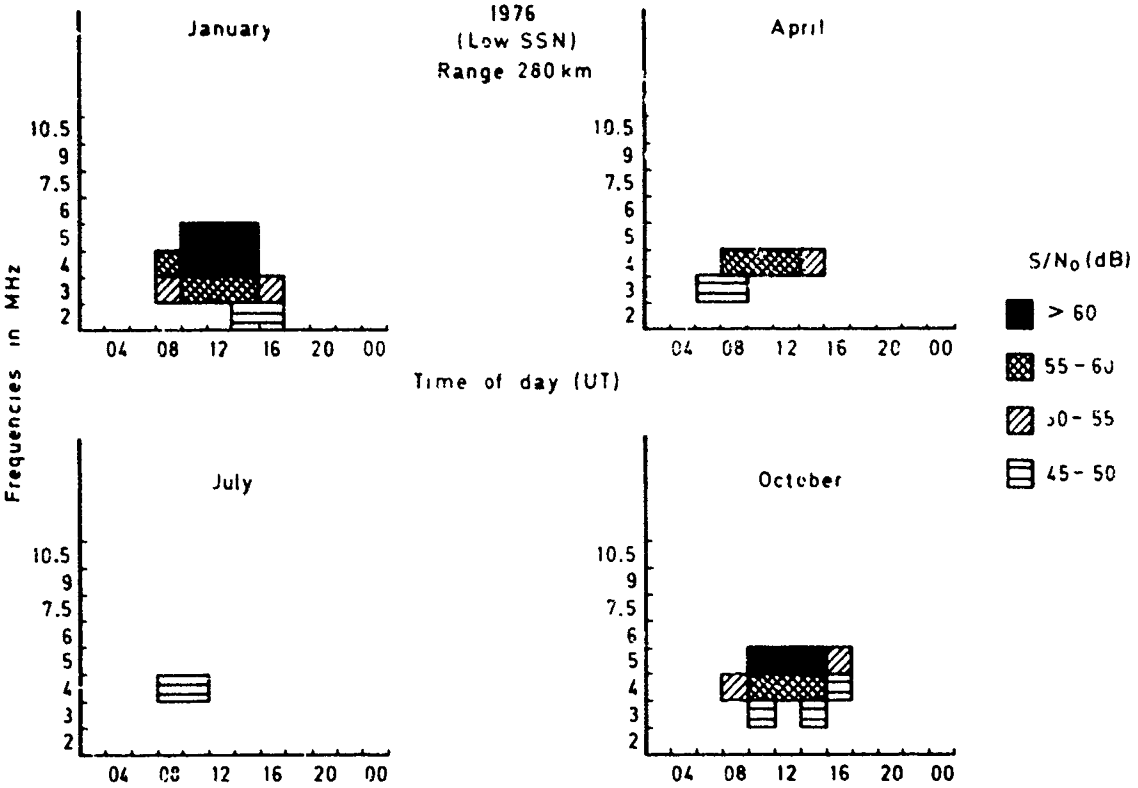


Fig.1 Short range communication (aircraft to station W)

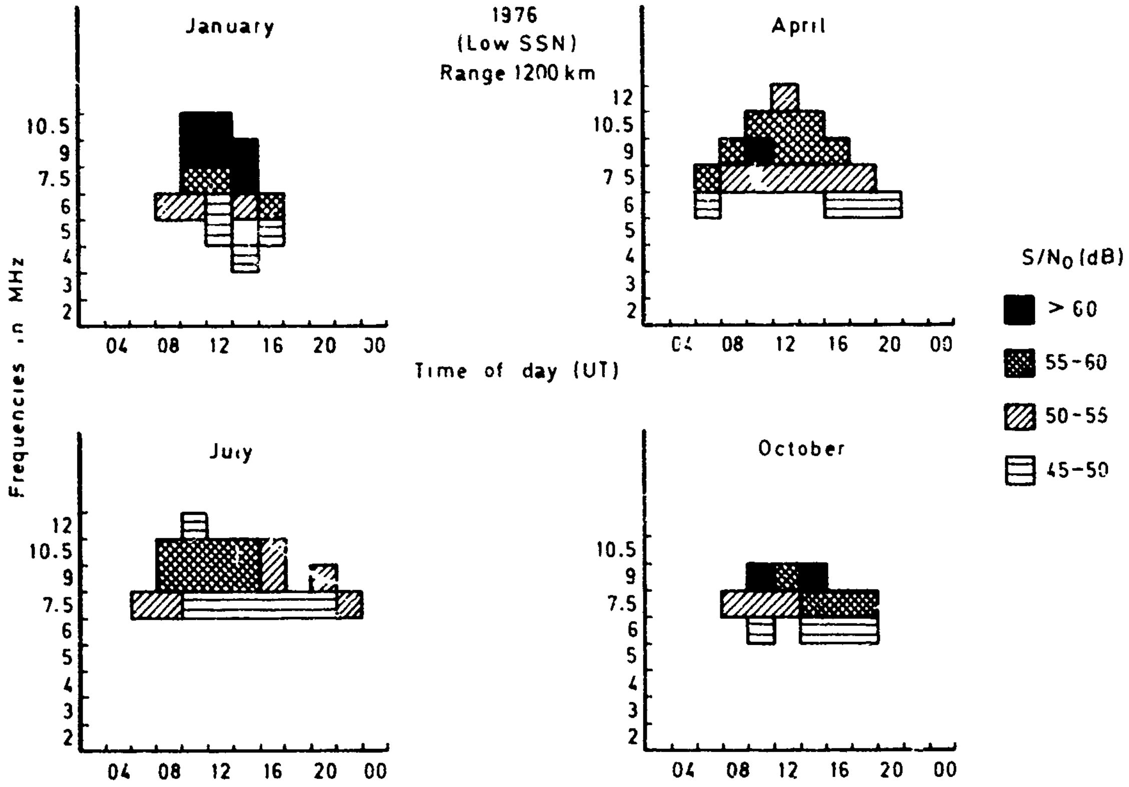


Fig.2 Medium range communication (aircraft to station X)

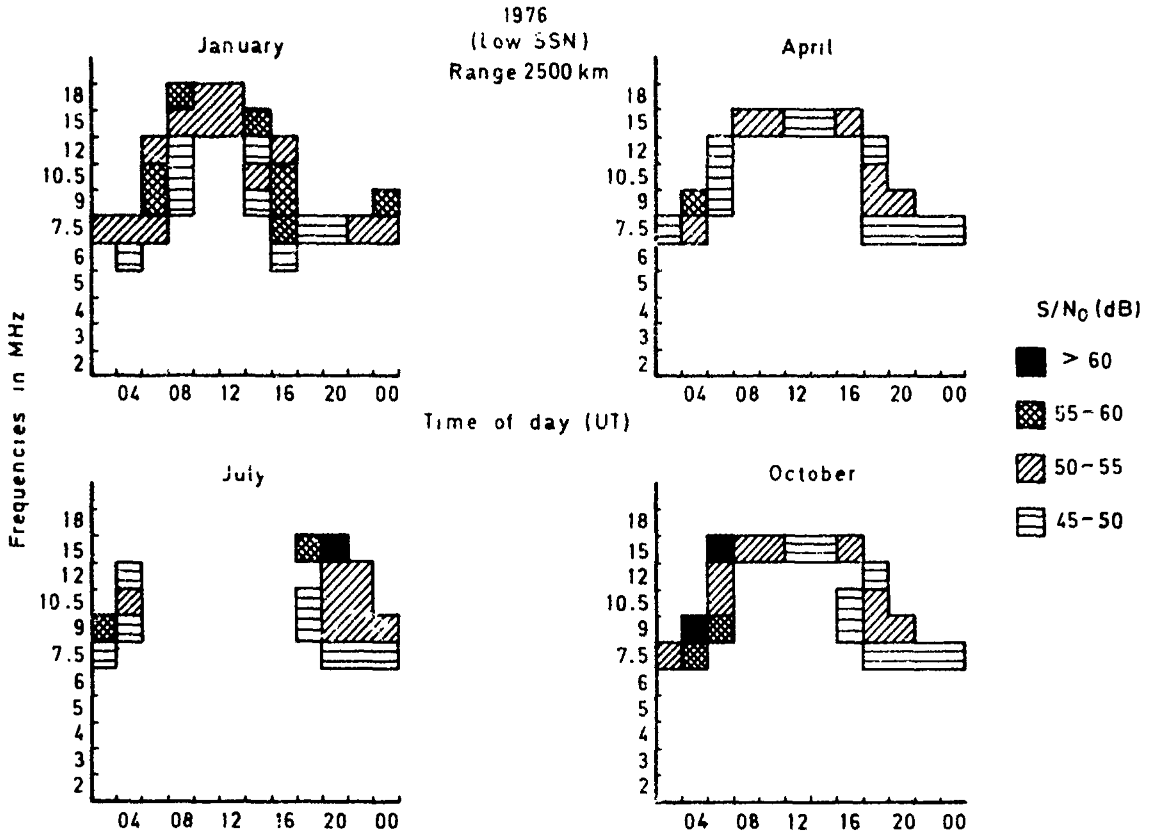


Fig.3 Long range communication (aircraft to station Y)

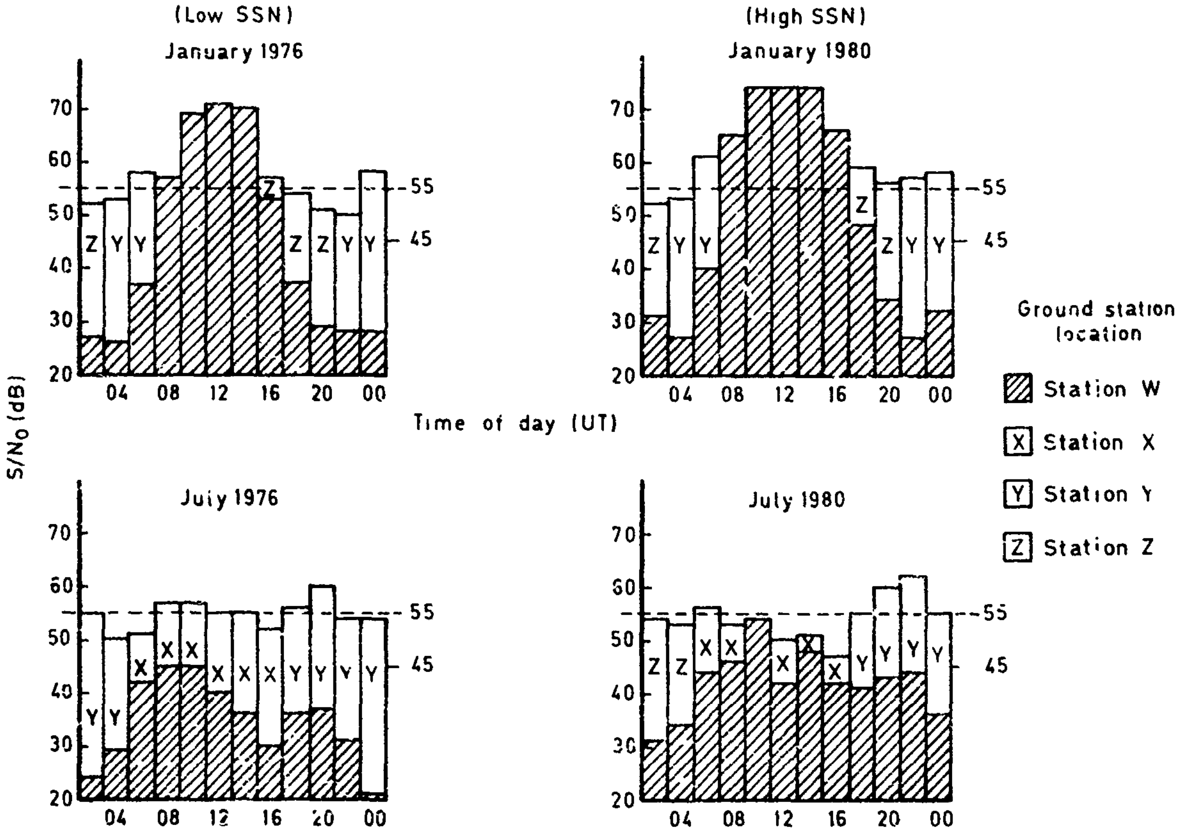


Fig.4 Communication from aircraft to various ground stations. Optimum frequencies are used but no ground antenna directivity or improvement in aircraft antenna efficiency is assumed

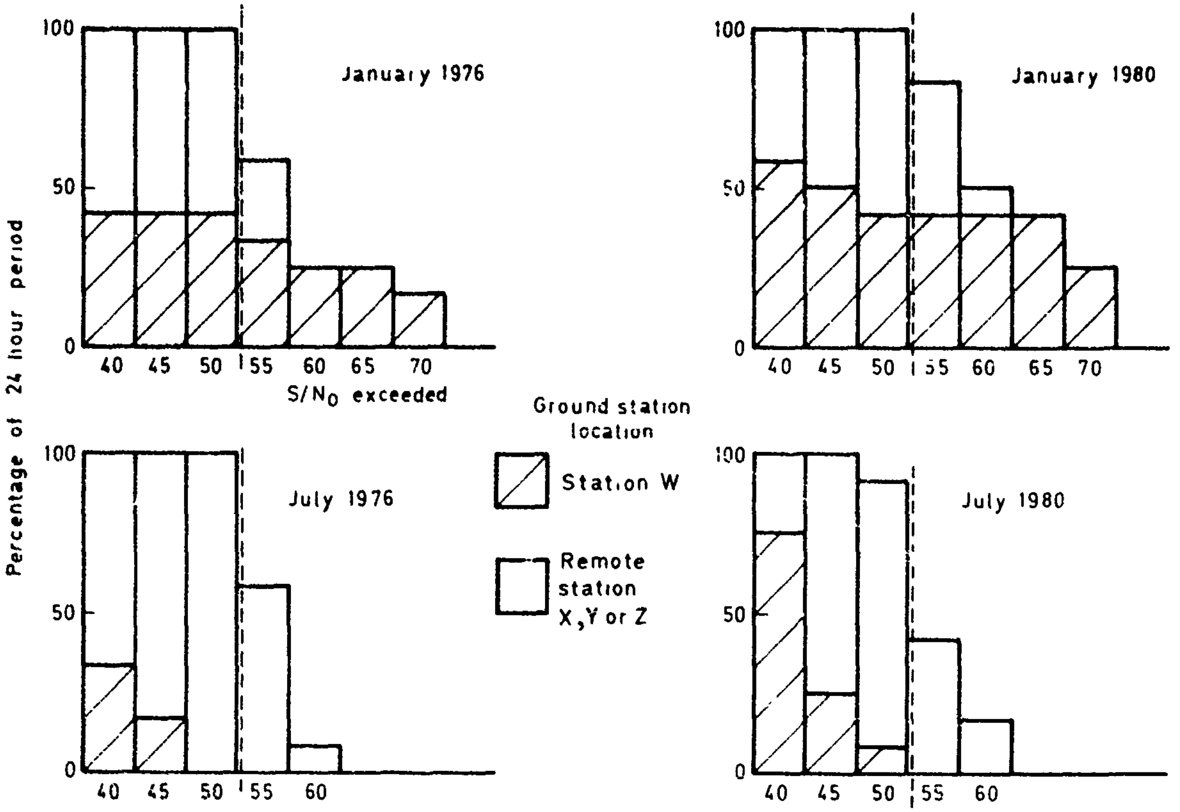


Fig.5 Percentage of the day a given S/N$_0$ is exceeded for aircraft to ground station links (No ground directivity or improved aircraft antennas assumed)

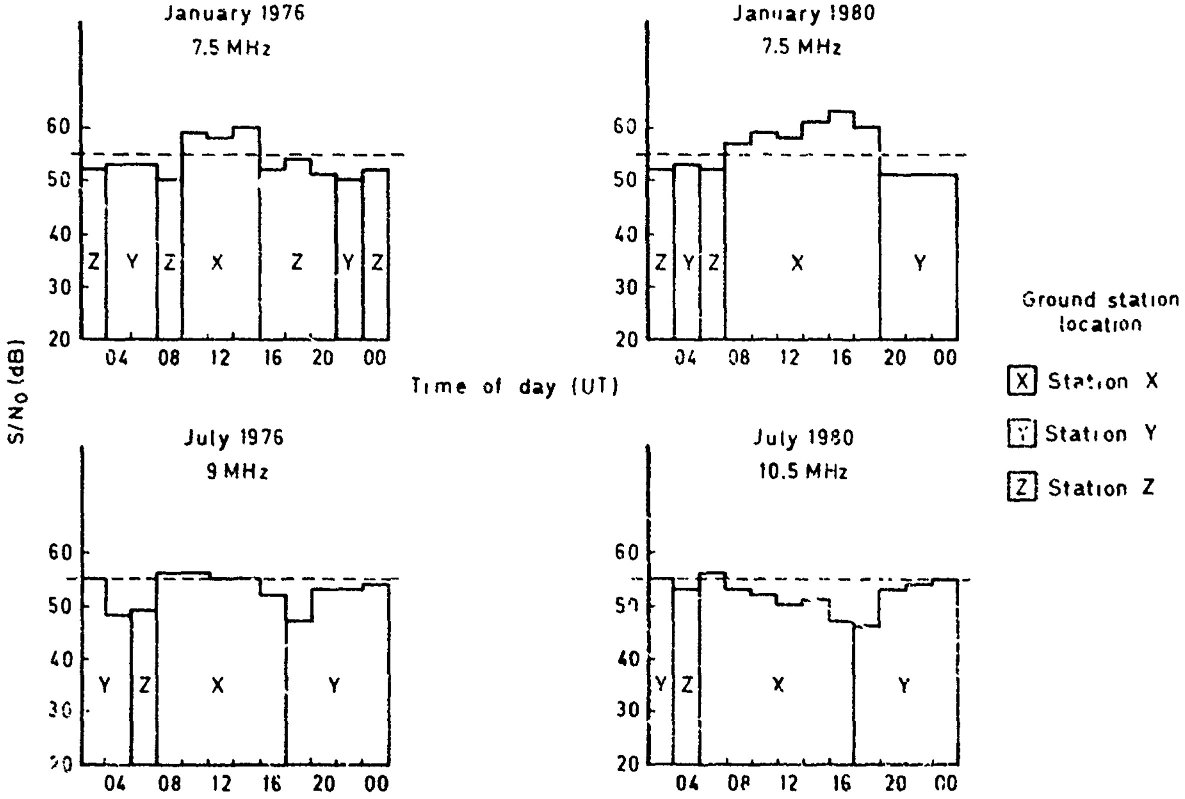


Fig.6 Communication from an aircraft to various ground stations. A constant frequency is used over a 24 hour period. (No ground directivity or improved aircraft antennas assumed)

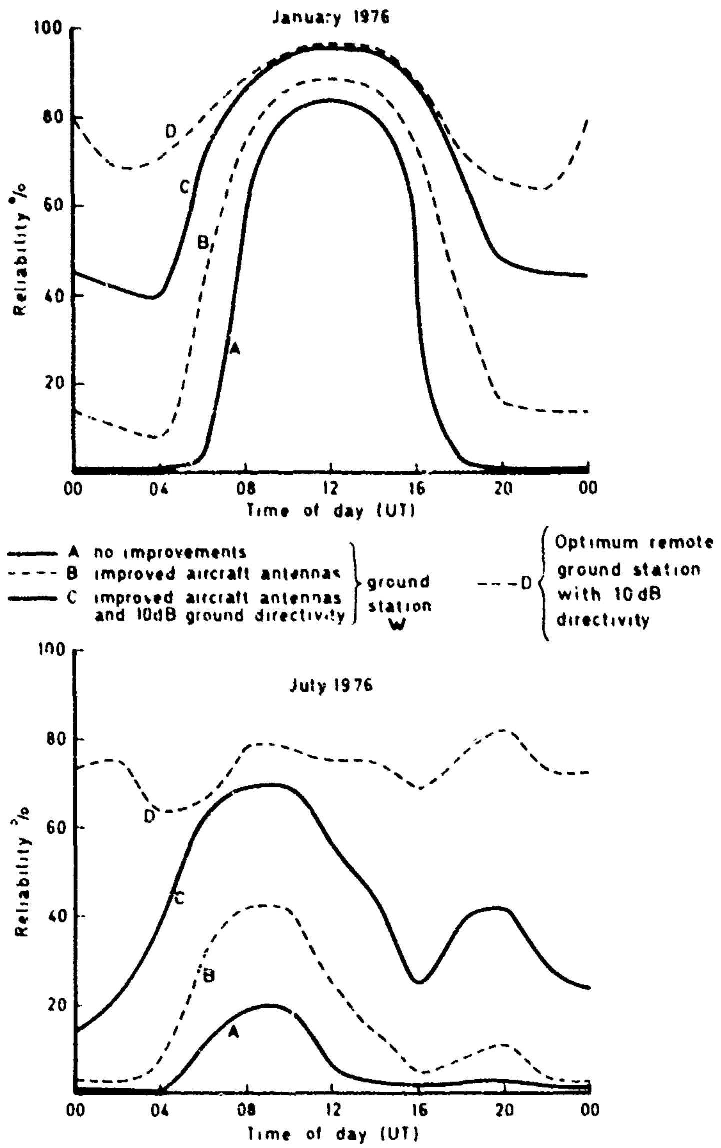


Fig.7 Summary of HF air-to-ground communications reliability

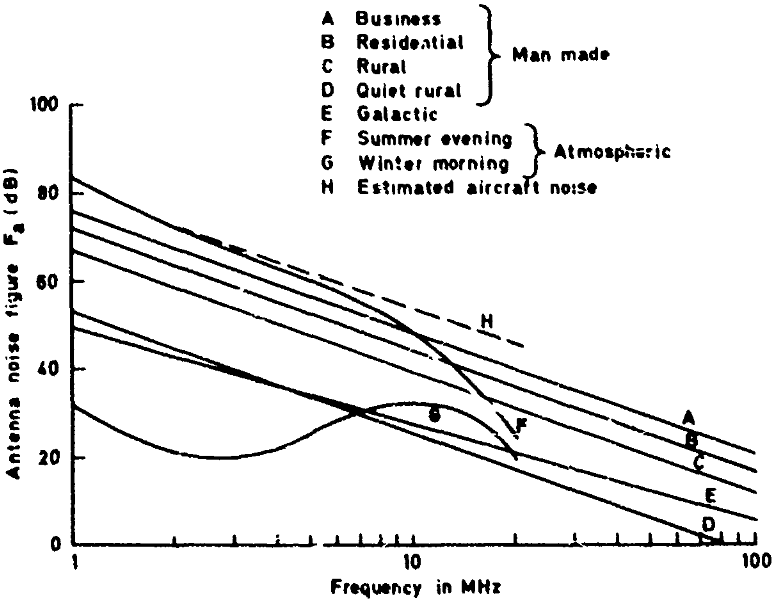


Fig.8 Antenna noise figures for external noise sources in the HF band

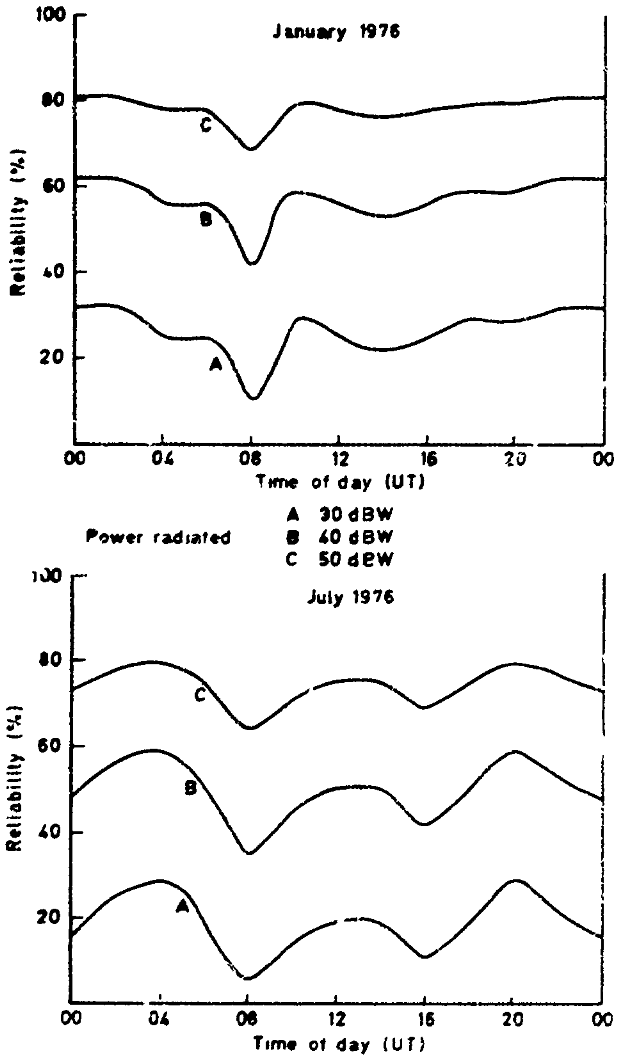


Fig.9 Summary of HF ground-to-air communications reliability

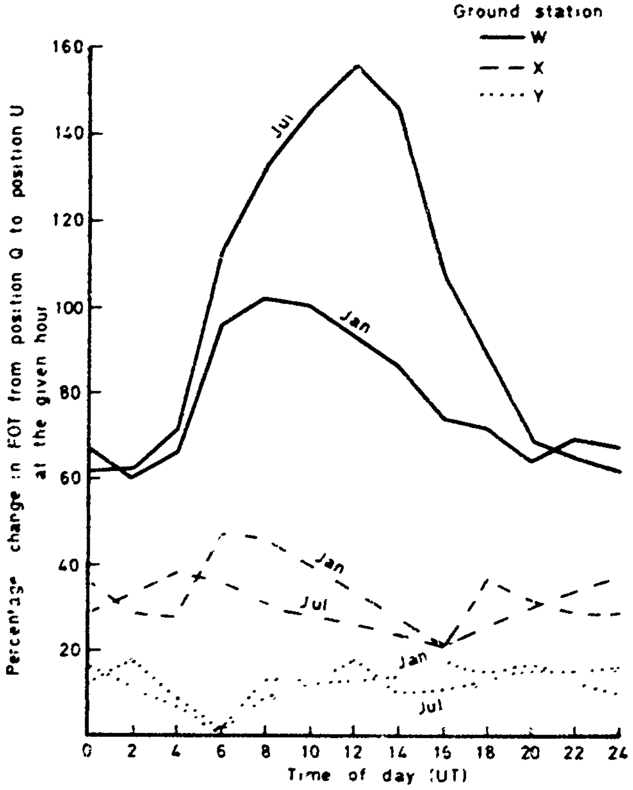


Fig.10 Effect of aircraft range upon operating frequencies for various ground stations in a low sunspot number year Position Q to position U is assumed to be 750 km

Appendix

RECEPTION PROBLEMS WITH SMALL AIRCRAFT

The reception of HF signals by small aircraft is subject to a number of sources of degradation and it is necessary to know how these can affect the overall circuit reliability on a ground-air link. There are two important features to consider.

(a) The degradation of the signal-to-noise ratio as the signal passes through the receiver.

(b) The absolute level of the signal at the receiver input.

A.1 Degradation of signal-to-noise ratio

Consider an incoming signal of mean power S at the aircraft antenna and an incident noise power N_0 in a 1 Hz bandwidth. Let the (omni-directional) antenna have efficiency η at the frequency considered. Then after reception by the antenna the signal power is ηS, the noise power is ηN_0 and the signal-to-noise ratio remains S/N_0. At the receiver input other losses such as those due to imperfect matching reduce the signal strength to $\eta' S$ and the noise power to $\eta' N_0$.

As the signal passes through the receiver, the receiver noise power N_r per Hz contributes to the total noise power and the final signal-to-noise ratio becomes

$$\eta' S/(\eta' N_0 + N_r) \qquad \text{(A-1)}$$

if

$$N_r \ll \eta' N_0 \qquad \text{(A-2)}$$

then the signal-to-noise ratio at the receiver output is the same as at the antenna. Now (A-2) can be expressed in logarithmic units as

$$\eta' + F_a \gg F_r \qquad \text{(A-3)}$$

where η' dB is the loss in the receiving system, F_a is the effective antenna noise power factor (dB) and F_r is the receiver noise figure. Inequality (A-3) is the condition that the receiving system performance in the aircraft is limited by external noise. Provided that (A-3) is valid the receiving antenna efficiency is unimportant.

Suppose that $\eta = \eta'$, ie only the receiving antenna efficiency contributes to the power losses. For an antenna on a small aircraft, η may be very small. The receiver noise figure can be taken as 10 dB, which is typical of good current practice; typical antenna efficiencies are given elsewhere (PAVEY, N.A.D., 1973). Then the minimum required external noise to ensure that receiver noise is unimportant can be calculated.

With the best estimates of F_a available equation (A-3) is nearly always satisfied. Although the aircraft noise figures are estimates it is felt that they are unlikely to be in error by more than a few dB. Thus even if there are mismatch losses in the receiving system of 10 dB (corresponding to a VSWR of approximately 40:1), the signal-to-noise ratio should not be degraded on reception.

A.2 Absolute signal level

Because of the nature of the airborne receiver it is necessary for the input signal to the receiver to be of the order of 2 μV (6 dB μV). It has been shown that to achieve satisfactory communication for SSB voice channels a signal-to-noise ratio (S/N_0) of 55 db is required. Now if the noise power at 2 MHz is -132 dBW per 1 Hz bandwidth, the signal S at the antenna must be -77 dBW. Suppose at 2 MHz η is -38 dB and assume a 40:1 VSWR; then the value of η' is 48 dB, and $\eta' S$ is -125 dBW in the worst case. This is 12 dB μV (in 50 Ω) at the receiver input, which is quite acceptable. To protect against fading a margin of some 6 dB is required. Therefore even under the most unfavourable conditions the signal level should be adequate at the receiver. Note, however, to achieve -77 dBW at the aircraft would require 40 dBW or more of radiated power from the ground station.

MODERN HF COMMUNICATIONS FOR LOW FLYING AIRCRAFT

JOHN F. BRUNE and BERNARD V. RICCIARDI
Communication and Sensor Division
US Army Avionics Research and Development Activity
Fort Monmouth, New Jersey

SUMMARY

The objective of this paper is to describe an application of the HF propagation phenomenon that can provide relatively short range, reliable, terrain independent communications. It has been shown that communications out to ranges of 50 km, under varying terrain conditions, to and from low flying aircraft, is an extremely difficult problem especially for air mobile tactical forces. The US Army under the Nap-of-the-Earth Communications (NOE Comm) System program has conducted extensive tests and analysis using the HF media for tactical communications with low-flying aircraft. This paper, based on the tests performed, describes how the near-vertical-incidence skywave (NVIS) portion of the HF channel can be used for aircraft communications. After considering the results of the tests and analysis, the US Army is planning to procure and field a modern HF-SSB radio which will take advantage of this channel to improve NOE aircraft communications.

HF-SSB radio systems have the capability of operating in either a ground wave or NVIS mode. For the NVIS mode, the energy is directed vertically to the ionosphere and returned to the surface of the earth. The NVIS mode provides umbrella type coverage. Because of NVIS propagation, HF-SSB systems with appropriate antennas have the capability of providing communications coverage out to ranges greater than 50 km in any type of terrain. The NVIS mode is terrain independent.

This paper addresses primarily the characteristics of the HF NVIS mode and also notes the features required of a modern HF radio system to make efficient and practical utilization of the already overcrowded HF band.

1. INTRODUCTION:

1.1 BACKGROUND: Aircraft radios utilizing the frequency range of 30 to 400 MHz have been standard equipment on Army aircraft for more than a decade. These sets have provided an effective communication system with normal high flight altitudes; however, current battlefield tactics require very low levels (often in defilade) to reduce vulnerability to sophisticated antiaircraft weapons systems. The effectiveness of the communication links utilizing the VHF and UHF bands is severely impaired due to loss of line-of-sight (LOS) conditions. The present VHF and UHF systems are limited to short ranges which are often insufficient to support combat operations involving aircraft on the modern battlefield. Thus an alternate method is necessary for communications to, from and between low-flying aircraft.

A system that uses an HF skywave propagation path provides such an alternate method that functions independently of terrain features. This system is based upon a natural phenomenon whereby a radio frequency signal directed at the zenith within a limited band of frequencies is reflected back to the earth from one of several of the ionized layers with a minimal path loss. This propagation mode is identified as the Near-Vertical-Incidence-Skywave (NVIS) mode. The basic scientific principle of this approach has been observed and reported early in the days of electro-magnetic experimentation. The NVIS mode was used for Army ground-to-ground communications since World War II; however, the short range characteristic afforded by the NVIS mode had little practical application to Army aviation until recently. The classic application of the HF spectrum is for communication over long ranges (e.g., 200-8000 km). The current Army need is for a system that functions at a close range of 0 to 50 km, the range where VHF or UHF frequencies are normally employed. The choice of frequencies to achieve short range skywave paths are different than those normally used for a long range application. The best frequency of operation lies within a small pass band (sometimes only 1 to 2 MHz wide) within the 2 to 10 MHz frequency band. The position of the pass band or "window" is a function of the time of day, season of the year, degree of sun spot activity, geographical location, etc. Modern computer propagation prediction techniques or an area ionospheric sounder can readily provide the guidance necessary to develop an acceptable frequency management plan.

Use was made of the NVIS principle for ground-to-ground communications by the Armed Forces during World War II, when transmission difficulties were encountered with VHF equipment operating in jungles and mountainous terrain.[6] The principle was further pursued and used during the Vietnam conflict.[5] Some helicopters in Vietnam were equipped with HF-SSB equipment (including an open wire zig-zag antenna). This approach did not receive widespread favor as a solution to the NOE Communications problem for reasons such as:

a. Difficulty in predicting and using the best frequency of operation to permit frequency assignments to follow the "window".

b. The available aircraft radio did not contain state-of-the-art technology which resulted in a poor communications channel (e.g. no squelch or preset channelization).

c. The aircraft antenna for the 2 to 10 MHz range was inefficient for NVIS propagation and mechanically undesirable.

d. Aviators lacked understanding of the HF media.

e. The pilot burden in flying aircraft at low levels is too great for him to devote the time required to communicate properly using the HF media.

1.2 TESTING:[1,8,10] Tests on the NOE Comm low flying aircraft communications program were both of an engineering and operational nature. The engineering tests were conducted at Fort Monmouth, and in the Central, Western and Northern, New Jersey areas which are characterized as both moderately rolling and hilly terrain; in the rugged White Mountain area of New Hampshire; and in both the hilly and mountainous areas of Oahu, Hawaii. The operational tests were conducted in the moderately rolling terrain of Fort Hood, Texas.

1.2.1 Fort Nonmouth, New Jersey Area Tests: The tests in the Fort Monmouth, New Jersey area primarily centered around small and medium size helicopter antenna design and an investigation of NVIS propagation characteristics. The aircraft antenna was a shorted loop type antenna, made out of 2.5cm aluminum tubing, 4.5m long, spaced 30cm from the fuselage and grounded at the rear of the helicopter near the vertical stablizer (see figure 1). This antenna exhibited good mechanical characteristics and provided the desired NVIS radiation characteristics, with as high an aircraft radiation efficiency as practical. Because it was of the current fed variety, it was not noticeably susceptible to detuning effects (e.g., rotor modulation). The antenna pattern which was predicted using computer modeling prediction techniques[7] was validated during the test program. It provided the desired high angle radiation pattern needed for NVIS propagation.

Communications range and field strength measurements were made. It is interesting to note that at a range of 35 km and with the proper choice of frequency for the particular time-of-day operation, the NVIS skywave signal was consistently 25 to 35 dB greater in field strength than that expected of a ground wave signal. Furthermore, the field strength at an extended range of 167 km was within 2 dB of that taken at 35 km. This uniform coverage is one of the most significant advantages in using this mode of transmission; however, it may be a disadvantage regarding undesired signal interception and susceptance to interference. The terrain independence characteristics of the NVIS mode were demonstrated in flight tests to the western and northern parts of New Jersey.

1.2.2 White Mountain Area of New Hampshire Tests: The New Hampshire National Guard (NHNG) provides a search and rescue service for vacationers in the rugged mountainous terrain in New Hampshire. Conventional LOS - limited radios do not work in this area and multiple relay stations were not practical. In response to a request to the Army for communications assistance, a NVIS system was provided to the NHNG. Based on frequency predictions and tests in this area, a reliable NVIS helicopter communication channel was established. A solid channel was established throughout all aircraft test maneuvers in many locations within the State, even with a very low power output ground based portable transmitter (20WPEP). The tests were conducted in daylight hours on a low noise, clear channel chosen from the best available allocation. The conclusions reached from these tests were that the NVIS channel is terrain independent and that relatively low radiated power output can be used to communicate during the day if the operating frequency is properly chosen.

1.2.3 Oahu, Hawaii Tests: Similar NVIS tests were performed in a different type rugged mountainous terrain as found in Oahu, Hawaii. Here more engineering test planning was used to assess the effectiveness of the HF NVIS channel as compared to that of the current LOS - limited tactical VHF-FM aircraft system and of a prototype Army aircraft satellite communications system.

The NVIS and satellite channels performed well over the test course. In the few areas where the aircraft-to-satellite channel was obstructed by the terrain the NVIS channel still provided satisfactory communications. The VHF-FM channel, on the other hand, provided very limited communications coverage, and it typically showed outages at ranges of 5 to 8 km.

The overall conclusion was that the HF-SSB system, operating in a NVIS mode, and UHF-FM (satellite) system both provided a solution to the NOE communications problem for operation over any part of the Hawaii NOE flight course terrain, or terrain similar thereto, during daylight hours, whereas VHF-FM is generally unsatisfactory for conducting a low flight mission over that type of terrain.

1.2.4 Fort Hood, Texas Tests:[8,9,10]

1.2.4.1 Test Program Purpose: A large scale combined operational and engineering test was conducted at Fort Hood, Texas, to quantitatively assess the performance and effectiveness of nine candidate radio systems (both the tactical VHF-FM currently fielded system and HF-SSB systems) and communication modes. The experiment design considered variables including range, altitude, terrain, time-of-day, frequency, and power that affect the radio channel signal to noise ratio (SNR). The tests were designed to determine how the performance of the non-LOS and LOS radio systems depended on these major variables. The test, conducted over a three month period in 1976, involved over 100 personnel, 1000 hours of flight testing, and utilized over 10,000 random alphanumeric (A-N) test messages to determine and evaluate quantatively the effectiveness of the voice channels of the radio systems.

1.2.4.2 Measure of Effectiveness: To evaluate the communications effectiveness of the HF channel, randomly selected alpha-numeric (A-N) characters were sent through the radio channel. Communication effectiveness was defined as the percent of the A-N characters sent one way without repeats through the communication channel which were correctly received. This measure provided a quantitative comparison of the channels tested as a function of the range and other test variables. (i.e., power output, time-of-day, link, range, altitude, terrain.) Communications effectiveness was the primary measure of effectiveness.

1.2.4.3 Test Message: A 30-character test message of randomly selected (and equally probable) letters and numbers were used. This was called an A-N test message. The A-N test messages were formatted and transmitted in a tactical spot report format by the tester. These messages operationally resemble target grid coordinates of a type that helicopter pilots routinely transmit over radio systems. Spot reports in this format sent one way through the channel without repeats were a demanding test of the communication channel. Finally, these A-N messages were recorded in the helicopter by a test observer and graded at the end of the mission. A word consists of six randomly selected A-N characters. In this message the characters and numbers were sent using the phonetic alphabet. Thirty random A-N characters were sent in groups of six A-N words per report; that is, "Spot report numbers 1 - AC90X4" would be transmitted as "Spot report number One, Alpha, Charlie, Niner, Zero, X-ray, Four". These messages were copied down on answer sheets, graded and used as the primary measure of effectiveness for the tests. This method of measuring communications effectiveness was used exclusively for the field tests performed at Fort Hood, Texas and Oahu, Hawaii.

1.2.4.4 Test Results Summary: The mean A-N intelligibility scores are summarized in Table 1 for the 25 and 50 km ranges. The communications effectiveness is shown by link (i.e., aircraft to aircraft (A-A), aircraft to ground station (A-G), and ground station to aircraft (G-A)) and time of day for the various HF equipment rated power outputs (400, 200, 100 and 40 WPEP). The times of day are defined: day (0900-1100 hr LT), dawn (0530-0730 hr LT) and night (0200-0400 hr LT). The frequencies used ranged from 2.240 to 4.370 MHz; and, once selected per the test plan, they were not changed for the duration of a test run. That is, if interference appeared on the test channel, or if HF conditions changed during the test run, the operational testor was not allowed to change to a better channel. The mean A-N scores shown; however, do not take into consideration A-N=0, or no communications link, scores. The mean includes only those scores where a communications link was established (A-N>0) to show channel performance. The zero scores (A-N=0) will be treated as part of the probability of successful communication discussion in section 2.3.2 of this paper. Table 1 shows the suitability of the NVIS mode for tactical communications and supports the conclusion reached in earlier tests in New Jersey, New Hampshire, and Hawaii.

2. NVIS CHANNEL:

NVIS propagation is dependent on the frequency of operation and the antenna characteristics (e.g., polarization and antenna directivity). The main difference between NVIS and long-range HF usage lies in the angle of the incident signal on the ionosphere and the resultant frequency band of operation. For NVIS the best angle of incidence is greater than 80 degrees and the frequency band is below 10 MHz. The ionospheric layers return the NVIS signal and create an umbrella type range coverage with no "skip zone" or "quiet zone" effects. Any ground wave present with the skywave signal results in wave interference effects. In the low altitude tests conducted at Fort Hood, the ground wave interference was constructive and enhanced signal performance at close in ranges. The tests conducted in the Fort Hood, Texas area showed ground wave enhancement out to about 20 km. In more rugged terrain (e.g., Oahu, Hawaii) the ground wave was noted as being important to ranges less than 10 km.

2.1 ANTENNA CONSIDERATIONS:

2.1.1 Ground Antenna: A horizontally-polarized antenna is preferrable for NVIS propagation. On short range circuits (0 to greater than 50 km) higher gain at radiation angles near the zenith is desired.[1,3,4,10,11] A ground based, unbalanced half-wave horizontal dipole, placed 10 to 12 meters above ground provides good results for NVIS. It is noted that an unbalanced antenna, as currently used by the tactical military forces, provides both horizontal and vertical components which enhances close-in signal performance. Table 2 shows the calculated gain of a dipole antenna at 10m elevation above good ground for a radiation angle of 85 degrees. For NVIS a ground based dipole can be expected to provide 4 to 5 dB of gain toward the zenith when it is elevated one-eighth wavelength to one-quarter wavelength above ground.

2.1.2 Aircraft Antenna: An HF shorted-loop antenna (as tested on the OH-58 and UH-1 aircraft) is near-optimum for all Army aircraft that will use the NVIS mode. This antenna radiates both a horizontally polarized skywave component toward the zenith, and also a vertically polarized groundwave component off the aircraft nose and tail. The pattern toward the zenith is essentially omnidirectional for ranges out to greater than 50 km. The groundwave pattern has maxima off the nose and tail of the aircraft, and nulls off the sides of the aircraft (but not a perfect figure-eight pattern). The shorted-loop antenna as shown in figure 1 has a predominately inductive characteristic when the operational frequency is less than a quarter wave. It can be matched to a 50-ohm impedance level by capacitive elements to achieve high efficiency in the coupling transition. Results of tests of the shorted loop installed on the electrically small OH-58 show an efficiency as shown in Table 3. These gain estimates are considered to be the best practical attainable efficiency on small aircraft. The antenna couples energy into the whole airframe to provide radiation at the lower frequencies. The amount of radiation is physically limited by the area of the loop and by the wavelength-airframe size ratio. The efficiency of about 3% (-15dB at 2.1 MHz) does not appear to be good; however, when compared to the estimated efficiency of present antennas on small aircraft of only 0.3% we note an order of magnitude improvement. The shorted-loop efficiency is adequate for communications when the "window" is open and noise and interference levels are normal, as will be shown later in this paper under the "System Margin" section.

2.2 FREQUENCY CONSIDERATION: The best operating frequency to use at a particular time is generally an interference-free frequency with the lowest propagation loss. The presence or absence of interference is difficult to predict, but, relatively reliable predictions of ionospheric propagation and atmospheric noise are available. An example of the output from a computer propagation prediction service program used during a field test to develop a frequency plan is shown in Table 4. This Table shows the statistical monthly reliability of certain designated test frequencies at two-hour intervals over a 24-hour period for operation over a specific 50-km path at Fort Hood. The Maximum Usable Frequency (MUF) and Frequency of Optimum Transmission (FOT) are also shown. In general, the trend for the best operating frequency, as shown during the month of November 1976, and during the daytime was 4 to 6.5 MHz, 2.7 to 3.5 MHz during the night, and 2.6 to 3.0 MHz just prior to sunrise. Experience has shown that the best quality circuit is attained when the operating frequency is close to the predicted FOT. However, for the Fort Hood test program, primary and alternate frequencies were selected for the three testing intervals of night, dawn and day. These frequencies were selected on the basis of a predicted reliability (of propagating) greater than 90 percent e.g., for 90 percent of the days in the month, the given frequency would propagate with a reliability of 90 percent or greater. The frequencies selected from Table 4 for the three test time periods were:

PERIOD	HOURS (LT)	PRIMARY (MHz)	SECONDARY (MHz)
Night	0200-0400	2.240	2.489
Dawn	0530-0730	2.240	2.489
Day	0900-1100	4.370	4.089

The experiment was not designed to permit selection of best operating frequency. If the best operating frequency is selected by a combination of propagation predictions as shown in Table 4, updated channel and traffic analysis information, a more effective communications channel would be available.

Figure 2 was derived from experimental station to station field strength data. It shows an average example plot of dB extra loss in the strength of a radiated signal versus choices of frequencies from the best choice operating frequency (f_o). It is recognized that the extra system loss may not always be symetrical as shown and that the slopes do vary depending on the environmental conditions;[12] however, this figure is included to illustrate the fact that an extra system loss can be expected when not operating on f_o. A variance of only ±1 MHz in the early morning period could mean an increase in systems loss of 10 dB. If real time ionospheric sounding is used the best choice frequency could be determined from empirical analysis. However, under tactical conditions it is not always possible to choose the best frequency, and the penality payed to overcome the additional losses is by the selection of higher power output and alternate modulation methods.

2.3 CHANNEL CONSIDERATIONS:

2.3.1 Relationship Between A-N Scores and Operational Channel Suitability: Table 5 is included to show the relationship between alphanumeric (A-N) score and operational suitability of the channel for aircraft voice communications. Based on laboratory tests, an A-N score of 90% can be achieved by pilots for a circuit merit 3 (CM3) channel. A CM4 channel in the laboratory degrades to a CM3 channel for a pilot flying at low altitudes primarily because of the aircraft man-machine environmental interface. The CM3 channel is equivalent to an FM channel with a 10dB SNR and an HF channel with a 6dB SNR. Operation below these SNRs usually requires disabling the radio squelch control. It is noted, that squelch disabling is often used to detect a message in the noise from a distant station; however, this results in a poor-quality circuit requiring many repeats to achieve a 100 percent score for a random A-N message.

2.3.2 Probability of Successful Communication Analysis: For the analysis of the test results at Fort Hood, a point estimate of the probability of successful communications (P_S) was computed from the multiple observations at each range by determining the fraction of the attempts when the A-N score was equal to or greater than 70 percent.[10] Zero A-N scores (A-N=0) were taken into consideration in determining P_S. Zero scores occurred and were recorded during the HF test primarily because of either severe channel interference or equipment malfunction and operator problems (e.g., improper squelch adjustment). Table 6 is a summary of the data. These results are considered realistic and typical for HF NVIS tactical operation with current frequency management techniques. From this table the following conclusions can be drawn:

a. The 400 W transmitter produced the highest P_S.

b. The P_S for both ranges considered are essentially the same.

c. The mean value of $\overline{P_S}$ shows that communications can be expected to be better during the day than at either the dawn or night time periods.

d. The speech processing added to the 100 W transmitter produces a higher P_S than the 200 W transmitter without speech processing.

e. Communications will be difficult during the dawn and night time periods with a 40 W transmitter.

A significantly higher channel reliability would have been achieved for the HF-SSB systems if other test frequencies had been used and if such frequency changes had been permitted during a given test period. If we assume that performance on single HF-SSB frequency is interference-limited, that the interference is narrowband, that two frequencies are available, and that interference on channel A and B are uncorrelated then:

P(A,B) = probability of successful communication on either channel A or B

P(A,B) = 1 - p (failure)

= 1 - p (both channels fail)

= 1 - p (channel A fails) X p (channel B fails).

For Example:

let P(A) = 0.67 (assumed)

P(B) = 0.80 (assumed)

P(A,B) = 1 - (1 - 0.67) (1 - 0.80)

= 1 - (.33) (.20)

= 0.93

In summary, two-frequency operation increases reliability of HF-SSB in the presence of narrowband interference (from .67 to .93) for the example and the assumptions given. This example points out the need for primary and an alternate channel assignments in lieu of a real time on-line frequency assignment management tool.

2.4 SYSTEM MARGIN.

2.4.1 Definition: The probability of successfully communicating at a distance for a given system, environment and operational deployment condition can be defined in terms of a system margin, M(d).[2,10] The system margin compares the measured or computed predetection signal-to-noise ratio (SNR) at a distance d, SNR(d), with the required receiver predetection SNR, R_r. The system margin M(d) is.

(1) $M(d) = SNR(d) - R_r$

For values of M(d), ≥ 0 an acceptable communications circuit is achieved. The actual values of M(d) show the relative circuit margin available.

The system margin M(d), in its expanded form is:[10]

(2) $M(d) = P_T - L_T + G_T - L_b - F_a - B - R_r + 176$ (dB)

These terms are defined in the following sub-section and in the Appendix. This equation assumes that the receiving antenna circuit losses, L_c (in dB), are approximated by 2dB minus the receiving antenna gain (in dBi).

2.4.2 Assumptions for Analysis: To analyze the expected system margin for a NVIS signal emanating from either a small helicopter or a fixed ground station the following general assumptions are made:

a. Frequency of Operation (f): f = 2MHz at 0700 hr LT

f = 4MHz at 1000 hr LT

b. Times of Day (TOD): Dawn = 0700 hr LT

Day = 1000 hr LT

c. Ranges: 25 km and 50 km

d. Average height of ionosphere: 300 km

e. Transmitter Power Output: P_T = 56 dBm (400W)

P_T = 53 dBm (200W)

P_T = 46 dBm (40W)

f. Transmission Line Loss: L_T = 1dB for ground

L_T = 0.5dB for aircraft

g. Antenna Gain G_T:

	2MHz (dBi)	4MHz (dBi)
Aircraft	-15	-10
Ground	+4.3	+4.9

h. Basic Transmission Path Loss L_b where $L_b = L_d + L_{TA}$:

TOD	f (MHz)	25 km (dB)			50 km (dB)		
		L_d	L_{TA}	L_b	L_d	L_{TA}	L_b
Dawn	2	94.5	5	99.5	94.6	5	99.6
Day	4	100.5	20	120.5	100.6	20	120.6

i. Antenna Environmental Noise Figure F_a:

F_a = 48dB Rural Noise at 4MHz

F_a = 65dB Atmospheric Noise limited at 2MHz

j. Bandwidth B = 10 log b:

B = 35dB for a 3000Hz voice channel

B = 25dB for a 300Hz CW channel

k. Required SNR, R_r = 6dB for a CM3 voice channel achieving A-N scores greater then 70 percent. (Fading not considered.) R_r = 0dB was chosen for CW in this example.

l. Terrain and aircraft low altitude are not considered.

2.4.3 Analysis: A summary of the analysis for the system margins for a signal eminating from either the aircraft transmitter (A/C XMTR) or the ground transmitter (GND XMTR) are shown in Tables 7, 8, 9 and 10. The Tables show the sensitivities of time of day and transmitter power output for the ranges of 25 km and 50 km. Tables 7 and 8 were derived for a voice channel with a typical HF-SSB bandwidth of 3000Hz. Tables 9 and 10 were included to show how the margin would be improved if the bandwidth was reduced to 300Hz for CW type operation.

2.4.4 Conclusions: From this analysis the following conclusions, which were born out during the testing program, are made:

a. Variable power output provides needed system margin especially for dawn and nighttime operations.

b. Transmission from the ground based antenna can provide more margin because of better gain.

c. CW operation could provide up to 16dB system margin over voice circuits because of bandwidth reduction and less susceptable to interference effects.

d. The margin at 25 km and 50 km is substantially the same.

e. Low power output (40W) voice operation at dawn and at night is difficult and will result in a noisy communications channel.

f. Low power output operation during the day should provide adequate margin if operating on the best frequency.

g. Improper choice of the best frequency for operation can be overcome by increased power output. This is a reason why conventional type prediction methods can provide an adequate frequency plan under normal conditions.

3. CONCLUSIONS:

An HF-SSB radio with modern features operating in the NVIS mode can be used successfully to provide satisfactory communications for low flying tactical aircraft over a 50 km (or greater) range in vituraly any type of terrain condition. By proper frequency selection, with the best results being obtained during daytime operations, lower power output can be used. At night time and during dawn, high power output is frequently required because of the presence of noise and interference in the channel. The frequency of operation for the NVIS mode is from 2MHz (radio equipment and aircraft antenna limitations) to about 10MHz (expected upper limit for NVIS frequency supportability). Frequencies selected near the FOT should be used. The predominant advantage of the NVIS mode is that once the fundamental path loss and ambient noise factors are overcome by the transmission system in a given geographical area, successful communications can be expected between two or more other points out to a radius of at least 50 to 100 km, with complete independance from terrain features. The predominant disadvanrage is the vulnerability to undesired interception and interference.

The channel quality of HF-SSB is inferior to that of the present VHF-FM tactical channel; however, it is operationally acceptable if used with an HF-SSB radio with modern features. We have shown that a channel with a CM3 rating can be used to reliably communicate A-N messages such as spot reports, targeting information, status reports, brevity codes, etc. However, if an aviator listens to this type channel (with CM3) for a period of time, pilot fatique will result due to the continuous background noise (and sometimes the interference) present in the channel. An optimum receiver squelch is required. A selective addressing squelch will be used to ensure positive communications contact. (The aviator will listen and reply only on a selected channel when being addressed.) This feature, coupled with frequency scanning of multiples preset channels, provides added system flexibility with simplicity of operation.

To take maximum advantage of the NVIS mode other factors must be considered in the modern HF-SSB radio. Single side band suppressed carrier modulation on either/or the upper (USB) and lower (LSB) side band with audio sylabic speech processing (about 3 to 4 dB) enhances the aviators talk power with minimum loss in speaker recognition. The choice of USB or LSB provides two channel assignments on the same frequency, and this choice can be used to minimize narrowband co-channel interference. Furthermore, the reliability of the HF-SSB radio channel will be increased further by assigning two (or more) frequencies, i.e., a primary and a secondary frequency, to support a communications net for a given time of day. This gives additional capability to avoid narrowband frequency-selective interference. The frequency scanning feature combined with selective addressing will accomodate this procedure.

On-line ionospheric sounding could provide the real-time selection of the best choice of frequency, the choice for the power output required resulting in the lowest power output being used for a transmitter with the lowest possible environmental interference effects, the highest possible channel circuit reliability, and efficient utilization of the crowded frequency spectrum. As an alternative to on-line ionospheric sounding, multiple frequency (channel) assignments which are based on prediction services information must be used. The user in this case either manually or automatically selects the desired frequency for operation. It is envisioned that the modern HF-SSB radio will be controlled by a micro-computor which could be programmed to provide full automatic frequency decision making and selection, if desired. Current plans include a limited manual capability.

The addition of a manual-keyed continuous wave (MCW) capability for the transmission of brevity codes will provide at least 10dB more system gain primarily because of a bandwidth reduction of 10 (3000 to 300 Hz). It is noted that with 10dB more system gain, the very low power output can be used effectively The radio will also have a data transmission capability for further communications enhancement with future appliques.

The results of this test program, supplemented by extensive analysis, were a contributing factor in the decision of the US Army to procure an HF-SSB system for use on tactical low-flying aircraft.

4. REFERENCES:

1. BRUNE, J.F.; REILLY, J.E., November 1975, "Compact HF Antenna (with Propagation Studies)", R&D Technical Report ECOM-4366, US Army Avionics Research and Development Activity, Ft. Monmouth, New Jersey.

2. HAGN, G.H., March 1979, "Radio System Performance Model for Predicting Communications Operational Ranges in Irregular Terrain", 29th IEEE Vehicular Technology Conference Record, Arlington, Illinois.

3. HAGN, G.H., July 1973, "On the Relative Response and Absolute Gain Toward the Zenith of HF Field-Expedient Antennas - Measured with an Ionospheric Sounder", IEEE Transactions Antennas and Propagation, Vol. AP-2, No. 4, pp 571-574.

4. HAGN, G.H.; VAN DER LAAN, J.E.; LYONS, D.J.; KREINBERG, E.M., September 1966, "Ionospheric-Sounder Measurement of Relative Gains and Bandwidths of Selected Field-Expedient Antennas for Skywave Propagation at Near-Vertical Incidences", Special Technical Report 18, prepared by Stanford Research Institute under Contract DA-36-039 AMC 00040(E), Order No. 5384-PM-63-91 for Advanced Research Projects Agency, ARPA Order No. 371, available from NTIS, Springfield, Virginia as AD 489-537.

5. HAGN, G.H.; VINCENT, W.R., May 1974, "Comments on the Performance of Selected Low-Power HF Radio Sets in the Tropics", IEEE Transactrons Vehicular Technology, Vol. VT-23, No. 2, pp 55-58.

6. HERBSTREIT, J.W.; CRICHLOW, W.Q., 1964, "Measurement of the Attenuation of Radio Signals by Jungles", Radio Science J. Res., National Bureau of Standards, Vol. 68D, No. 8, pp 930-906.

7. MEDGYESI-MITSCHANG, L.N., July 1976, "Prediction of HF Antenna Radiation Patterns", ECOM Report 75-0907-F, McDonnell-Douglas Research Laboratories, available from USA AVRADA, Ft. Monmouth, New Jersey.

8. PINSON, LTC J.M. et al, March 1977, "Nap-of-the Earth Communications (NOE COMM) System", Final Report FM-320 tests, RCS-ATCD-8 TCATA, TRADOC, Ft. Monroe, Virginia.

9. RICCIARDI, B.V.; TUPPER, B.C.; HAGN, G.H., October 1977, "Design of A Full-Scale Test for US Army Helicopter Nap-of-the-Earth (NOE) Communication Systems", ARO Report 78-2, Proceeding of the 23rd Conference on the Design of Experiments in Army Research Development and Testing, pp 427-438.

10. TUPPER, B.C.; HAGN, G.H., June 1978, "Nap-of-the-Earth (NOE) Communications for US Army Helicopters", Final Report Contract DAAB07-76-C-0868, AVRADCOM Technical Report 76-0868-F, SRI International, available from NTIS, Springfield, VA as AD 76-0868F.

11. HAGN, G.H.; VAN DER LAAN, August 1970, "Measured Relative Responses Toward the Zenith of Short-Whip Antennas on Vehicules at High Frequency", IEEE Transaction Vehicular Technology, Vol. VT-19, No. 3, pp 230-236.

12. SOICHER, H., February 1979, private communications.

5. ACKNOWLEDGEMENTS:

It is the authors pleasure to acknowledge the contributions of Messrs George H. Hagn and Bruce Tupper of SRI International, whom we had the distinct pleasure of working with for the duration of the NOE Communications System Test program. They provided the encouragement, guidance and technical expertise which formulated the basis for this paper. In addition, acknowledgement is made to Dr. Haim Soicher of the CENCOMS Propagation Research Team for his invaluable contributions in respect to results of the most current investigation of the ionospheric phenomenon and to Mr. George Lane of the US Army Communications Electronics Engineering Installation for his work in providing meaningful propagation prediction charts.

APPENDIX DEFINITIONS

P_T = transmitter output power (dBm)

L_T = insertion loss of transmission line for the transmitter (dB)

G_T = antenna gain for the transmitting antenna to include coupler losses (dBi)

L_b = basic transmission path loss (dB)

L_d = total NVIS free space distance loss (dB)

L_{TA} = turn around loss (dB)

F_a = environmental antenna noise figure (dB)

B = noise power bandwidth (dB)

R = required predetection SNR (dB)

176 = is a constant term derived from the noise figure ($n = 1\ k\ t_o\ b$) referred to the terminals of a loss-free antenna. Receiver antenna system considerations (dB) are also included.

MEAN ALPHANUMERIC (A-N) INTELLIGIBILITY SCORES

ALTITUDES 0m AND 10m COMBINED (NO FREQUENCY CHANGES PERMITTED)

TRANSMITTER POWER OUTPUT (w)	MODE PATH	25km RANGE (% A-N CORRECT)			50km RANGE (% A-N CORRECT)		
		DAY	DAWN	NIGHT	DAY	DAWN	NIGHT
400	A-A	98	95	94	82	93	97
400	A-G/G-A	96	93	96	99	93	96
200	A-A	96	85	82	82	87	*
200	A-G/G-A	96	89	87	80	75	96
100**	A-A	96	96	81	93	91	96
100**	A-G/G-A	92	94	90	95	91	86
40	A-A	99	*	90	*	80	75
40	A-G/G-A	85	91	98	94	85	87

SOURCE: TCATA FM-320 NOE COMM SYS TEST DATA

*Sample size too small, results inconclusive.
**Include 3 to 4dB syllabic speech processing.

TABLE 1

THEORETICAL NVIS GAIN OF DIPOLE

REFERENCE ISOTROPIC

10 METER HEIGHT - 85° INCIDENT ANGLE

GOOD CONDUCTIVITY EARTH

FREQ MHz	ABSOLUTE GAIN (dBi)
2	+4.3
3	+4.8
4	+4.9
5	+5.0
6	+4.9
7	+4.8
8	+4.5
9	+4.1
10	+3.6

TABLE 2

NVIS GAIN OF SHORTED LOOP
ON OH-58 AIRCRAFT

Frequency (MHz)	Absolute Gain (dBi)
2.1	-15
2.3	-12
2.6	-11
3.2	-8
4.0	-10
4.9	-8
6.0	-4

SOURCE: Ref 10, p 216

TABLE 3

US ARMY COMMUNICATIONS-ELECTRONICS ENGINEERING INSTALLATION AGENCY

FREQUENCY RELIABILITY TABLE

PROJECT 341 PPA 761.51.0 DCA
GROUND STATION 2 TO FORT HOOD, TX
31.48N - 97.87W 31.03N - 97.87W
TYPE OF SERVICE 3A3A
XMTR 2+30 HORZ HW DIPOLE (H 9.14)
RCVR 2+30 HORZ HW DIPOLE (H 9.14)
POWER = .200KW 3MHz MAN-MANDE NOISE = -166DBW

SSN 12.1 NOVEMBER 1976

AZIMUTHS	MILES	KM.
180.00 .00	31.1	50.0

MINIMUM ANGLE = .0 DEGREES
(L -.50) (A -0) OFF AZ 0
(L -.50) (A -0) OFF 0
REQD. S/N = 50DB

FREQUENCIES IN MHZ

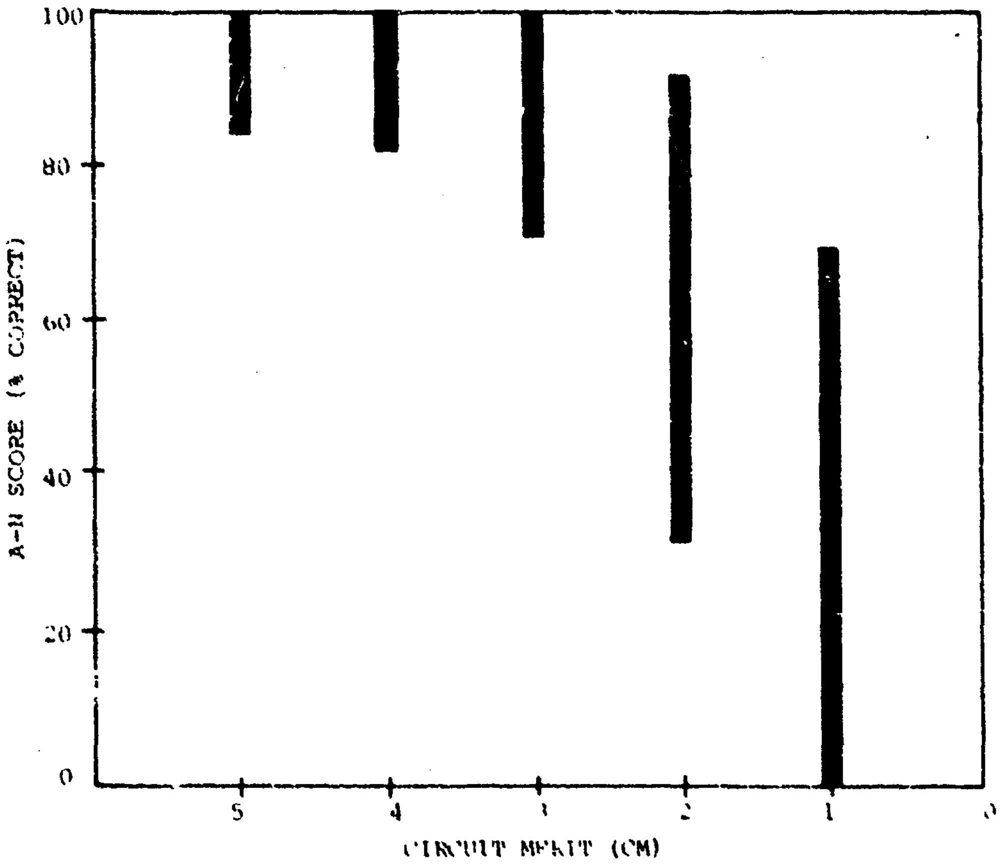
FT HOOD LT	2.2	2.5	3.3	4.1	4.4	5.1	5.8	6.1	6.5	7.4	7.7	
20	.92	.90	.62	.21	.12	-	-	-	-	-	-	REL.
22	.90	.89	.63	.22	.13	-	-	-	-	-	-	REL.
29	.86	.88	.76	.33	.23	.07	-	-	-	-	-	REL.
02	.85	.87	.88	.59	.44	.18	.07	.04	-	-	-	REL.
04	.85	.84	.69	.34	.24	.07	-	-	-	-	-	REL.
06	.92	.90	.68	.30	.21	.06	-	-	-	-	-	REL.
08	.98	.98	.99	.99	.99	.93	.68	.49	.28	.04	-	REL.
10	.99	.99	.99	.99	.99	.99	.98	.95	.88	.54	.42	REL.
12	.99	.99	.99	.99	.99	.98	.95	.92	.85	.59	.49	REL.
14	.99	.99	.99	.99	.99	.98	.95	.92	.86	.60	.51	REL.
16	.99	.99	.99	.99	.99	.98	.92	.85	.71	.29	.20	REL.
18	.96	.96	.98	.89	.79	.33	.05	-	-	-	-	REL.

FT HOOD LT	18	20	22	24	02	04	06	08	10	12	14	16
MUF	4.8	3.5	3.5	[illegible]	4.3	3.9	3.7	6.1	7.5	7.6	7.7	6.9
FOT	4.1	2.7	2.8	3.1	3.5	2.8	[illegible]	5.2	6.4	6.3	6.3	5.9

DASHES IN RELIABILITY LINES SIGNIFY RELIABILITIES OF 00 PERCENT

TABLE 4

RELATIONSHIP BETWEEN ALPHANUMERIC (A-N) SCORE
AND OPERATIONAL SUITABILITY OF CHANNEL FOR AIRCRAFT COMMUNICATIONS



5. Excellent. First-time reliability. High-quality circuit. Excellent channel.

4. Good. Good-quality circuit; small number of repeats. Good channel.

3. **Marginal**. Poor-quality circuits; frequent repeats. Between A-N = 70-30, a marginal channel exists, which is adequate for communication.

2. Unacceptable. Poor-quality circuit; many repeats required. Background-noise limited. Squelch disable operation.

1. Unacceptable.

0. Unacceptable. No channel

TABLE 5.

OBSERVED PROBABILITY OF COMMUNICATIONS SUCCESS, P_s,
FOR A-N TEST SCORES ≥ 70 PERCENT
(No Frequency Changes Permitted)
Altitudes 0m and 10m Data Combined

TRANSMITTER POWER OUTPUT (w)	MODE PATH	25 km RANGE P_s			50 km RANGE P_s		
		DAY	DAWN	NIGHT	DAY	DAWN	NIGHT
400	A-A	1.00	.92	.87	.69	.82	.92
400	A-G/G-A	.96	.93	.96	1.00	.93	1.00
200	A-A	.87	.80	.74	.73	.62	*
200	A-G/G-A	.97	.94	.72	.97	.72	.58
100**	A-A	.98	.79	.79	.96	.79	.73
100**	A-G/G-A	.89	.90	.88	.97	.64	.73
40	A-A	.95	*	.85	*	.33	.50
40	A-G/G-A	.87	.59	.73	.90	.31	.38
$\overline{P}_s$ (MEAN VALUE)		.94	.84	.82	.89	.65	.69

Source: TCATA FN-320 NOE COMM Sys TEST DATA.

*Sample size too small, results inconclusive
**Includes 3 to 4dB syllabic speech processing

TABLE 6

VOICE CHANNEL SYSTEM MARGIN - M(d)
d=25 km

P_T (W)	M(25)-A/C XMTR (dB)		M(25)-GND XMTR (dB)	
	DAY	DAWN	DAY	DAWN
400	12.0	6.0	26.4	24.8
200	9.0	3.0	23.4	21.8
40	2.0	-4.0	16.4	14.8

TABLE 7

VOICE CHANNEL SYSTEM MARGIN - M(d)
d=50 km

P_T (W)	M(50)-A/C XMTR (dB)		M(50)-GND XMTR (dB)	
	DAY	DAWN	DAY	DAWN
400	11.9	5.9	26.3	24.7
200	8.9	2.9	23.3	21.7
40	1.9	-4.1	16.3	14.7

TABLE 8

CW CHANNEL SYSTEM MARGIN - M(d)
d=25 km

P_T (W)	M(25)-A/C XMTR (dB)		M(25)-GND XMTR (dB)	
	DAY	DAWN	DAY	DAWN
400	28.0	22.0	42.4	40.8
200	24.0	19.0	39.4	37.8
40	18.0	12.0	32.4	30.8

TABLE 9

CW CHANNEL SYSTEM MARGIN - M(d)
d=50 km

P_T (W)	M(50)-A/C XMTR (dB)		M(50)-GND XMTR (dB)	
	DAY	DAWN	DAY	DAWN
400	27.9	21.9	42.3	40.7
200	24.9	18.9	39.3	37.7
40	17.9	11.9	32.3	30.7

TABLE 10

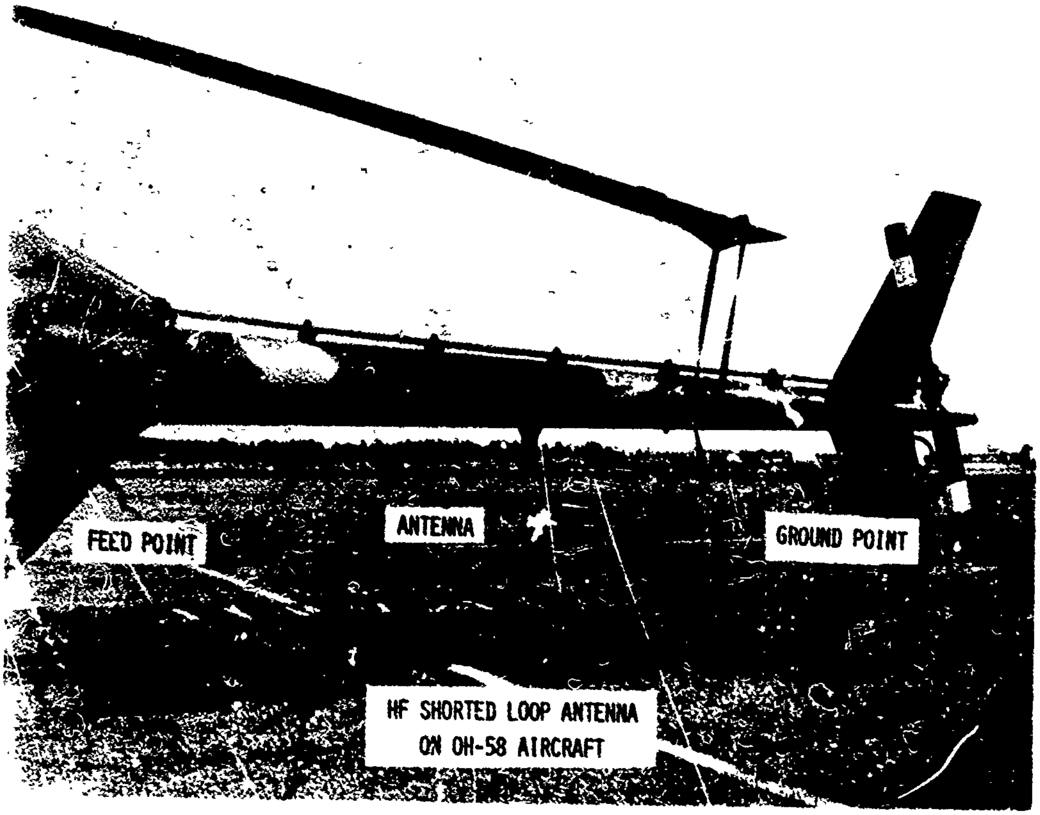

FEED POINT
ANTENNA
GROUND POINT
HF SHORTED LOOP ANTENNA
ON OH-58 AIRCRAFT

Figure 1

EXTRA EXPECTED LOSS WHEN OPERATING AT A FREQUENCY
CLOSE TO THE BEST CHOICE FREQUENCY (f_o)

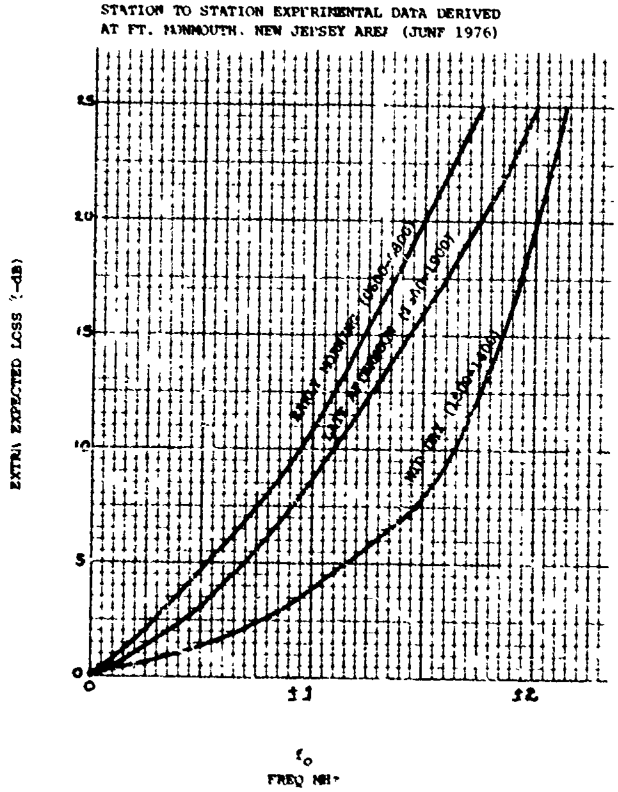


Figure 2

DISCUSSION

E.R. Schmerling, US

Voice bandwidths and ionospheric communications need careful frequency management. Tactically, this can be very difficult. The moment that narrow band (teletype or equivalent) is allowed, satellite communications at very high frequencies using non-directional antenna provide reliable communications and avoid the frequency management problem.

Author's Reply

Satellite communications to US Army aircraft provides a good quality communications channel. (See paper ref. 10). However, there are many key tactical limitations that preclude present consideration by Army Aviation. They are: very limited band of UHF frequencies assigned to Army for priority traffic; no tactical UHF ground equipment; aircraft directional antennas are too bulky and too heavy; satellites provide a prime modal communications system which is highly undesirable for tactical purposes; and availability for tactical operations is poor.

M. Soicher, US

You mentioned the importance of frequency management in circuit reliability? How were the frequencies used chosen at the various test sites? How did you assess ionospheric conditions at time of test?

Author's Reply

From the computer generated frequency predictions (sample given in paper) a communications plan was established with at least two frequencies (primary and alternate) being chosen from the assigned test frequencies. They were chosen based on the best predicted probability of success. However, prior to any test run a station to station "on-the-air" sounding was made to determine the better of the two channels. Therefore, a combination of a-priori knowledge of the best frequency of the assigned frequencies (via prediction services) and on-line limited type sounding (via listening and talking over the channels) made the plans workable.

Prof. Shearman, UK

The use of ground wave propagation is proposed as giving an alternative to NVIS of particular value at the higher frequencies. Is the antenna proposed suitable for this role? Ground wave loss is lower at low frequencies and, in the daytime, noise level is low at the low frequencies, so the proposal to use high frequencies is surprising.

Author's Reply

To answer the question let me define the terms used which may be misleading:

(a) NVIS propagation mode - Band of HF frequencies (2 to about 10Mhz) where we have high angle radiation ionospheric supportability. This produces terrain independant coverage. Any ground wave generated is of secondary nature and tends to enhance short range communications. The skywave signal is what is used for communications.

(b) Ground wave propagation mode - Band of HF frequencies where we do not have NVIS supportability (about 10Mhz to 30Mhz). Here we use low angle radiation to produce extended line-of-sight coverage which is terrain dependent. The proposed antenna is suitable, but not optimum, for this purpose. The use of HF-SSB in this particular mode must be compared to that of the tactical VHF-FM (30 to 88Mhz) band to fully understand the advantages to be gained for array operations. (Reference 10 of this paper provides a good source).

T. Damboldt, Ge

Did you make a record of the cases when communications were disturbed by unintentional interference from other stations using the same frequency? I have the feeling that this interference may sometimes be the limiting factor rather than atmospheric or man-made noise.

Author's Reply

During the test program limited records were kept on unintentional interference from stations using the same frequency. The main disadvantage of the NVIS communications mode is in the area of non-directional interference. Your assumption is correct; however, it is noted that once atmospheric and/or man-made noise is overcome an acceptable channel as described in this paper can be established (i.e. 6dBSNR, A-N 70%, with repeats required to achieve 100%). To overcome the noise high transmitter power output is required together with the ability to change frequencies rapidly.

REAL-TIME ADAPTIVE HF FREQUENCY MANAGEMENT

Dr. Robert B. Fenwick
BR Communications
Post Office Box 61989
Sunnyvale, CA 94088 USA

Mr. Terence J. Woodhouse
Tactical Communications Area (AFCS)
Langley AFB, VA 23665 USA

SUMMARY

In 1972 the United States Air Force began a program aimed at improving tactical HF communications. The approach involved (1) measurement in real time of the important unknowns: propagation, noise, and spectrum occupancy, and (2) adapting operating frequencies in real-time to the conditions measured. A potential problem with this approach is causing harmful interference to other spectrum users.

An exercise called TROPHY DASH III was conducted to assess the ability of a particular technique to permit real-time selection of frequencies while yielding acceptably low interference. TROPHY DASH III results are viewed as significant. First, in spite of extensive notification of other spectrum users, negligible interference was reported.

Based upon these results, real-time adaptive sharing of the HF spectrum in the tactical theater appeared both valid and feasible for a limited number of high-priority users. and procurement of the AN/TRQ-35(V) Tactical Frequency Management System was authorized.

1. INTRODUCTION

Vigorous efforts to obtain highly efficient spectrum sharing are vital at HF given its unique combination of widely varying propagation and noise conditions, extremely limited bandwidth, and far beyond line-of-sight range. Yet in many respects spectrum sharing techniques have differed little from those used for other parts of the radio spectrum where propagation, noise, and occupancy can be predicted with comparatively high accuracy.

Satisfactory HF communications can be obtained by the largest number of users only if real-time, adaptive frequency management is employed. Technology now makes possible effective real-time, adaptive frequency management to those users who can justify its cost. Certain military communicators have a mission of sufficient priority to justify this expense. While these military communicators can undoubtedly optimize their own communications, the problem exists of how they can choose frequencies while at the same time minimize harmful interference to other spectrum users. The new AN/TRQ-35(V) Tactical Frequency Management System (TFMS) has been designed to help reach this goal, and the concepts embodied in the AN/TRQ-35(V) have been tested in an exercise called TROPHY DASH III.

2. DESIGN OF THE TROPHY DASH III EXERCISE

2.1 Background

Historically, U.S. military exercises have been preceded by assignment of the available discrete frequency assets, by link, according to circuit or net priority. This system lacks flexibility, however, in that there is no provision, nor in most cases any capability, for modifying those assignments in the field after deployment. Since propagation and interference conditions change constantly, there is no guarantee under this rigid system that the most important links will always have the best available frequencies. To make possible this kind of guarantee, two pieces of information are essential: real-time propagation information and current, as well as historical, channel occupancy information. To obtain this information, the AN/TRQ-35(V) Tactical Frequency Management System has been developed.

The intent of the design of the AN/TRQ-35(V) is to provide as complete information as is practical to obtain on optimum frequency selections and at the same time minimize interference to other spectrum users. Relevant factors measured, and the techniques employed, in the AN/TRQ-35(V), are:

Propagation - "Chirpsounder" oblique sounder
Noise - "Spectrum Monitor"
Occupancy - "Spectrum Monitor" receiver with channel-occupancy memory

That the Chirpsounder can provide an adequate measure of propagation conditions is widely accepted. However, whether or not a "Spectrum Monitor" can determine satisfactorily if a frequency is "in use" is open to question. To provide an answer to this question, the TROPHY DASH III exercise was conducted in early 1975.

Ideally it would be desirable from the military communicators' point of view to place no frequency restrictions on him. This is unrealistic, however, since under such conditions harmful interference would undoubtedly occur, particularly on the usually-quiet emergency channels and on those channels where low-power communications take place -- especially portable/mobile. The frequency assignment process used in the TROPHY DASH III exercise was first to assume that ALL frequencies were available for military use (2-30 MHz) -- then eliminate frequencies or bands of frequencies where harmful interference was judged most likely to occur.

To determine what frequencies were not to be used, all U.S. government agencies were given the opportunity to request specific deletions. At the end of this process, 70% of the 2-30 MHz spectrum was authorized to TROPHY DASH III for a one-time test. This available spectrum primarily comprised International Broadcasting, Fixed, and parts of the Maritime Mobile Bands.

2.2 Objectives

Objectives of the TROPHY DASH III exercise were: (1) an assessment of the feasibility of using "band" assignments, rather than specific frequencies, in the tactical theater as determined by the number and type of complaints received from other spectrum users and (2) to arrive at final specifications for a TFMS Spectrum Monitor.

2.3 Exercise Configuration

The geographical configuration of the test is shown in Figure 1, which depicts a simulated Tactical Air Control System (TACS). MacDill AFB was the location of the simulated Control Center; the other locations communicated only with the Control Center.

A Chirpsounder receiver and Spectrum Monitor were located at MacDill AFB, while Chirpsounder transmitters were deployed at the other locations. The widely varying path lengths were selected to enable assessment of the Spectrum Monitor's ability to determine occupancy at varying distances.

Figure 2 shows block diagrams of the equipment used at the various test locations. Full-duplex, ISB communications, 1 kW or 2.5 kW, took place on all three paths, 6A9b emission throughout the test, using intermittent voice on one sideband and continuous 16-tone multiplex teleprinter signals on the other sideband.

2.4 Exercise Schedule

The exercise was conducted 24 hours per day, for five days, in January/February 1975. The schedule called for changes of both frequencies on each duplex link every 30 minutes in order to maximize spectrum usage. In addition, it was felt that some spectrum users might not report intermittent voice interference. Hence voice and teleprinter sidebands were interchanged 15 minutes into the half-hour spent on each frequency. A voice identification statement was transmitted at least once every 15 minutes.

2.5 TFMS Characteristics and Operating Procedures

During the test period, TFMS personnel monitored the propagation and interference conditions over the three HF links across the bands of frequencies authorized for test use. Operating frequencies were chosen within these bands on the basis of propagation quality and freedom from interference at the test network terminals, in order to optimize communications over the test links. Complete logs of the operation were kept.

Chirpsounder information was used to maintain the highest possible signal-to-noise ratio on each path. As much as possible MacDill transmit frequencies were kept below the corresponding receive frequencies on each path to ensure that "MUF failure" did not result in the loss of frequency control from MacDill. Figure 3 illustrates the Chirpsounder transmitters and Figure 4 the receiver in their final AN/TRQ-35(V) package. Basic specifications are given in Tables I and II, respectively, and sample record is shown in Figure 5, illustrating the capability of the Chirpsounder to separately measure mode structure (i.e., line-of-sight, surface-wave, sporadic E, E-layer, F-layer, and multiple-hop modes) and received power at all frequencies within either 2-16 or 2-30 MHz.

Spectrum usage information was provided by a specially built "breadboard" Spectrum Monitor which was designed to measure spectrum occupancy in the 2-20 MHz frequency range (production version 2-30 MHz). This frequency range was divided into 6000 "channels" of 6 kHz width (typical occupancy of the AN/TSC-60(V) communications systems in ISB mode) and 3 kHz spacing of center frequency. The TD III spectrum monitor measured the amplitude of signals in each of the 6000 channels each 12 seconds, at which time it was determined which of four adjustable amplitude thresholds had been crossed. Statistics of threshold crossings up to 30 minutes old were compiled by a minicomputer and displayed on a CRT having a TV-raster format.

The Spectrum Monitor of the AN/TRQ-35(V) does not differ significantly in important specifications from the TROPHY DASH III "breadboard" other than for expanded frequency coverage. The R-2093 Spectrum Monitor is shown in Figure 6, and basic specifications are given in Table III. Figure 7 shows examples of the displays and data obtained during TROPHY DASH III, which depict percent-of-time each amplitude threshold has been exceeded in a 30-minute period. The lowest threshold is typically 3 dB above atmospheric noise. Display widths of 100 kHz and 500 kHz are available, as shown.

TABLE I

BASIC ELECTRICAL SPECIFICATIONS OF T-1373/TRQ-35(V) CHIRPSOUNDER TRANSMITTER

Frequency Range	2-16 or 2-30 MHz
Output Waveform	linear FM-CW
Sweep Rates	50 kHz/sec (2-16 MHz) 100 kHz/sec (2-30 MHz)
Transmitting Times	selectable transmit/no transmit each 5 minutes in 1 hour
Output Power	0.2W or 2W antenna diplexed; 10W or 100W direct into antenna
Noise and Spurious (non-harmonic)	>55 dB below fundamental measured in 10 kHz bandwidth >150 kHz from fundamental
Harmonics	>60 dB below fundamental
Frequency Blanking	16 frequency bands 30 kHz wide may be programmed as no transmit
Self Test	test of synthesizer lock and standby battery supply

TABLE II

BASIC ELECTRICAL SPECIFICATIONS OF R-2081/TRQ-35(V) CHIRPSOUNDER RECEIVER

Frequency Range	2-16 or 2-30 MHz
Input Waveform and Sweep Rates/Format	compatible with transmitter
Number of Transmitters Received	can be programmed to automatically receive and store data from up to three Chirpsounder transmitters
Time Delay Window	10 msec (2-16 MHz); 5 msec (2-30 MHz)
Time Delay Resolution	100 μsec nominal (2-16 MHz)
Time Synchronization with Transmitter	automatic if sweep started within ± 1 sec of the transmitter
Propagation Data Displays	(1) RF signal level vs. frequency over 60 dB range above -110 dBm (2) time delay vs. frequency stored for up to 3 paths
Self Test	test of receiver sensitivity, synthesizer lock, audio spectrum analyzer, CRT display and standby battery supply

TABLE III

BASIC ELECTRICAL SPECIFICATIONS OF R-2093/TRQ-35(V) SPECTRUM MONITOR

Frequency Range	2-30 MHz
Number of Frequencies ("Channels") Analyzed	9333
Channel Spacing	3 kHz
Analysis Bandwidth (3 dB)	6 kHz
Analysis Filter Shape Factor (60:3 dB)	2.0 maximum
Type of Analysis	analyzes frequency of occurrence of crossings of 4 received-power thresholds spaced by 10 dB ± 1 dB
Data Memories	last 10 sec, current 5 minutes, last complete 5 minutes, last 30 minutes
Threshold Adjustment Range	lowest threshold automatically adjusted from -110 dBm to -62 dBm in 3-dB steps
Lowest Threshold Level	in most sensitive setting lowest threshold is typically 3 dB above background noise level
Period Between Amplitude Samples, per Channel	11 sec nominal
Channel Sampling Interval	3.2 msec
Spurious Responses	response to two -30 dBm signals <-110 dBm outside ± 150 kHz of each signal; response to one -40 dBm signal <-110 dBm outside ± 30 kHz of signal
Monitor Modes	audio monitoring of USB, LSB, AM, FM
Monitor Tuning	± 3 kHz from indicated center frequency
Display Type	Bar graphs of refreshed 5" CRT (TV raster display, refresh rate >50/sec)
Display Formats	"CURRENT" giving most recent thresholds crossed; updated each 10 seconds "CURRENT 5 MIN" giving histogram for each threshold during current 5 minute block "LAST 5 MIN" giving percent of time threshold crossed in last 5 min block - the final values in "CURRENT 5 MIN" "LAST 30 MIN" gives percent of time each threshold crossed in last 30 min, weighted to favor more recent data; rise and decay approximates exponential with 7 min time constant
Display Frequency Range	Selectable between nominal 100 kHz (33 channels) and nominal 500 kHz (167 channels)
Display Indicators	Digital indication of center frequency on CRT; signal threshold in dBm of CURRENT scan; digital indication of faults found in memory and receiver tests
Computer Fault	Red switch light indicating computer fault
Self Test	test of receiver sensitivity, memory, and battery supply

The procedure used at the time of each frequency change (which occurred every 5 minutes throughout the test -- 6 links, each changed every 30 minutes) was as follows:

a. Determine from Chirpsounder data, duplex separation requirements, and "MacDill frequency lower" requirement in which 500 kHz band the assignment is required. Figure 5 shows a sample sounder display which was used for this purpose. Note that the display is in two parts: the upper portion depicts received signal strength and the lower shows propagation mode information. Operating frequencies were chosen on the basis of maximum amplitude (first priority) and minimum multipath, with a bias toward higher frequencies to maximize signal-to-noise ratio.

b. Set center frequency of spectrum monitor to center of 500 kHz band and call up CURRENT 5 MIN display for lowest threshold.

c. Identify frequency or frequencies where threshold is exceeded a small fraction of the time (say 5% or less).

d. Verify that candidate frequency is within assigned band. If not, find another which is, using above procedure.

e. Switch to 100 kHz display and center candidate frequency display.

f. Switch display to LAST 5 MIN and verify no occupancy in this period.

g. Switch display to 30 MIN and verify no occupancy in this period.

h. Switch display to CURRENT and verify no present occupancy.

i. Request operator at distant end of circuit to verify no occupancy on candidate channel, using his spare receiver of the AN/TSC-60(V).

j. If favorable report is received from distant operator, change to new frequency.

2.6 Effectiveness Criteria

Several indicators were considered as test success criteria. First, of course, communications of high quality should be obtained at all times. Second, the number of complaints received from other spectrum users was considered an indication of the viability of the spectrum sharing concept. All agencies on the U.S. Interdepartment Radio Advisory Committee were advised of the test and several gave special notification to their subordinate units. Some of the U.S. Government spectrum users initiated special frequency watch procedures in order to provide contact reports to the test team in the event one of their frequencies was used. A special telephone number, directly into the TFMS van and manned continuously by test team personnel, was made known to all IRAC members in order to facilitate the interference reporting procedure.

3. RESULTS AND CONCLUSIONS

Figure 8 shows examples of ionograms obtained during the test on the three paths. Note at the top of Figure 8 the change in Maximum Observed Frequency from 9.3 MHz at 0300 to approximately 13 MHz an hour and 15 minutes later, a change of almost 4 MHz. The predicted MUF (obtained from standard prediction techniques) changed less than 400 kHz during this period.

Figure 8c illustrates an occurrence encountered frequently during TROPHY DASH III. The sporadic-E propagation shown in Figure 8c made possible transmission at much higher frequencies than predicted for that time of day -- up to 12 MHz could have been used when the predicted MUF was only 4.4 MHz. In contrast, Figure 8d shows F layer propagation which was almost one MHz (nearly 25%) below the predicted MUF for that time of day. Figure 8e again shows propagation via sporadic E of frequencies in excess of 8 MHz (the predicted MUF was 4 MHz). Note the contrast with normal daytime conditions in Figure 8f.

All of these illustrations have two things in common: they document propagation conditions which were typical during the test, and the real-time information shown differs substantially from the predicted values. Without the real-time information provided by the sounder these propagation conditions would never have been fully exploited in support of communications, and optimum spectrum usage would not have been possible.

At no time during TROPHY DASH III was there a circuit outage due to propagation or interference. Signal-to-noise ratios were very high at all times. Also, at no time was interference from other spectrum users experienced on any channel.

It was concluded that use of a single Spectrum Monitor is satisfactory in a theater of diameter of a few hundred km, where operation is conducted below the vertical-incidence critical frequency, but for paths outside this area (e.g., the 1760 km Altus path of the test) spectrum monitoring should be performed at both ends.

During the test, 1049 frequency changes were made using 745 discrete center frequencies. In spite of this expanded use of normally unavailable spectrum, only two valid interference reports were received -- both involving receiving stations monitoring quiet channels. In addition, two invalid interference reports were received (i.e., the interference was being caused by someone else). Based upon these results it has been concluded that a spectrum sharing scheme similar to that employed on TROPHY DASH III would be feasible in the tactical theater for units equipped with a TFMS.

During TROPHY DASH III, a 24 hour study was undertaken to determine the availability of assigned tactical training frequencies. A total of 56 HF frequencies were assigned to USAF Tactical Air Command (TAC) which could have been used by the test team without the special U.S. government authorization.

Each hour, the Chirpsounder was used to determine the limits of the propagation window. Then, all of the TAC frequencies that fell within this window were investigated using the spectrum monitor to determine how many of them were unoccupied, and were therefore available for use on a non-interference basis. The results are illustrated by Figure 9. During the 24-hour period, there were six hours when none of the 56 authorized frequencies were available to Altus. A similar condition existed for six hours to Robins and for seven hours to Homestead. For three of the 24 hours, no frequencies were available interference-free to any of the sites. During an actual exercise where only TAC frequencies were available the situation would have been far worse, as the average exercise has 6 to 10 duplex HF links to support with the same number of frequencies, instead of only three as operated on TROPHY DASH III. It can therefore be concluded that the seemingly generous frequency assignments normally available to support this kind of exercise operation -- even when as few as three links are used -- are inadequate if harmful interference is to be avoided.

From the results of TROPHY DASH III it can be concluded that the concept of sharing the HF spectrum in a manner similar to that used during the test, is not only feasible but is necessary and viable in the tactical environment. The test demonstrated that in spite of heightened awareness on the part of the government sector of HF users, this sharing of frequencies can be accomplished with a minimum of interference. The concept of this type of sharing is very attractive in that it makes more efficient use of a scarce global resource.

Aspects which need to be defined in follow-on testing programs include sharing limitations, the possible need for automation (expanded effectiveness vs cost), the ramifications of operating more than one TFMS in a theater and the interfaces required by such a configuration, the practical security advantages of frequency changes made possible by the TFMS, and the possible ECM/ECCM applications of the Spectrum Surveillance Monitor. It appears that implementation of the TFMS sharing concept is one means of achieving at the theater level the U.S. policy of more effective management of the frequency resource.

4. SUBSEQUENT OPERATION OF AN/TRQ-35(V)

Since the completion of TROPHY DASH III the AN/TRQ-35(V) has become a part of the USAF inventory and has been used with great success by all of the four U.S. military services. Applications have included ground-ground, air-ground, ship-shore, and ship-ship communications. Figures 10-15 show typical data obtained by the AN/TRQ-35(V) in these various applications.

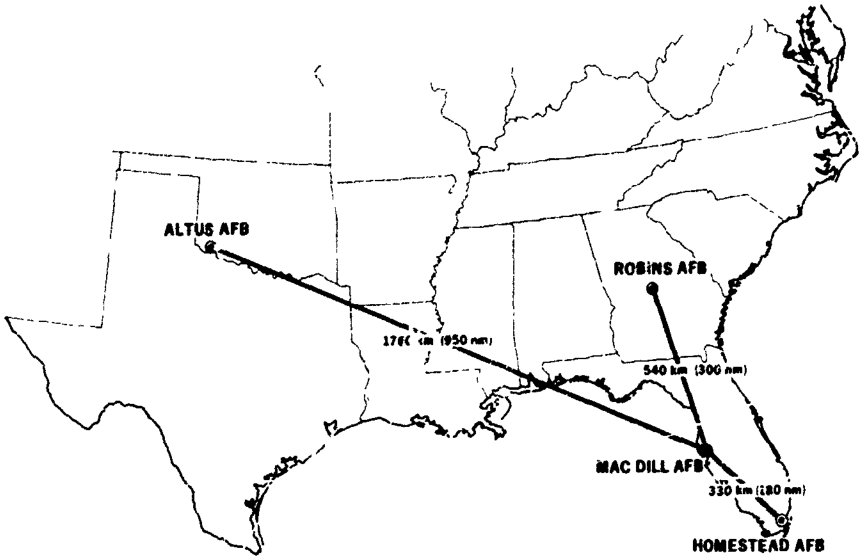
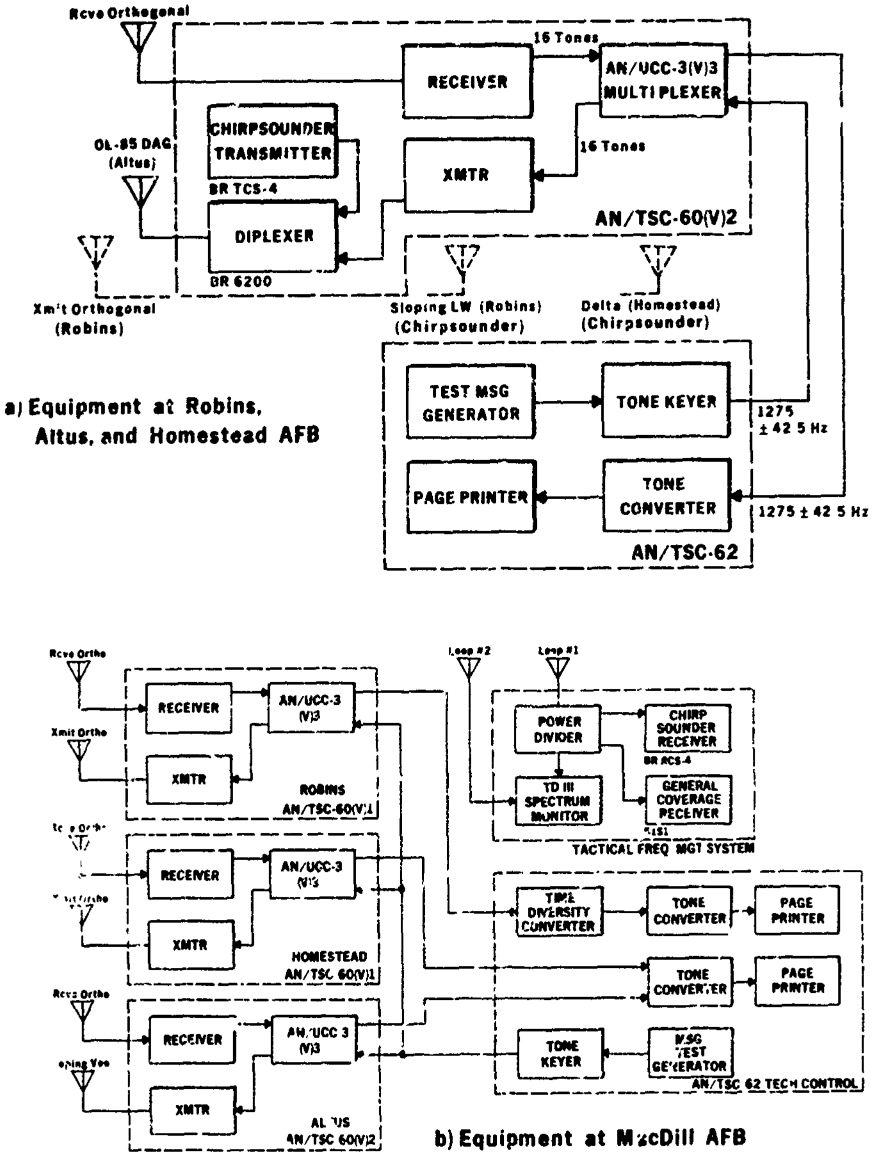

ALTUS AFB
ROBINS AFB
1760 km (950 nm)
540 km (300 nm)
MAC DILL AFB
330 km (180 nm)
HOMESTEAD AFB

Fig.1 Geometry of TROPHY DASH III Exercise. TFMS receiving hardware was located at MacDill AFB


Rcve Orthogonal
RECEIVER
16 Tones
AN/UCC-3(V)3
MULTIPLEXER
CHIRPSOUNDER
TRANSMITTER
BR TCS-4
OL-85 DAG
(Altus)
16 Tones
XMTR
DIPLEXER
BR 6200
AN/TSC-60(V)2
Xm't Orthogonal
(Robins)
Sloping LW (Robins)
(Chirpsounder)
Delta (Homestead)
(Chirpsounder)
TEST MSG
GENERATOR
TONE KEYER
1275
± 42.5 Hz
PAGE PRINTER
TONE
CONVERTER
1275 ± 42.5 Hz
AN/TSC-62
Rcve Ortho
RECEIVER
AN/UCC-3
(V)3
Xmit Ortho
XMTR
ROBINS
AN/TSC-60(V)1
Loop #2
Loop #1
POWER
DIVIDER
CHIRP
SOUNDER
RECEIVER
BR RCS-4
TD III
SPECTRUM
MONITOR
GENERAL
COVERAGE
RECEIVER
TACTICAL FREQ MGT SYSTEM
RECEIVER
AN/UCC-3
(V)3
XMTR
HOMESTEAD
AN/TSC 60(V)1
TIME
DIVERSITY
CONVERTER
TONE
CONVERTER
PAGE
PRINTER
TONE
CONVERTER
PAGE
PRINTER
Rcve Ortho
RECEIVER
AN/UCC 3
(V)3
XMTR
ALTUS
AN/TSC 60(V)2
TONE
KEYER
MSG
TEST
GENERATOR
AN/TSC 62 TECH CONTROL

a) Equipment at Robins, Altus, and Homestead AFB

b) Equipment at MacDill AFB

Fig.2 Block diagram of equipment used during exercise

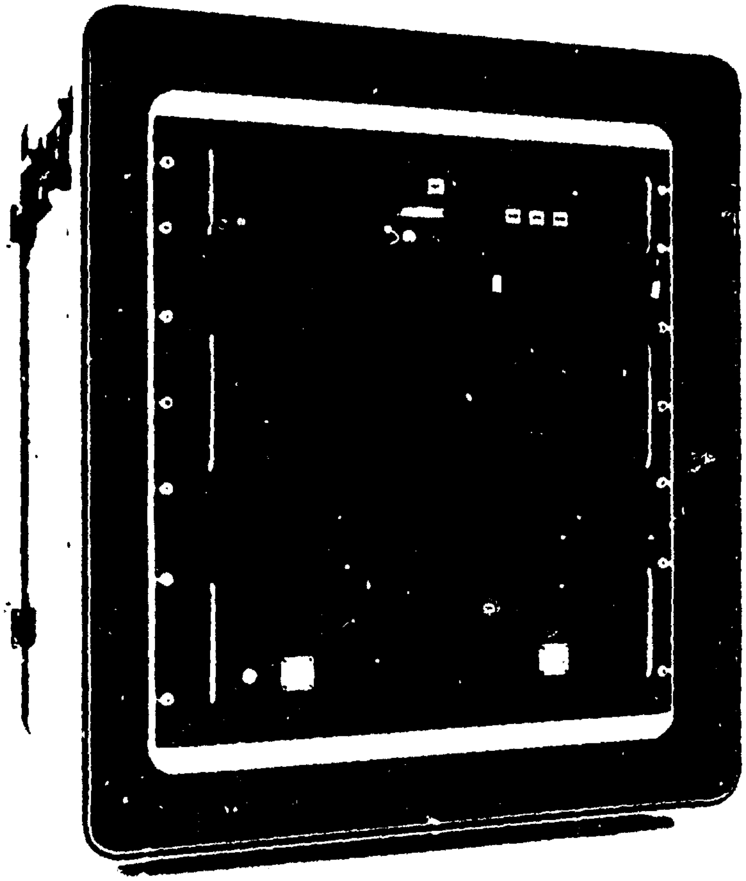


Fig.3 T-1373/TRQ-35(V) Chirpsounder transmitter

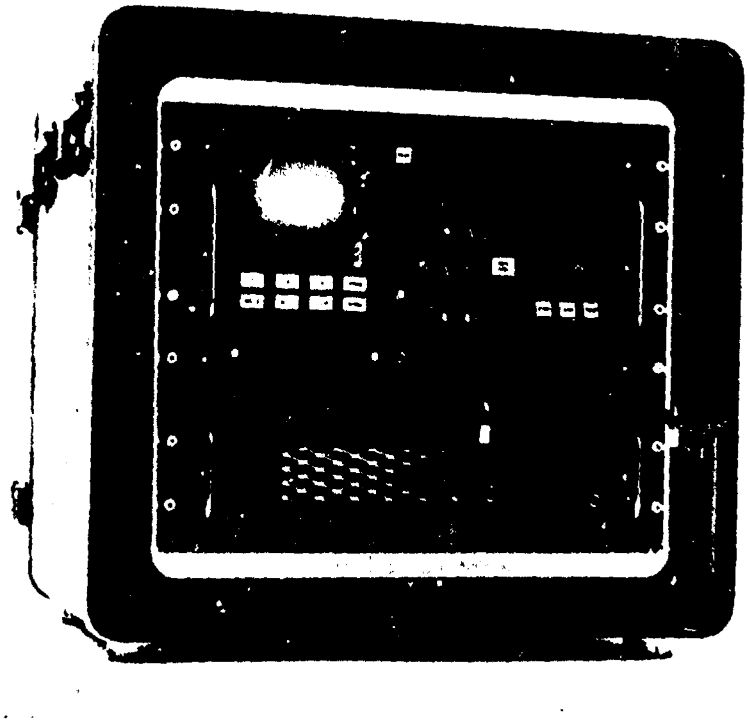


Fig.4 R-2081/TRQ-35(V) Chirpsounder receiver

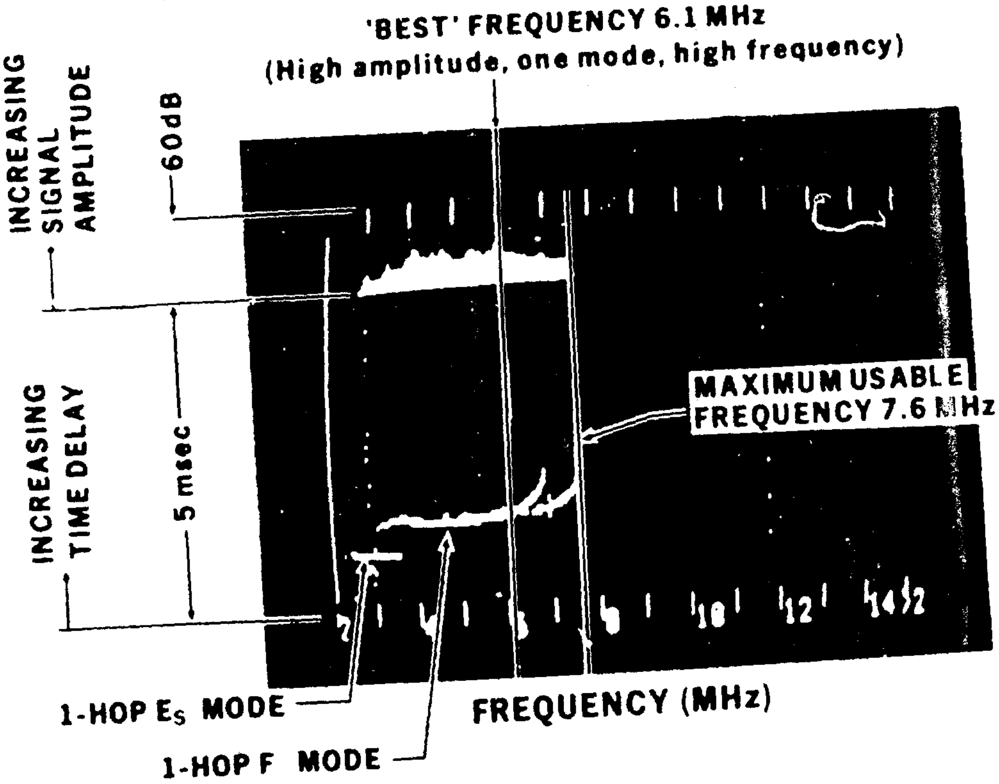


Fig.5 Examples of record made by Chirpsounder receiver

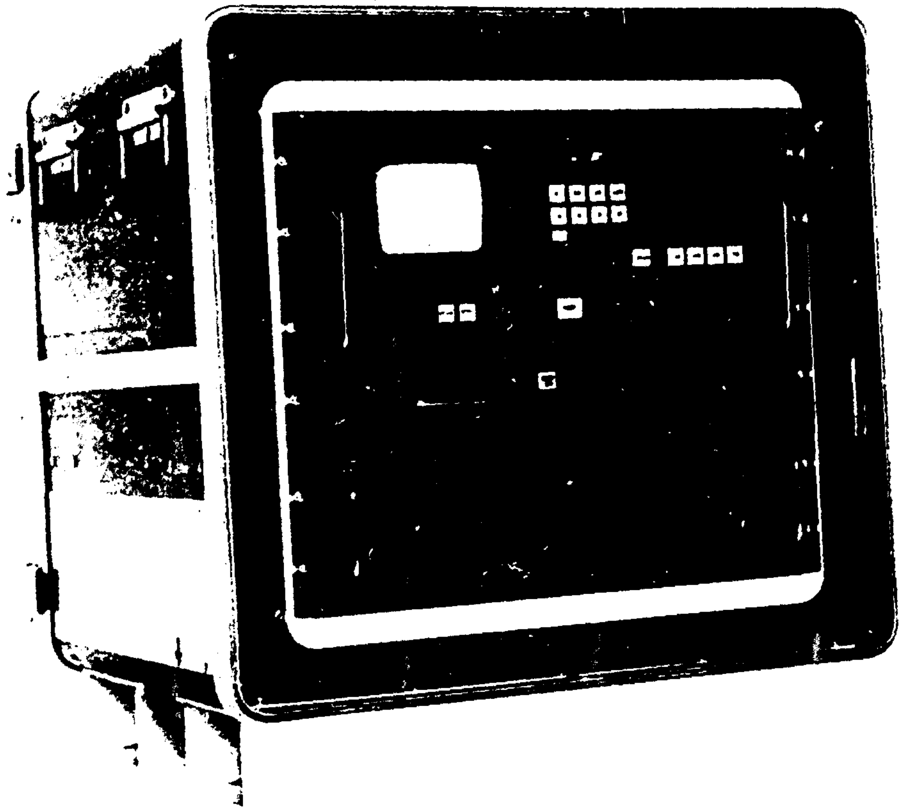
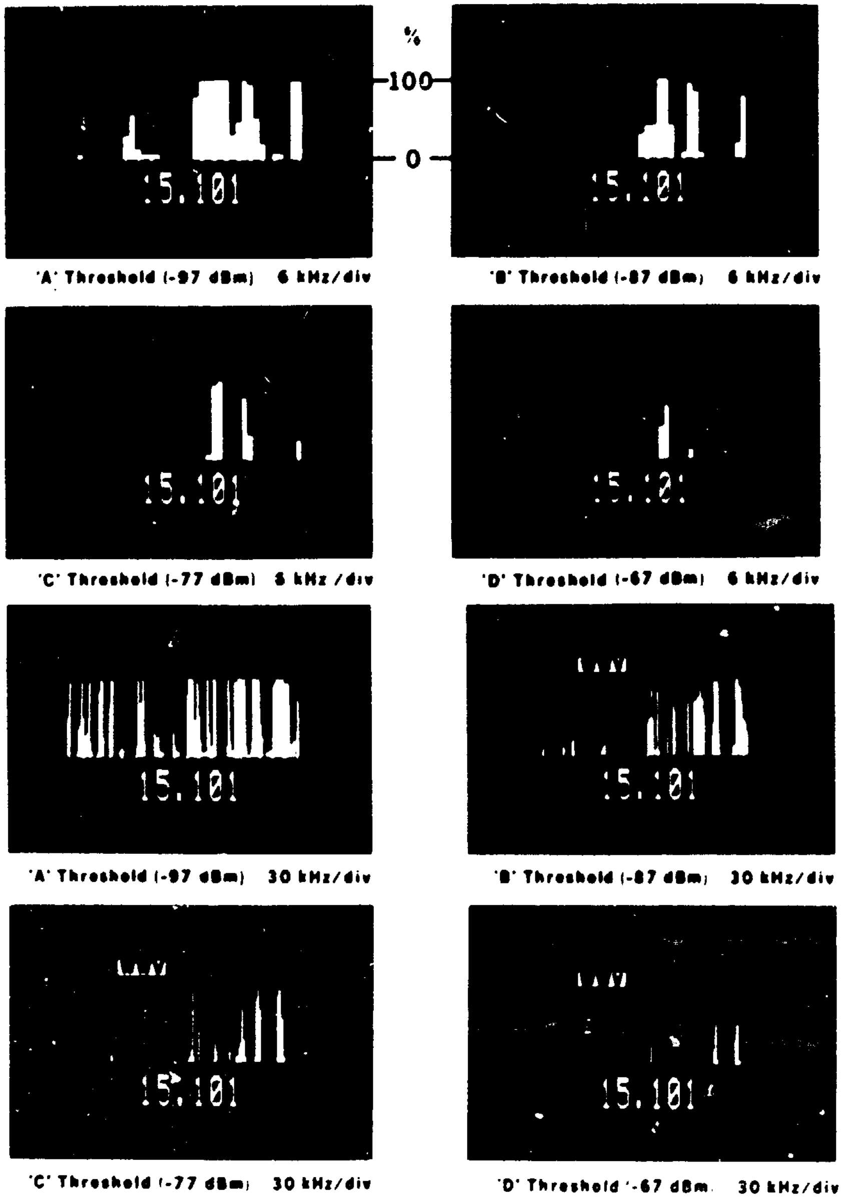


Fig.6 R-2093/TRQ-35(V) spectrum monitor



'A' Threshold (-97 dBm) 6 kHz/div

'B' Threshold (-87 dBm) 6 kHz/div

'C' Threshold (-77 dBm) 6 kHz/div

'D' Threshold (-67 dBm) 6 kHz/div

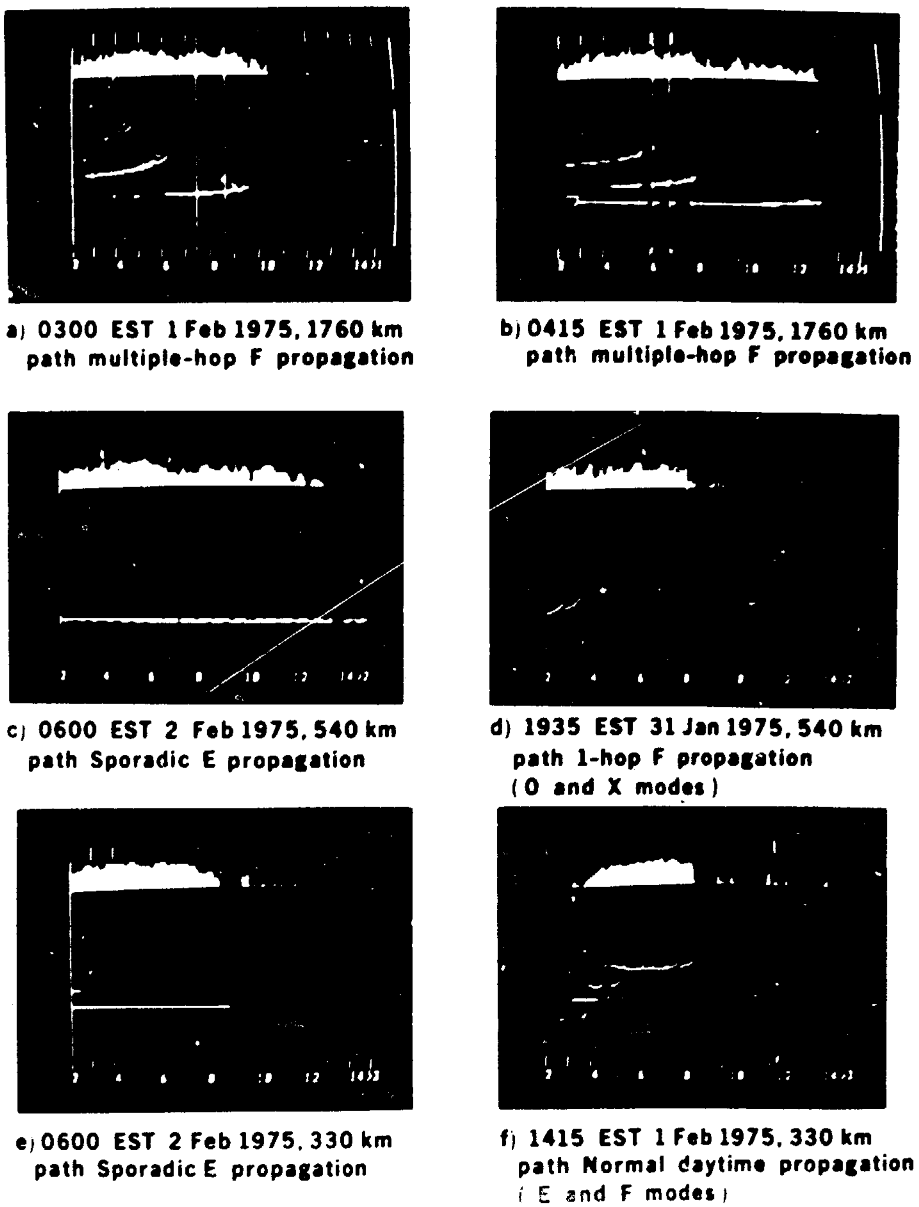
'A' Threshold (-97 dBm) 30 kHz/div

'B' Threshold (-87 dBm) 30 kHz/div

'C' Threshold (-77 dBm) 30 kHz/div

'D' Threshold (-67 dBm 30 kHz/div

Fig.7 Examples of 30 minute spectrum monitor displays circa 15.1 MHz, midday made during the TD III exercise. Note strong signals in international broadcasting band above 15.1 MHz and light occupancy in aeronautical mobile (OR) band between 15 MHz WWV and 15.1 MHz



a) 0300 EST 1 Feb 1975, 1760 km path multiple-hop F propagation

b) 0415 EST 1 Feb 1975, 1760 km path multiple-hop F propagation

c) 0600 EST 2 Feb 1975, 540 km path Sporadic E propagation

d) 1935 EST 31 Jan 1975, 540 km path 1-hop F propagation (O and X modes)

e) 0600 EST 2 Feb 1975, 330 km path Sporadic E propagation

f) 1415 EST 1 Feb 1975, 330 km path Normal daytime propagation (E and F modes)

Fig.8 Examples of different types of Chirpsounder records made during the TROPHY DASH III Exercise

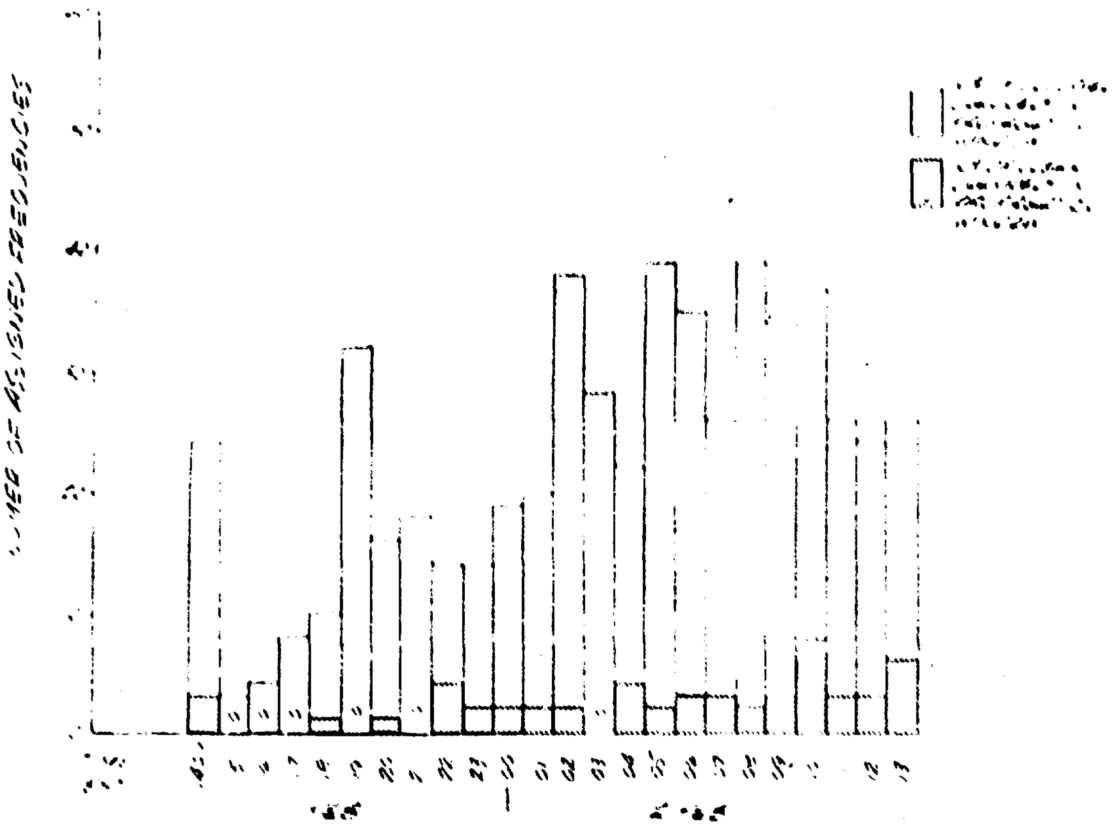


Fig.9 Analysis of utility of TAC frequency assignments in a 14 hour period on Altus path

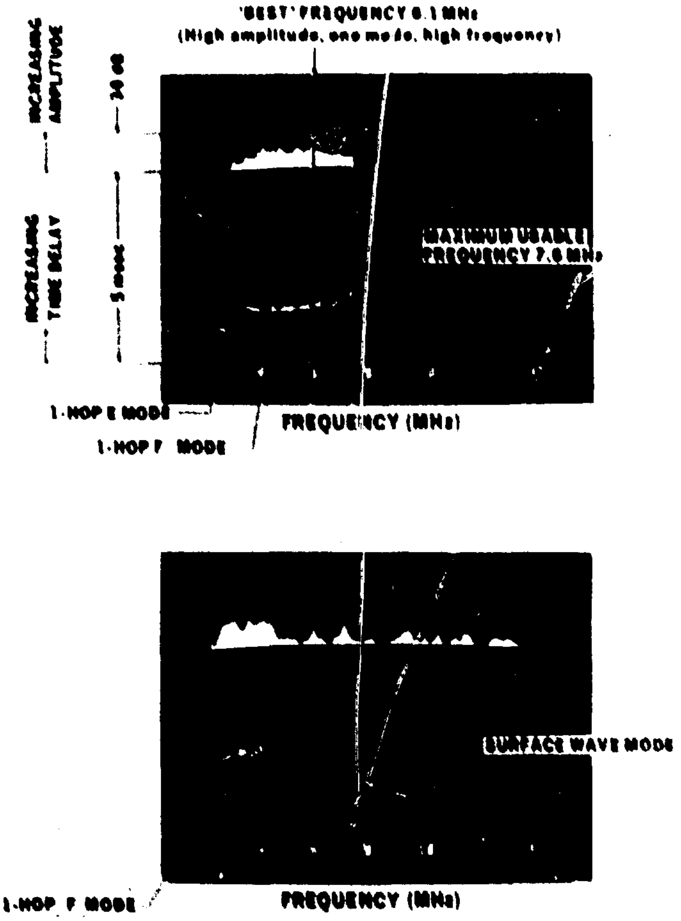


Fig.10 Examples of Chirpsounder records made in Norway in September, 19'6.
Upper: daytime, excellent conditions; lower: ship-to-shore, 40 km path

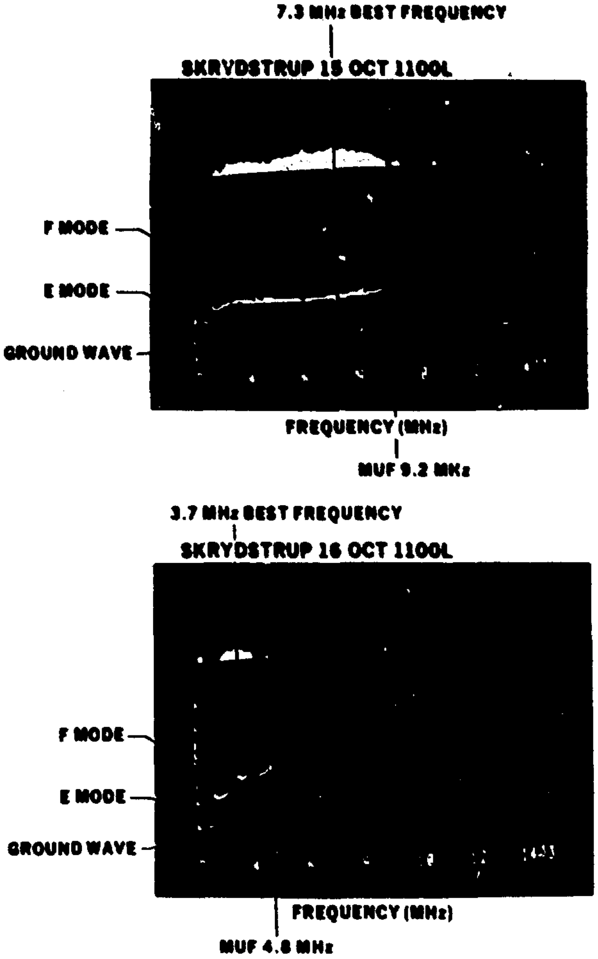


Fig.11 Example of day-to-day variability of propagation, records made in Denmark, Skrydstrup-Vandel path, 1100 LT on 15 and 16 October 1976

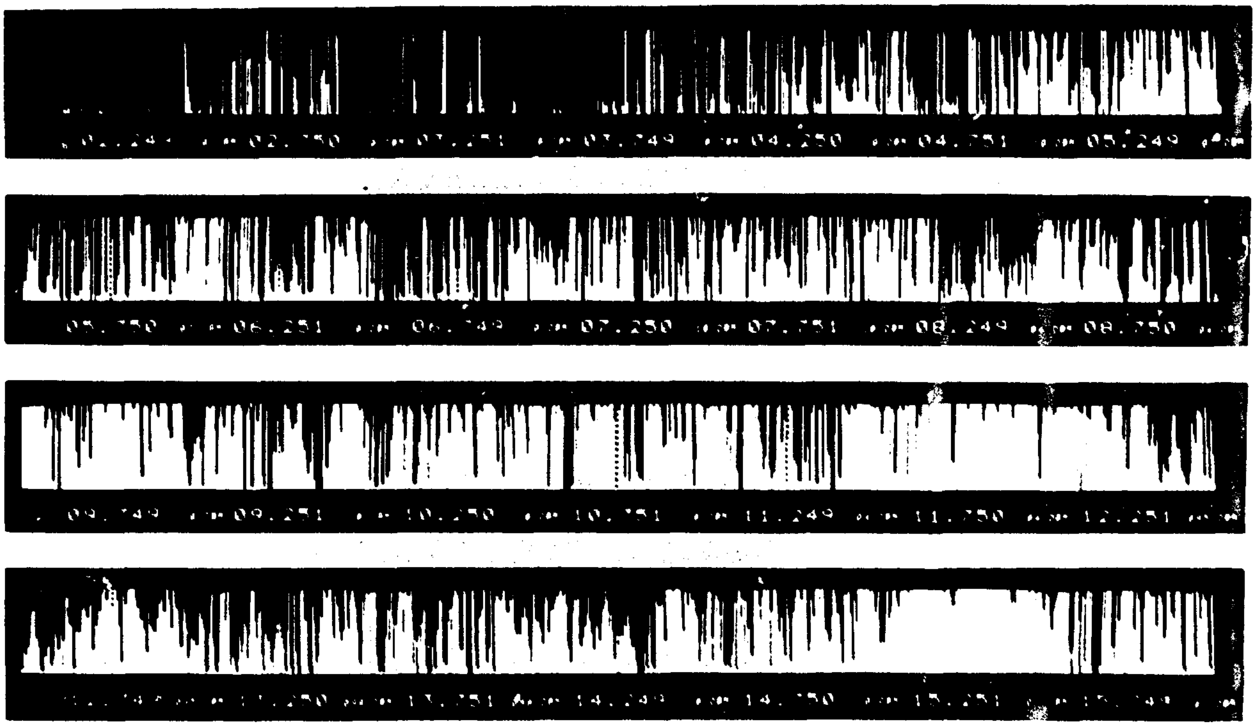


Fig.12 Example of European daytime spectrum occupancy measured with R-2093 spectrum monitor; Germany, 1100 LT, 22 September 1978

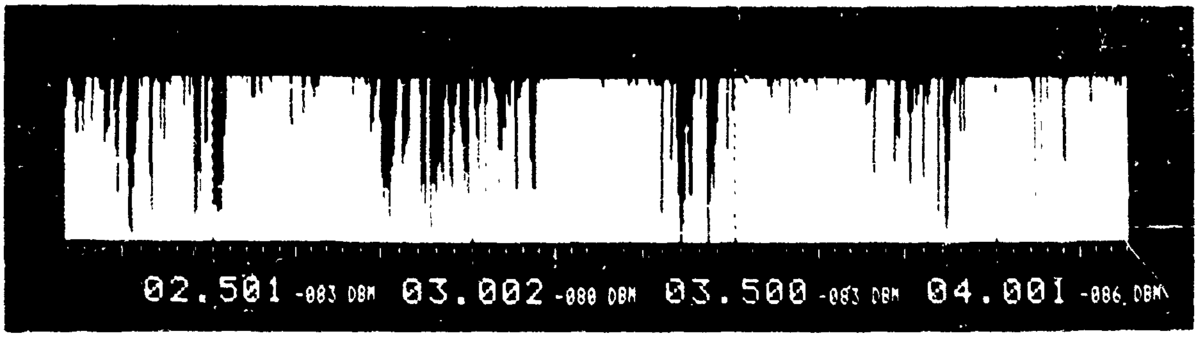
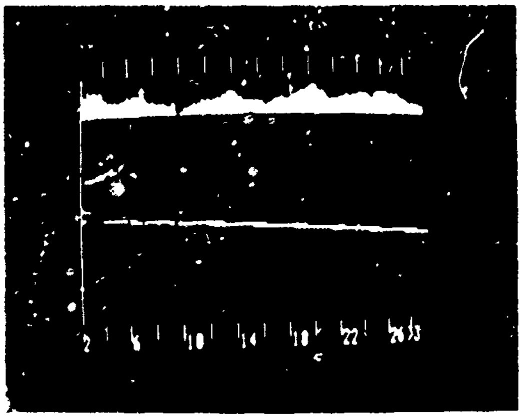


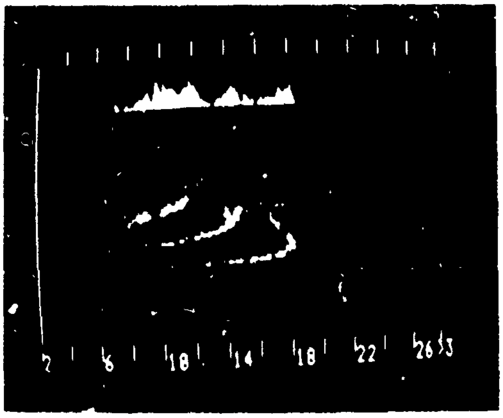
Fig.13 Example of European nighttime spectrum occupancy measurement with R-2093 spectrum monitor, Belgium, 2030 LT, 25 October 1977



1910Z **200 km**

Fig.14 Example of ground-to-air Chirpsounding, Croughton, England, to aircraft (EC-135K). Note range change with time of the line-of-sight mode (lowest trace)

1800 GMT --- 6000 KM



5 APRIL 1978

Fig.15 Example of long-path Chirpsounder record: 1800 GMT, Sultanate of Oman – England, 6000 km, 5 April 1978

DISCUSSION

Prof. Shearman, UK

The approach to channel selection adopted in the paper is to see whether the channel was received at the receiving station of the system is clear of interference. However, this test does not guarantee that transmission on this frequency will avoid interference with a remote receiver For instance if that remote receiver is listening to a transmitter within the skip range of our station, we will only receive weak back-scattered signals and may consider the channel clear.

Author's Reply

While freedom from interference cannot be guaranteed by the method proposed the probability of interference is greatly reduced by the feature of statistical compilation over a 30 minute period In the method proposed it has been recommended that spectrum monitors be deployed at both ends of a path if operation is to be above the vertical-incidence critical frequency Note further that a channel appears as "occupied" if signal-to-noise ratio exceeds 3dB in a 6kHz bandwidth. Thus ground backscatter from beyond the skip zone will very frequently be received and displayed by the spectrum monitor in the situation described by Prof. Shearman

ASSESSMENT OF HF COMMUNICATIONS RELIABILITY

by

N. M. Maslin

Radio and Navigation Department
Royal Aircraft Establishment
Farnborough, Hampshire, England

SUMMARY

The concept of circuit reliability for an HF sky-wave link is discussed. Its frequency dependence is considered together with the relationship to the median received signal-to-noise ratio for a given HF circuit. The importance of a study of this kind is the ability to be able to quantify changes that could be made in an HF circuit, in terms of reliability improvements, and hence to make decisions of the cost effectiveness of increasing transmitter power, antenna efficiencies, directional antenna design, etc.

This study should prove particularly useful for HF communications to mobiles such as ships or aircraft. High transmitter powers and high gain antennas of point-to-point land fixed HF communication circuits are not available and the power budget is more critical. Furthermore, cost-effective improvements are often of prime concern when a large number of shipborne or airborne equipments are involved. The assessment of communications reliability, particularly to military aircraft, can have an important bearing upon mission effectiveness.

1 INTRODUCTION

High frequency (2-30 MHz) radio waves that propagate via the ionosphere are used extensively for long range point-to-point communications and broadcasting. Circuit planners need to know what frequency ranges their systems should be capable of covering, what transmitter powers are necessary to overcome the background noise at the receiver, and what antenna configurations would be most suited to the applications required. These are all factors dictated by propagation considerations and long term prediction techniques are available for this purpose (BRADLEY, F.A., 1975; BARCLAY, L.W., 1974). Characteristics of such links have consequently been extensively studied; commercial services are available for prediction of optimum working frequencies and quality of communications at these frequencies. Most point-to-point land fixed HF communication circuits use high-gain rhombic or log-periodic antennas, whilst arrays of horizontal dipoles, also with significant directivity, are popular for broadcasting using the sky-wave mode. In principle, therefore, the capability of these links may be optimised by good engineering design and practice in respect of the equipment and antenna systems, whilst powerful transmitters are often available.

Much more difficult problems are presented, however, by HF communication to mobiles such as ships or aircraft, and some examples of air-ground links are given later. Often communication is required at ranges from a few kilometres to several thousand kilometres over a wide variety of terrain, and this implies different modes of propagation according to range. Difficulties arise from the physical constraints which are placed upon the aircraft antenna, from serious excess noise as well as from additional multipath propagation mechanisms and Doppler frequency shifts. To achieve satisfactory results over an HF link of this kind, careful consideration must be given to the terminal radio equipment, the planning of operational links and the management of the frequency to be used over those links.

Faced with problems of this complexity, the communications system analyst seeks to isolate a number of problem areas for study. This paper endeavours to clarify the steps necessary to evaluate the reliability of an HF link and shows the value of such an assessment for air-to-ground communication links in which sources of degradation in the circuit can be very troublesome. It does not specifically address the problems caused by multipath features, time and frequency dispersion and the conclusions reached apply assuming that such effects are not major. It is hoped, however, that the following procedure may help the communications engineer to decide what improvements are needed and how they can be made in the most cost effective manner.

2 CONSIDERATIONS FOR THE TOTAL ASSESSMENT PROBLEM

A typical radio communication circuit is required to perform a function. Usually this function involves the transmission of information from one location to another. The measurement of how well the circuit performs involves both the volume of information that can be transmitted during a given period of time as well as the accuracy with which the input information is reproduced at the output.

We wish to maximise information transfer rate, and minimise the required transmitted power. Nature and man are the sources of all engineering problems. Effects of nature are centred upon the existence of the Earth and its environment within the universe and solar system. Electrical noise impinging upon the Earth from solar and cosmic sources establish an ultimate bound on the information that can be conveyed throughout the HF radio spectrum. This bound is modified by radiation from noise sources within the troposphere, the terrestrial environment and primarily from man-made radio frequency sources. Long-term characteristics of this noise affect the required transmission power whilst short-term chatacteristics determine how the signal should be designed and detected to convey the desired information.

Fig 1 shows a concept of a communication circuit. The message is inserted in the transmitter system where it becomes part of the modulation of an EM wave radiated from the transmitting antenna. The receiving antenna picks up only an infinitesimal amount of the power radiated as well as unwanted radiations from many interfering sources. The ratio of the power input at the terminals of the transmitting antenna to the resulting signal power available at the receiving antenna is known as the system loss.

Copyright © Controller HMSO

The receiver itself contributes some thermal noise although at HF the receiver noise is not usually the limiting factor.

At the receiver output the signal carrying the desired information must be of sufficient energy relative to all of the unwanted signals and noise combined to provide for a sufficiently faithful reproduction of the original message. There are many sources of RF interference and noise. Some are in a sense controllable from the standpoint of adequate receiver design and siting, whilst others are likely to be present in all cases. After allowance for the system loss, sufficient transmitter power must be used to overcome the combined effect of all of the sources of unwanted RF power at the receiver. Diversity transmission and reception depends upon the degree of similarity in fading occurring simultaneously on propagation paths separated by space or frequency or with different polarisations. As modulation bandwidths are increased, selective fading becomes more of a problem.

To transmit information it is necessary to modulate the fundamental operating frequency of the transmitter. Regardless of whether the modulating signal is digital or analogue, or whether amplitude, frequency, phase or other type of modulation is used, the output signal contains many frequency components other than the fundamental.

A given communication system can only be as effective as the weakest link in the chain from transmitter to receiver. In the case of air-ground systems, therefore, the ground terminal is just as important as the aircraft terminal and must be specified accordingly. Since there are likely to be many more aircraft than ground stations, and all airborne electronic equipment penalises the aircraft performance because of its mass, volume and power consumption, it is important to keep the complexity of the airborne communication terminal as low as possible.

It may, of course, be expensive to realise communication reliability targets and cost-effectiveness studies must be carried out. This, in turn, implies a subsidiary study relating communications reliability to mission effectiveness. The main factors which degrade reliability need to be identified. Assessments can then be made of reliability improvement which may be achieved by realistic modifications to the total air and ground systems.

In the initial planning stages it is important to have an overall picture of the HF link. This may be obtained from a power level diagram, such as shown in Fig 2. The vertical axis is scaled in power, the horizontal axis is arbitrarily scaled but progresses from left to right as one goes from the transmitter output power P_T via the propagation path to the receiver input. Gains are indicated by positive slopes, losses by negative slopes. The space between the broken lines drawn parallel to the horizontal axis represents signal-to-noise ratio prior to the receiver demodulator. Fig 2 shows a long range sky-wave propagation path at night. Note that although the propagation losses are similar for 4 MHz and 9 MHz, the overall signal power at 4 MHz is reduced due to the poorer aircraft antenna radiation efficiency. Fig 2 clearly shows that the 9 MHz frequency would be satisfactory with a signal-to-noise ratio of about 25 dB; the 4 MHz frequency would be totally unacceptable, however.

3 BASIC DEFINITIONS AND CONCEPTS

3.1 Sky-wave availability

The upper limiting frequency of sky-wave propagation is governed by the distribution of electron concentration along the propagation path. This upper limit is known as the maximum usable frequency (MUF) and may be considered as the maximum frequency of a wave capable of propagating over a given sky-wave path. Because the properties of the ionosphere exhibit temporal variations, the MUF continually fluctuates; but predictions of ionospheric conditions are based upon monthly median conditions and some allowance must be made for day-to-day variability. Thus the predicted MUF is defined as that frequency for which signals are expected to be available for 50% of the days at a given hour within a given period, usually a month. It is denoted here by f_m. A note of caution should be made. The MUF concept can only be considered an approximation since many factors can lead to errors in estimates of the MUF. For example, when the ionosphere is highly structured spatially the HF signal may not disappear completely when the frequency is above 'the MUF'.

The sky-wave availability Q is defined as the percentage probability that radio signals can propagate at a given hour over a given sky-wave path. It may be derived in terms of the statistics of the day-to-day variations of the MUF.

3.2 Variation of received signal-to-noise ratio

The transmission loss of an HF sky-wave propagation path depends upon a number of factors, and these can have implications on various aspects of the communication link. Day-to-day variations in the electron concentration within the E and F regions of the ionosphere influence the direction which a given ray path takes through the ionosphere to reach the receiver. These fluctuations in ray path direction may, in turn, change the effective gains of the transmitting and receiving antennas. Different ray path directions can produce different effects of focusing, spatial attenuation, sporadic E losses, polarization losses and multipath phenomena. Changes of electron concentration in the D and E regions can greatly affect the value of ionospheric absorption. Day-to-day variations may also be produced in atmospheric noise intensities by these mechanisms.

The monthly median value of the mean signal power, S, for a particular mode may be combined with the monthly median of the mean noise power, N_0, within a 1 Hz bandwidth to produce an estimate of the monthly median value of S/N_0, the signal-to-noise ratio in a 1 Hz bandwidth. From a knowledge of the day-to-day variability of the signal and the noise it is then possible to determine the probability of exceeding some specified required S/N_0 ratio. We shall denote by P the percentage probability that the mean received S/N_0, whose monthly median is R_m, is above some specified level R_0.

3.3 Circuit reliability factor

This is defined (CCIR, 1974a) as the fraction of days that successful communication may be expected at a given hour within the month as a specific operating frequency. The circuit reliability factor is based upon monthly median estimates of propagation parameters and their distributions and represents the fraction of days in the month at the given hour that successful communication is expected.

The primary factor in determining the circuit reliability is the long term median S/N_0 ratio. This is directly associated with a grade of service, which in turn defines the type of communication desired; for example, the percentage of error-free messages in teletype transmissions or the intelligibility of voice transmissions. A minimum required S/N_0 ratio is associated with the desired grade of service. This ratio depends upon many factors such as modulation index, signalling rates and codes and includes effects of fading, error-correcting schemes, optimum modulation and detection techniques and diversity schemes.

Mathematically, the circuit reliability factor can be expressed (MASLIN, N.M., 1978) as a dimensionless quantity ρ given by

$$\rho = QP/100$$

where ρ, Q, P are expressed as percentages.

3.4 Dependence of reliability on frequency

The magnitude of ρ depends ultimately upon two opposing effects:

(i) P increases with increasing frequency. For a given sky-wave path the higher the frequency the less the absorption and the less the received noise; hence the greater the received S/N_0 ratio.

(ii) Q decreases with increasing frequency. The higher the frequency the less the chance there is of ionospheric reflection at that frequency.

A typical variation of reliability with frequency is shown in Fig 3.

3.5 Effect of signal-to-noise on reliability

Consider now the effect of increasing the received S/N_0 ratio of the HF link by some means such as increase of transmitter power, improvement of antenna efficiencies, reduction of received noise, etc. Then P increases, as shown in Fig 4 and hence ρ also increases (Fig 5). The factor Q, however, remains constant and shows the maximum value of ρ which is attainable. As Fig 5 shows, the Q curve forms an 'envelope' to the reliability curves.

Thus, for example, if the median received S/N_0 ratio, R_m, is such that curve E is applicable in Fig 5 and 90% reliability is required, then a horizontal line drawn from $\rho = 90\%$ cuts curve E at two values of f, denoted by f_1 and f_2. The lower value f_1 is known as the lowest usable frequency, LUF, the upper value f_2 is just below the optimum traffic frequency, FOT. For frequencies below the LUF, absorption is too great to give the required S/N_0 ratio; above f_2 the ionosphere cannot support the given frequency for the required percentage of time. Thus satisfactory operation can only be achieved between f_1 and f_2. Note that if the median S/N_0 ratio is reduced to that for curve D, there is no frequency which would satisfy the service criterion.

3.6 Assessment of reliability improvement

So far circuit reliability has been discussed in qualitative terms. To adequately assess the performance of a radio link, the communications engineer needs to know what improvement in circuit reliability can be achieved by a given increase in the received S/N_0 ratio. The increased value of ρ does not depend on how the improvement is made; for example it might be achieved by an increase in transmitter power, use of directional receiving antennas, improved transmitter antenna efficiency, etc.

Fig 6 shows the values of reliability calculated using the CCIR Bluedeck (BARGHAUSEN, A.F. *et al*, 1969; HAYDON, G.W. *et al*, 1976) program for a point-to-point link in February 1976. It is convenient to label the abscissa as 'dB above required signal-to-noise' and thus a direct comparison of the effect of different frequencies is made, although different initial transmitter powers or system gains are required to produce the same signal-to-noise value for different frequencies.

Thus, for example, if the signal-to-noise ratio on the link is improved by 10 dB by some means (increased transmitter power, improved efficiencies, antenna gain, etc) the corresponding improvement in reliability can be estimated from the curves. The following points should be noted:-

(i) For large values of $(R - R_0)$, ρ tends to the value Q since P tends to unity. This means that well below the MUF $\rho \approx 1$ while as f increases ρ decreases for a given signal-to-noise ratio. At the MUF $\rho = \frac{1}{2}$ and above it $\rho < \frac{1}{2}$.

(ii) For very large negative values of $(R - R_0)$, $\rho \approx 0$ for all frequencies, since the probability that the instantaneous S/N ratio exceeds the required level R_0 is very small.

(iii) Between the extremes (i) and (ii) the curves are approximately linear, passing through the value $\frac{1}{2}Q$ at $R = R_0$. The slope is determined by the statistical spread of absorption loss and external noise characteristics.

(iv) The slopes of the curves in Fig 6 are proportionately reduced by the value of Q. Thus the MUF has a slope of one half of that for frequencies which have $Q = 1$.

(v) Fig 6 appears to show that the lower frequencies are the most reliable. This is only true if the actual received signal-to-noise ratio for each frequency is the same. In this case the lower frequencies are advantageous because of the more favourable value of Q. However, the lower frequencies are often more heavily absorbed, particularly during the daytime.

3.7 The required signal-to-noise ratio

To assess the circuit reliability it is first necessary to adopt a minimum required S/N_0 ratio criterion for a given grade of service. For example, CCIR (CCIR, 1974b) recommend a 15 dB S/N ratio for marginal and 33 dB for good commercial quality of communications with HF SSB voice in a 3 kHz bandwidth. In practice, criteria must be chosen to suit the communication requirements; higher S/N_0 ratios than the minimum usually specified might be required to reduce data errors due to bursts of noise or to achieve intelligibility of unrelated words on speech transmission.

4 AIR-TO-GROUND COMMUNICATION LINKS

4.1 General features

From the foregoing sections it is believed that some guidance can be offered for the design and operation of air-ground HF communication links. It is first necessary to choose some reliability factor criteria, bearing in mind that these may be much lower for mobile communications than for commercial point-to-point links. For the present purposes $\rho = 20\%$ (nearly useless) and $\rho = 80\%$ (satisfactory) have been chosen.

In common with fixed point-to-point links it is most important to choose an operating frequency appropriate to the propagation and noise conditions prevailing at the time. Provided that this is done, the margin between nearly useless and satisfactory communications has been shown (MASLIN, N.M., 1977) to be from 20 dB to 35 dB depending upon the time of day. Frequency allocations and interference may limit the choice of frequency; unfortunately, operators do not always strive to achieve optimum conditions.

4.2 Modification of the reliability curves

Consider now how the properties of, for example, the aircraft antenna would change the shape of the reliability curve in Fig 3. A typical graph of aircraft antenna efficiency against frequency is shown schematically in Fig 7a. The frequencies f_{10}, f_{20} ..... designate antenna efficiencies of -10 dB, -20 dB ..... and are different for large and small aircraft. The modified reliability curve for $R_0 = 0$ dB in Fig 7b is then given by the chain curve, which passes through the $R_0 = 10$ dB curve at $f = f_{10}$, $R_0 = 20$ dB at $f = f_{20}$ and so on. The effect of the frequency variation of antenna efficiency is therefore to shift the $R_0 = 0$ dB curve towards the higher frequencies and this restricts the frequency range capable of providing a given reliability, ρ', from $(f_2 - f_1)$ to $(f_2 - f_1')$, see Fig 7b. The width of the frequency 'window' is reduced. The centre of this 'window' moves along the frequency axis according to the prevailing conditions. For example, a high sunspot number year tends to increase Q for a given frequency because higher frequencies can be propagated as a result of the increase in MUF. The value of P, however, may decrease for a given frequency because of the greater ionisation which enhances absorption. Thus the centre of the frequency 'window' moves to higher frequencies than those for low sunspot number years. The width of the 'window' also changes. Diurnal variations also radically affect both the width and centre frequency of the window.

Based on the example of the Norway-UK link the calculated results for large and small aircraft are given in Fig 8. In these examples the efficiencies at 4 MHz of the antennas have been taken as 10% and 0.1% respectively, which are typical results (PAVEY, N.A.D., 1973). The effect of the poorer antenna efficiency on the small aircraft is twofold:

(i) the usable frequency range is decreased; from 4.4-11.8 MHz to 6.3-11.8 MHz for $\rho = 80\%$,

(ii) the reliability at the lower frequencies is decreased; at 4.4 MHz from 80% to 27%.

The example chosen is considered to be fairly representative of a long range air-ground link. However, as mentioned in the introduction, communication in a variety of circumstances is commonly required; both ground-wave (for short distances) and higher angle sky-wave (for medium distances) propagation paths need also to be considered if an effective overall assessment of communications reliability is to be made. These cases may be treated initially in a similar manner to the analysis in this paper. The value of R_0 needs to be established for particular cases and this may be influenced by several factors associated with the aircraft.

4.3 Example of a small aircraft

The dependence of reliability upon frequency is shown in Fig 9, which gives predicted values for a small aircraft operating over two short range links at noon in January 1976. Note that at the lower frequencies there is a rapid decrease in reliability due to the decrease in antenna efficiencies (see Fig 7a). The curves D to A show examples of progressive improvements which might be made. Curves C and D show the effect of 10 dB ground directivity factors whilst curves A and B show how progressively increasing the antenna efficiencies could increase reliability.

4.4 Example of a large aircraft

The purpose of this section is to illustrate how conditions are likely to change and how these changes can affect the overall reliability for a typical mission profile of a long range maritime patrol aircraft.

Consider air-ground voice communications utilizing 10 dB ground receiving antenna directivity and assume that the mission is of 10 hours duration. The ground station is in North Scotland and the aircraft flies northward at 450 kn for 3 hours, remains at this range for 4 hours and then returns, again at 450 kn, during the last 3 hours of the mission.

Four curves are given: the MUF, the frequency which provides the best reliability (this is not necessarily the optimum working frequency), the reliability factor for this best frequency and the reliability factor for a given fixed frequency.

Fig 10 shows a mission starting at 0200 hours. The optimum frequency is a compromise between sky-wave availability and signal absorption. Note that the optimum reliability attainable is approximately constant over the whole mission. For a constant (3 MHz) frequency the January results show that this would be a reasonable frequency for most of the mission, but when the MUF increases (around 1000 hours) the reliability of 3 MHz decreases and the optimum frequency rises to about 8 MHz. Results for April show that the MUF is greater than for January and hence 3 MHz is not a good choice since the MUF is above 8 MHz for most of the mission; 3 MHz signals are too heavily attenuated to be useful.

5 DISCUSSION

It has been shown that the reliability of an HF air-ground sky-wave link is a complicated function of a large number of factors. Recourse has therefore been made to representative examples to illustrate the main phenomena. Based on these examples the objective of this section is to provide guidance for the design and management of such links.

Consider firstly the link management aspects. The most crucial factor is the choice of operating frequency for a particular set of conditions. The upper frequency bound is determined by the availability of a suitable propagation path, which is governed primarily by the MUF. The lower frequency bound is realised by the absorption within the ionosphere, so that it is a function of the performances of the communication terminals. The useful frequency band available to the communicator depends not only on the signal strength and fading characteristics, but also on the dispersion caused by multipath propagation and ionospheric movements. The normal prediction methods should therefore ideally be supplemented by daily forecasts and perhaps ionospheric soundings. The system designer needs predictions of frequencies which have tolerable dispersion in addition to adequate signal-to-noise ratios. If a choice of frequency is available, the highest frequency which offers acceptable reliability is often to be preferred since this minimises multipath effects (*ie* f_2 is preferable to f_1 in Fig 5).

Ground-air and air-ground communications must be treated as separate situations in view of the different transmitter powers, radio frequency noise-field and antenna characteristics at the ground and airborne terminals. The MUF increases with range and if a choice of ground stations is available it may sometimes (LIED, F., 1967; MASLIN, N.M., 1979) be advantageous to work to the more remote station so that higher working frequencies can be used. Consequently the need for continual frequency changing would be eliminated; the sky-wave propagation modes would also be operable over longer periods of a flight compared to the shorter range link.

Particularly severe problems are presented by HF communications with small aircraft at moderate ranges, primarily as a result of very poor aircraft antenna efficiency in the low part of the HF band. Working over a longer range communication link increases the optimum traffic frequency, thus avoiding the poor antenna efficiencies and generally reducing the external noise levels. This is discussed more fully in a comparison paper (MASLIN, N.M., 1979).

If all-round coverage is not required by the ground receiving station, then an important parameter is the directivity (*ie* a geometrical factor) rather than absolute gain of the ground antenna.

Ground antenna directivity gives improved circuit performance when the system is limited by external noise, but is of little help when availability of the propagation path itself is a major constraint.

Sources of aircraft noise (BSI, 1973) are important factors for consideration and are often the predominant noise sources received in the aircraft. They include electrostatic discharges, rotating electrical machinery and switching transients so that the noise is likely to contain both Gaussian and impulsive components. Radiated interference from the avionic installation is picked up by the aircraft antennas; it is important to provide adequate attenuation of this noise by the aircraft airframe.

All the above features are important when assessing an HF air-ground link since the power budget is usually more critical than for a fixed point-to-point link. Suppose it is deduced from the foregoing discussion that a 10 dB improvement in median received signal-to-noise ratio would provide a worthwhile increase in communications reliability for a given air-ground link. It is then necessary to decide upon the most cost effective solution. It might be better, for example, to improve the ground station directivity factor by 10 dB rather than increasing the aircraft transmitter power tenfold or improving the antenna efficiencies of a large number of aircraft.

6 CONCLUSIONS

The assessment of circuit reliability has been shown to depend upon two important parameters.

(i) The ratio (f/f_m) of the operating frequency to the MUF. Since the MUF is fixed for a given time of day, season and latitude, the associated sky-wave availability Q is fixed and the reliability has a maximum value of Q.

(ii) The ratio of the received to required signal-to-noise ratio ($R_m - R_0$ in dB). The required ratio R_0 is fixed for a given modulation mode by the grade of service needed, whilst R_m can be changed by adjusting parameters within the link. The associated probability factor P is, therefore, to some extent controllable.

The range of reliability improvements which is possible for a given increase in $(R_W - R_0)$ has been discussed. In this way a communication link can be assessed to decide what improvements would prove cost effective.

The choice of a suitable frequency for the link for a given time of day involves an assessment of two conflicting factors.

(i) Higher frequencies are less liable to be propagated (Q is smaller).

(ii) Lower frequencies are more heavily attenuated (P is smaller).

Moreover atmospheric noise generally increases with decreasing frequency. Thus for a given situation there may be a band or 'window' of frequencies available within which the required reliability could be achieved for the particular link configuration employed.

This paper has concentrated on examples of an air-ground sky-wave link because the power budget can be very crucial; an assessment of the communications reliability of aircraft under operating conditions is an essential step in the evaluation of their effectiveness. It has been shown how the resultant received median signal-to-noise ratio, and in turn the circuit reliability, depends upon a number of factors such as transmitter power, propagation path loss, antenna efficiencies, etc. In the course of this study a number of simplifications have been made. Such effects as multipath features, time and frequency dispersion, the change in apparent aircraft antenna performance due to ground reflection, and uncertainties in the aircraft antenna polarisation characteristics have not been specifically addressed. However, it is believed that an understanding of the basic features, outlined here, should prove useful to the communications engineer concerned with the planning and assessment of HF sky-wave links.

REFERENCES

BARCLAY, L.W., 1974, "Ionospheric prediction techniques". The Marconi Review, 37. 51-66.

BARGHAUSEN, A.F., FINNEY, J.W., PROCTOR, L.L. and SCHULTZ, L.D., 1969, "Predicting long-term operational parameters of high frequency sky-wave telecommunication systems". ESSA Technical Report, ERL 110-ITS 78, US Government Printing Office, Washington.

BRADLEY, P.A., 1975, "Long-term HF propagation predictions for radio circuit planning". The Radio and Electronic Engineer, 45, 31-41.

BRITISH STANDARDS INSTITUTION, 1973, "General requirements for equipment for use in aircraft (BS3G100)". Part 4, section 2 "Electromagnetic interference at radio and audio frequencies". BSI, London.

CCIR, 1974a, Report 252-2, "Interim method for estimating sky-wave field strength and transmission loss at frequencies between the approximate limits of 2 and 30 MHz". ITU, Geneva.

CCIR, 1974b, Recommendation 339-3, "Bandwidths, signal-to-noise ratios and fading allowances in complete systems". ITU, Geneva.

HAYDON, G.W., LEFTIN, M and ROSICH, R., 1976, "Predicting the performance of high frequency sky-wave telecommunication systems". (The use of the HFMUFES4 Program). OT Report 76-102.

LIED, F. (Editor), 1967, "High frequency radio communications, with emphasis on polar problems" (p.112). AGARDOGRAPH 104, Technivision, Maidenhead, England.

MASLIN, N.M., 1977, "Factors governing the improvement of circuit reliability for an HF air-ground link". RAE Technical Report (unpublished), TR 77105.

MASLIN, N.M., 1978, "Assessing the circuit reliability of an HF sky-wave air-ground link". The Radio and Electronic Engineer, 48, 493-503.

MASLIN, N.M., 1979, "HF communications to small low flying aircraft". Paper 39, AGARD Symposium on 'Special Topics in HF propagation', Lisbon.

PAVEY, N.A.D., 1973, "Radiation characteristics of HF notch aerials installed in small aircraft". AGARD Conference Proceedings CP-139.

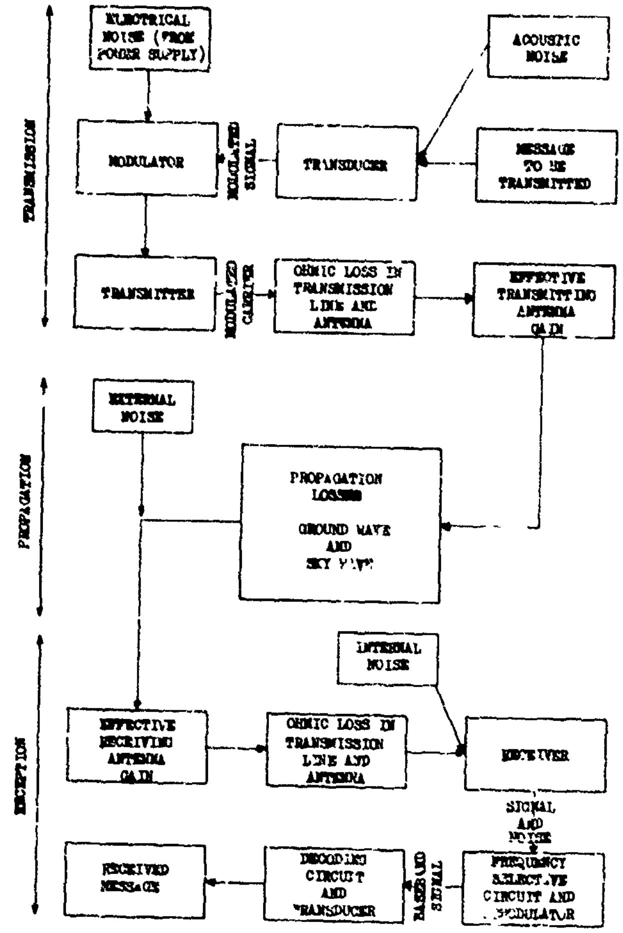

TRANSMISSION
ELECTRICAL NOISE (FROM POWER SUPPLY)
ACOUSTIC NOISE
MODULATOR
MODULATED SIGNAL
TRANSDUCER
MESSAGE TO BE TRANSMITTED
TRANSMITTER
MODULATED CARRIER
OHMIC LOSS IN TRANSMISSION LINE AND ANTENNA
EFFECTIVE TRANSMITTING ANTENNA GAIN
PROPAGATION
EXTERNAL NOISE
PROPAGATION LOSSES
GROUND WAVE AND SKY WAVE
RECEPTION
INTERNAL NOISE
EFFECTIVE RECEIVING ANTENNA GAIN
OHMIC LOSS IN TRANSMISSION LINE AND ANTENNA
RECEIVER
SIGNAL AND NOISE
RECEIVED MESSAGE
DECODING CIRCUIT AND TRANSDUCER
BASEBAND SIGNAL
FREQUENCY SELECTIVE CIRCUIT AND DEMODULATOR

Fig.1 Schematic diagram of a communications link

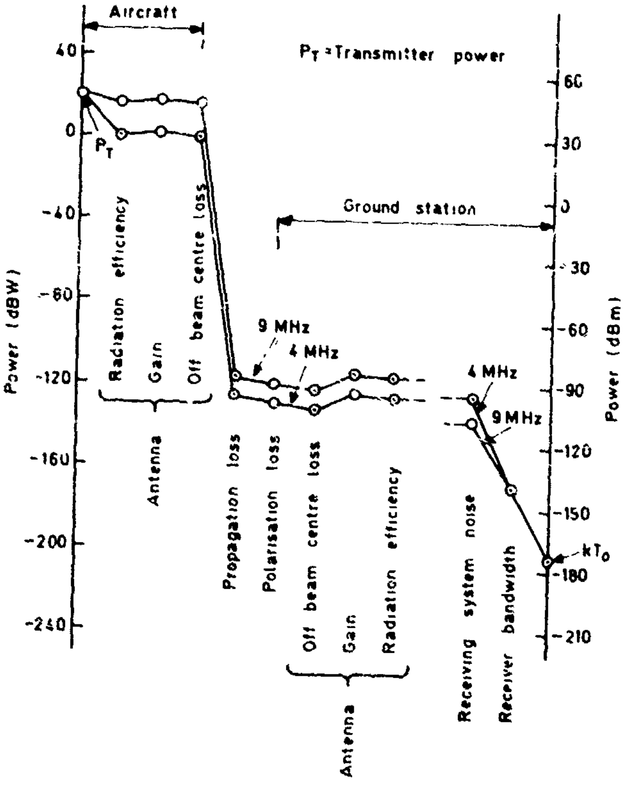


Fig.2 Power level diagram for air-ground link, range 2000 km, low sunspot number, winter midnight

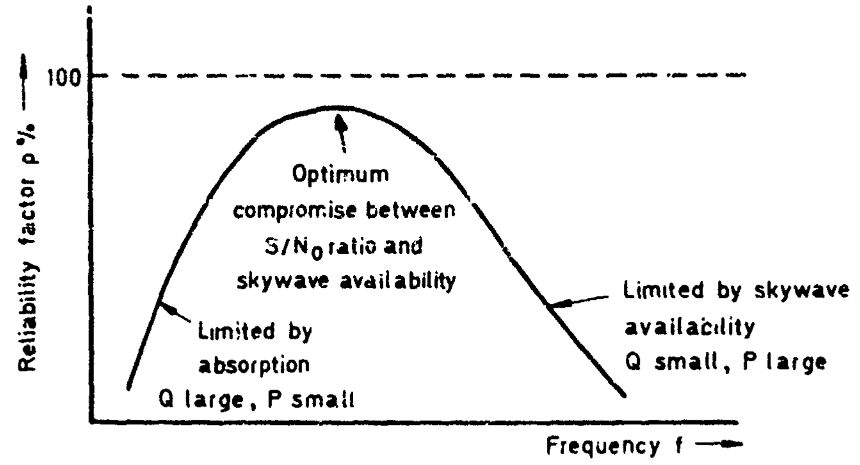


Fig.3 Reliability factor as a function of wave frequency

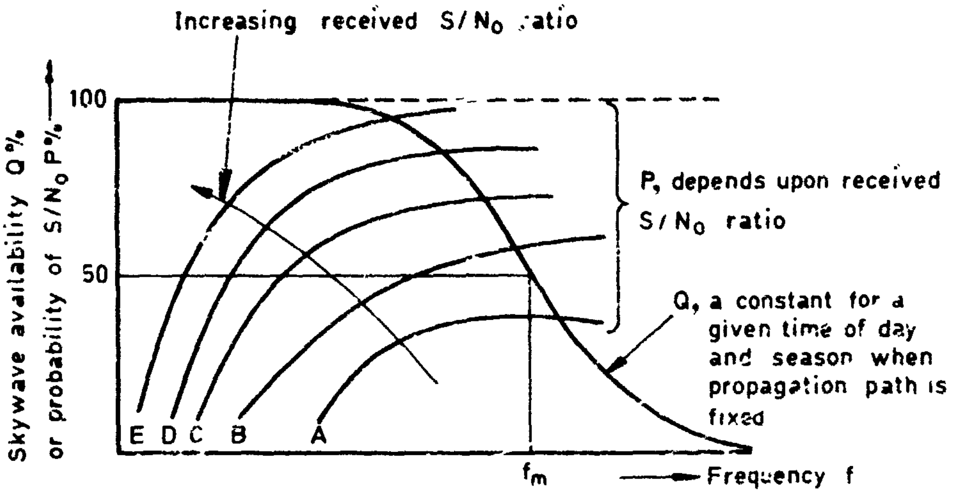


Fig 4 Dependence of Q and P on the wave frequency and the received signal-to noise ratio respectively

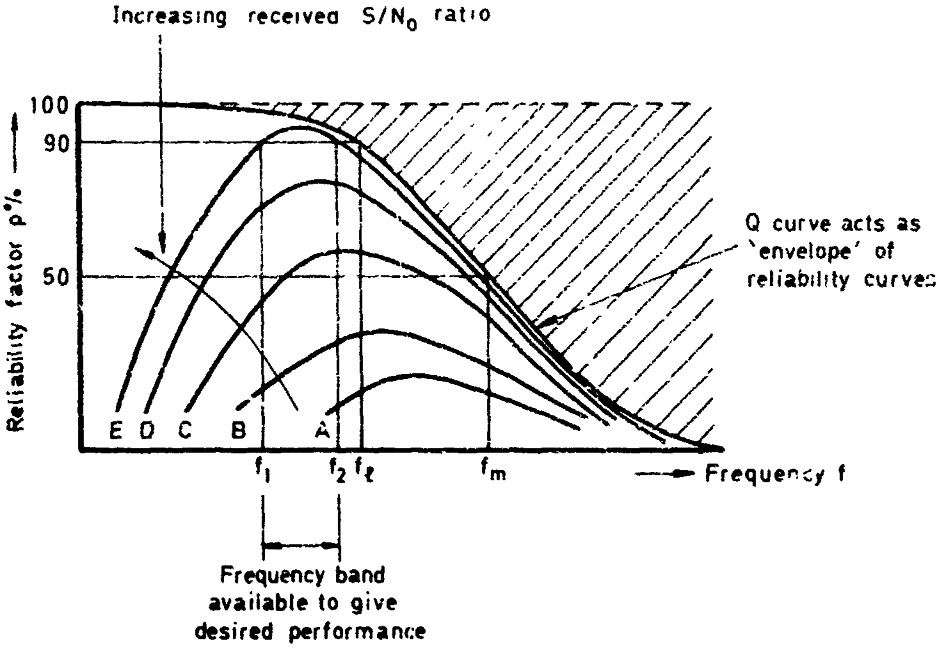


Fig.5 Reliability factor as a function of wave frequency and received signal-to-noise ratio

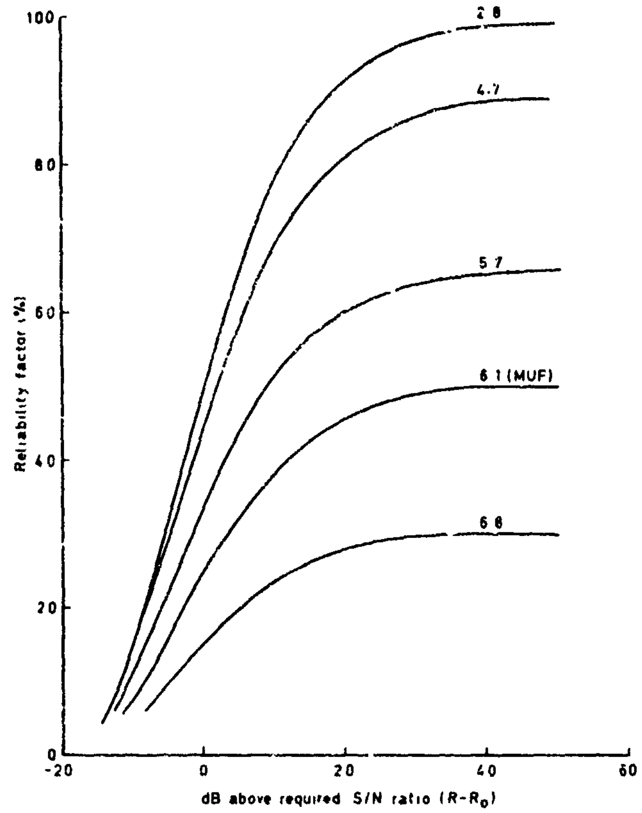


Fig.6 Reliability improvement curves for different frequencies (MHz) at 0300 hours Norway-UK link February 1976

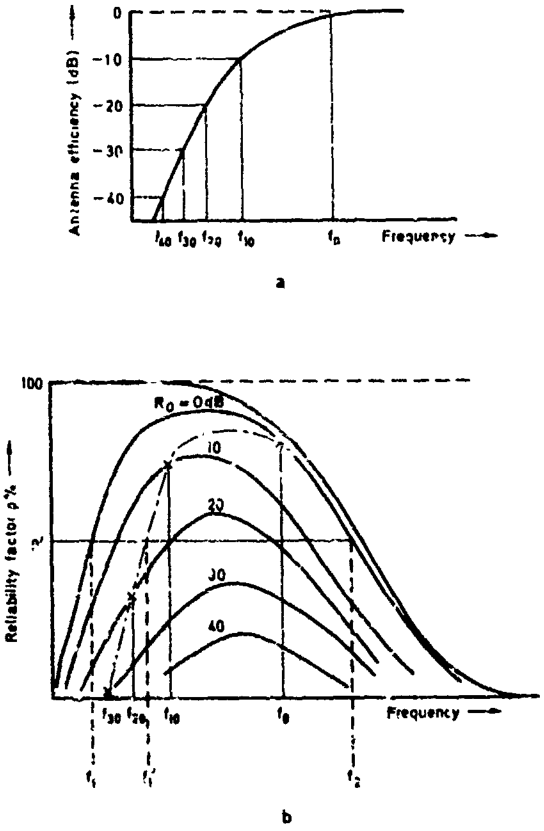


Fig. 7(a & b) (a) Schematic variation of aircraft antenna efficiency with wave frequency, (b) schematic variation of reliability factor with wave frequency. The chain curve shows the modified $R_0 = 0$ dB curve when the antenna efficiencies of (a) are used

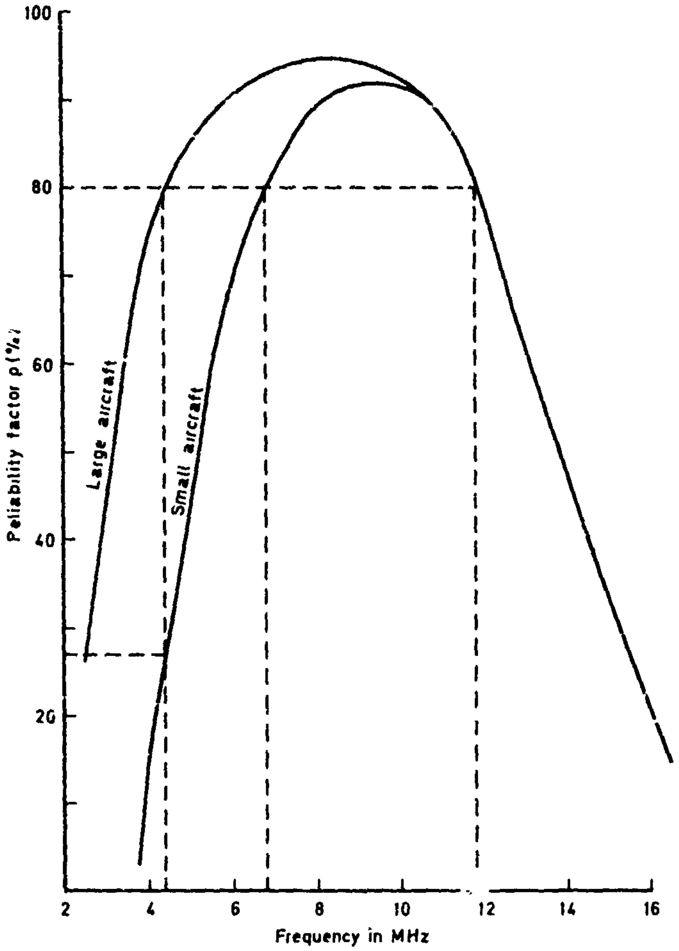


Fig.8 The predicted effect of aircraft antenna efficiencies on circuit reliability at 0900 hours: Norway-UK link February 1976

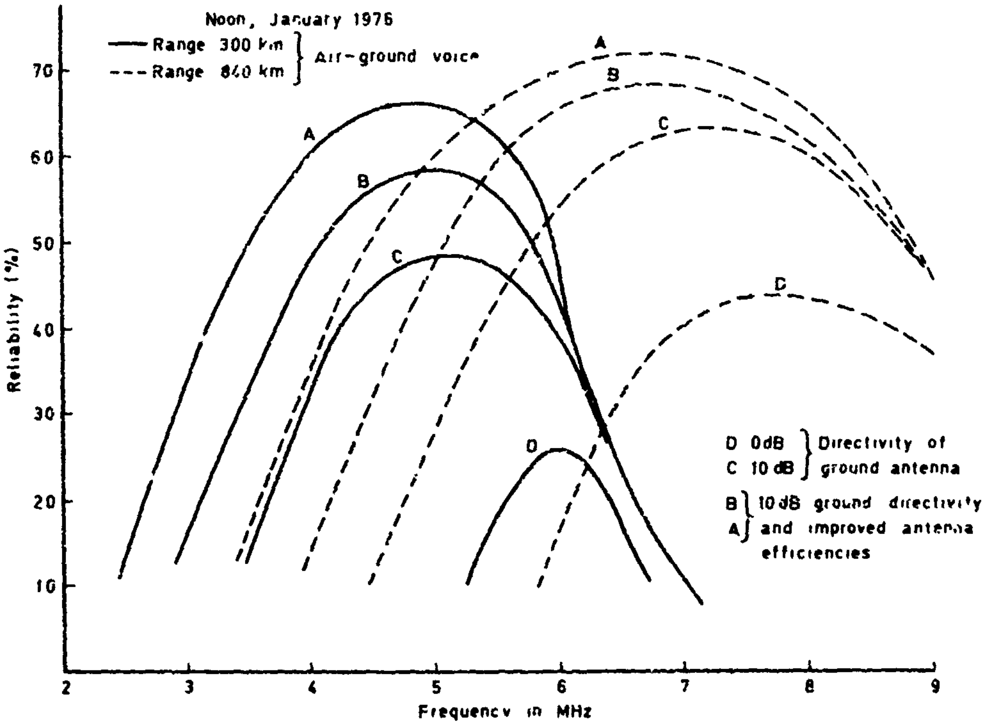


Fig.9 Variation of reliability with frequency for a small aircraft

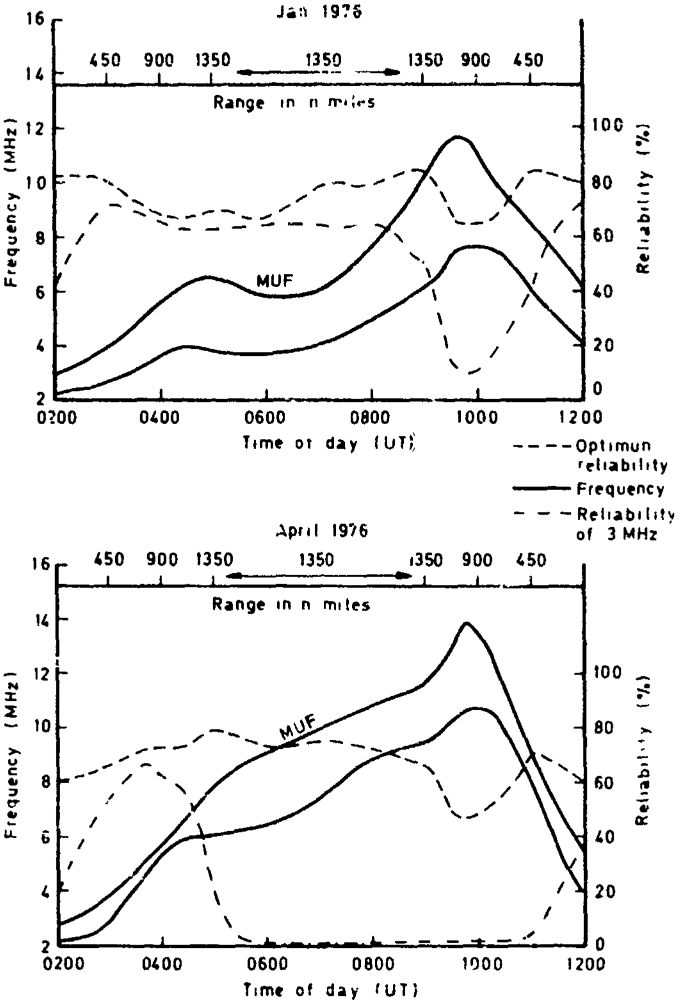


Fig.10 Flight profiles of a large aircraft at dawn

DISCUSSION

Comment by Dr Klaus Bibl, US

(a) A distribution is necessary if momentary MUF values are available on only monthly averages since geometric focussing near MUF is of great importance for signal-to-noise calculations

(b) Interference measurements and calculations dependent on height of aircraft and radio activity areas are important for HF communication reliability assessments, as our results in the airborne ionosondes have shown

Author's Comment

Yes we do agree with these two comments and in connection with the second we are proposing to undertake a measurement program as we shall mention in a later paper at this meeting.

COMPARISON OF MEASURED AND PREDICTED MUF'S AT A REMOTE LOCATION

Th. Damboldt
FORSCHUNGSINSTITUT DER DEUTSCHEN BUNDESPOST
P.O. BOX 5000
D-6100 DARMSTADT

SUMMARY

During the past years measurements of the signal strength of distant HF transmitters have been carried out at Lüchow, F. R. of Germany, by the Research Institute of the Deutsche Bundespost. The times at which the signals are received first and last by ionospheric propagation during the diurnal cycle define the times where the frequency of the transmitted signal is equal to the path MUF (maximum usable frequency). Values of these transition times are taken from the field-strength records of a path from England to Germany and then compared with the predicted values. A further comparison is made between the observed MUF transition times and the ionogram critical frequencies at the approximate path mid-point, i.e. de Bilt in Holland for the above mentioned path.

Good agreement is obtained between the ionogram critical frequencies and the measured MUF's. It is suggested that information obtained from signal-strength records can lead to substantial improvements in MUF predictions. This is of considerable importance since operational requirements usually exist for real-time situations and not for the average conditions computed by prediction programmes.

1. INTRODUCTION

The basis for this "passive sounding systems" is the continuous measurement of the signal strength of distant HF transmitters. A number of meteorological broadcast stations transmitting 24 hours a day is suitable for this purpose. For the work presented here, the station "Bracknell" in England was chosen which transmits on several frequencies simultaneously. The field strength of these transmitters is recorded on the following frequencies: 18.2 MHz, 14.4 MHz, 11.0 MHz, 8.0 MHz, 4.8 MHz and 3.3 MHz. The path length is about 800 km. Located within a few kilometres of the path mid-point is the ionosonde station de Bilt (Figure 1).

2. F-LAYER AND SPORADIC-E LAYER MUF TRANSITS

A typical example of a field-strength record with the sudden increase in field strength is shown in Figure 2. The scatter field strength is about 20 dB above the noise level. The MUF transit time denotes the time at which the propagation mechanism changes from scatter to ionospheric reflection or mode propagation. At this point the field strength increases by about 40 dB (Damboldt and Süßmann, 1976).

The gradient of the field-strength increase at a fixed frequency is dependent on the gradient of the critical frequency in the course of the day. Figure 2 shows a typical winter-day record, where the diurnal variation of the critical frequency is much more pronounced than on a summer day.

Generally a MUF transit is much less distinct when the sporadic-E layer MUF equals the path MUF. Figure 3 shows a typical example of a field-strength increase caused by a short enhancement of the sporadic-E critical frequency. An exact MUF transit time is difficult to determine and due to the sporadic behaviour of this layer, information about the critical frequency on one path can not be used to estimate it on another path.

3. COMPARISON OF PREDICTED MEDIAN MUF WITH OBSERVED MUF TRANSITS

The vertical critical frequencies and the MUF factor were taken from the CCIR Atlas of ionospheric characteristics (CCIR, 1978). They were then converted to an 800-km path with the transmitter at Bracknell and the receiver at Lüchow. The upper decile, median and lower decile MUF's are shown in Figure 4 for October 1978. Also shown is a number of MUF transit times at the frequencies recorded during this month. Good agreement between predicted and measured values is obtained, however the spread of the daily values - due to different propagation conditions - is obvious. Therefore, the daily measured values of the MUF transit times can be used to update the MUF curve in real time (Beckwith and Rao, 1975; Jones, Spracklen and Stewart, 1978).

4. COMPARISON OF MEASURED IONOGRAM CRITICAL FREQUENCIES WITH OBSERVED MUF TRANSIT TIMES

The ionosonde station de Bilt makes hourly measurements of the vertical critical frequency and other ionospheric parameters. Figure 5 shows the diurnal variation of the F-layer critical frequency on the day indicated. Also given are the times of MUF transits observed on the 800-km path at the frequencies indicated. It is obvious that the ionogram vertical critical frequencies can very well be inferred from the observed MUF transits and vice versa.

Figure 6 shows the diurnal variation of the F-layer vertical critical frequency (of the day indicated) of the ionosonde at Lindau, F. R. of Germany. This station is located about 200 km south of the receiving station. Unexpectedly good correlation between ionogram critical frequencies and MUF transits is obtained although the ionosperic reflection point for the oblique path is about 500 km distant from the ionosonde station. Under quiet ionospheric conditions there seems little difference between the critical frequencies at the two mentioned ionosonde stations which are about 500 km apart (Figures 5 and 6).

5. CONCLUSION

The conclusion which can be drawn from thes observation results is that - at least in a limited geographical region - measurements of ionospheric critical frequencies on one path (or at one location) can be used to update the critical frequencies for another path (or another location) in real time.

6. REFERENCES

Beckwith, R.I. and N. Narayana Rao, (1975) Real-time Updating of Maximum Usable Frequency Predictions for HF Radio Communications, IEEE Transactions on Communications Vol 23, 286-288

C.C.I.R. (1978), CCIR Atlas of ionospheric characteristics, Report 340-3, Geneva

Damboldt, Th. and P. Süßmann (1976), Erweiterung von Funkprognosen für kurze Entfernungen auf Frequenzen oberhalb der MUF, Kleinheubacher Berichte, 19, 527-532

Jones, T.B., C.T. Spracklen and C.P. Stewart (1978), Real-time updating of MUF predictions, AGARD Conference Proceedings, C.P. - 238, Paper 17

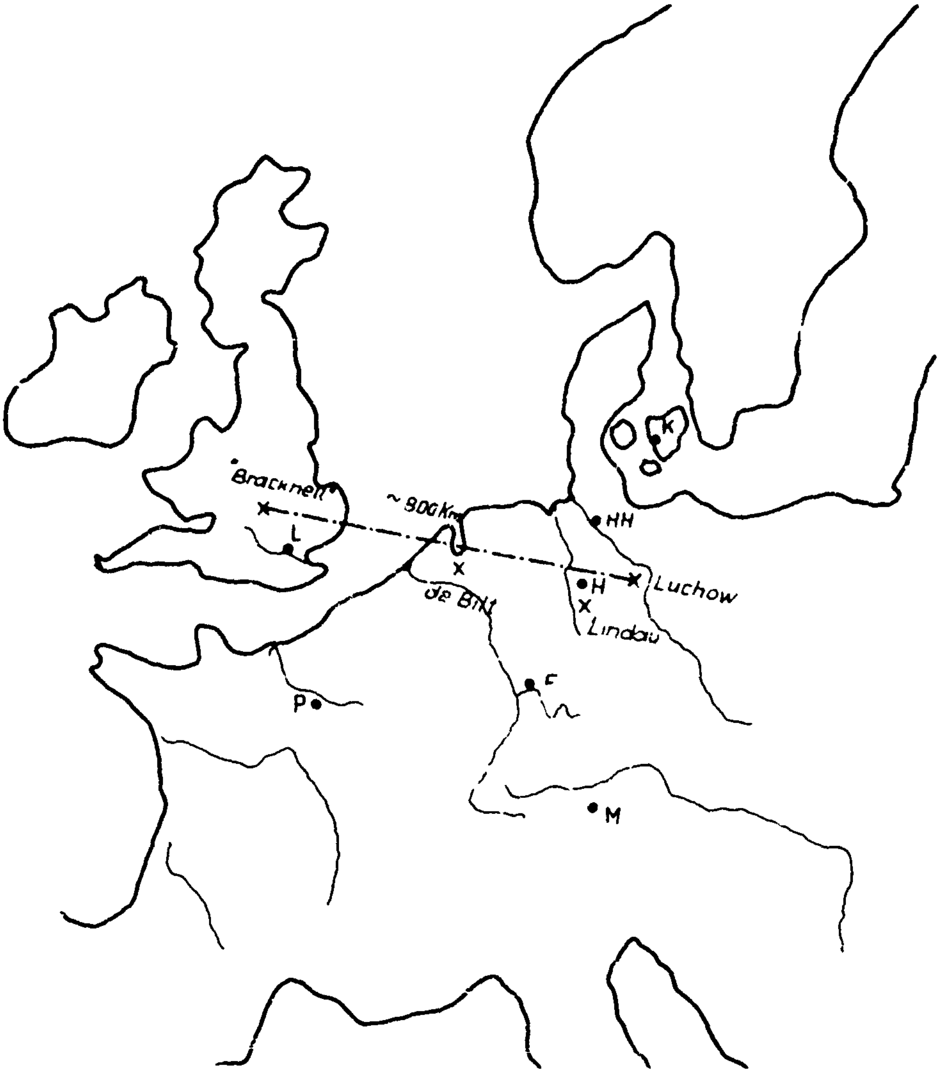

Bracknell
~800Km
L
x
de Bilt
HH
H
Luchow
x
Lindau
K
P
F
M

Fig. 1 Map of Europe showing the locations of transmitter, receiver and ionosonde stations.

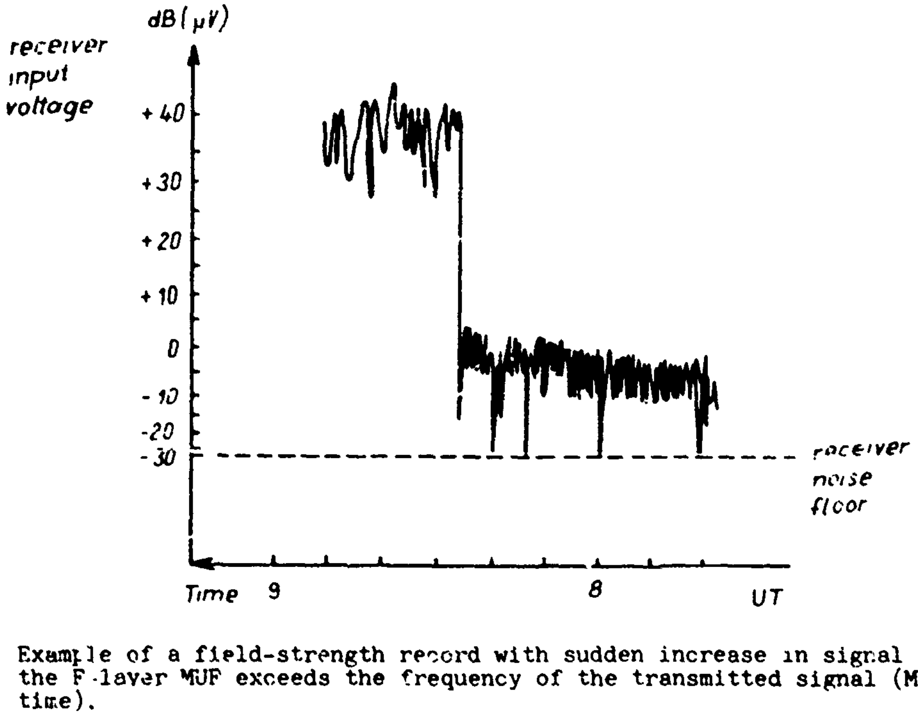


Fig. 2 Example of a field-strength record with sudden increase in signal strength when the F-layer MUF exceeds the frequency of the transmitted signal (MUF transit time).

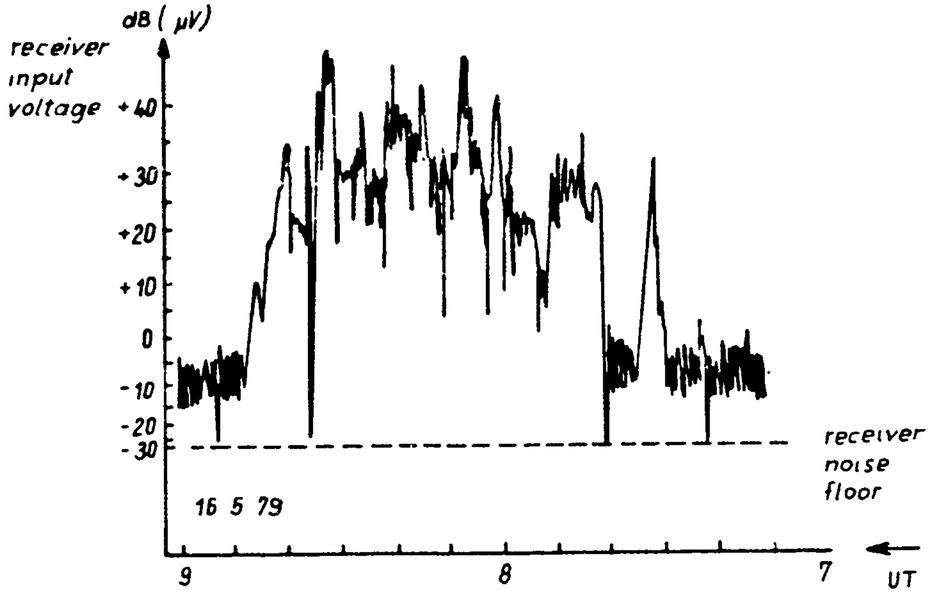


Fig. 3 Example of E_s-layer MUF increase

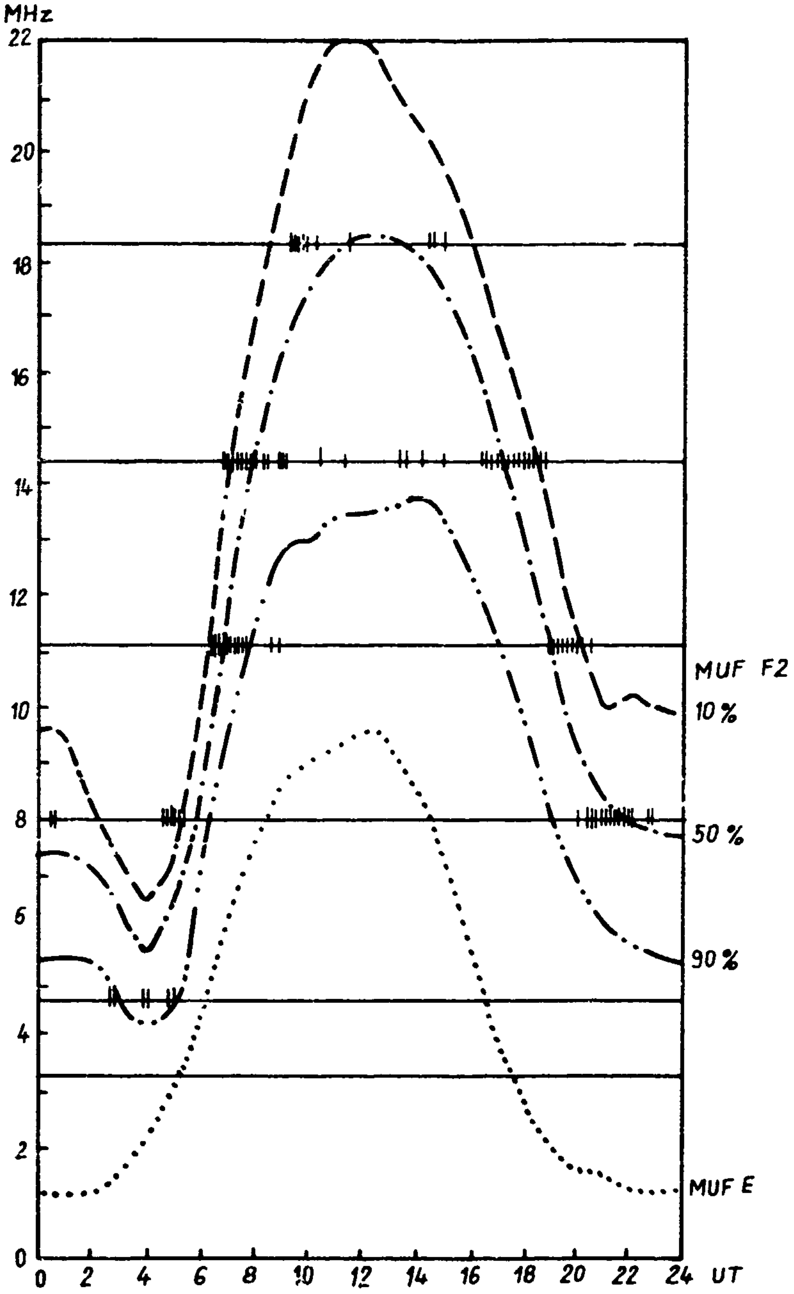


Fig. 4

Maximum usable frequencies for the 800-km path, showing median and decile values

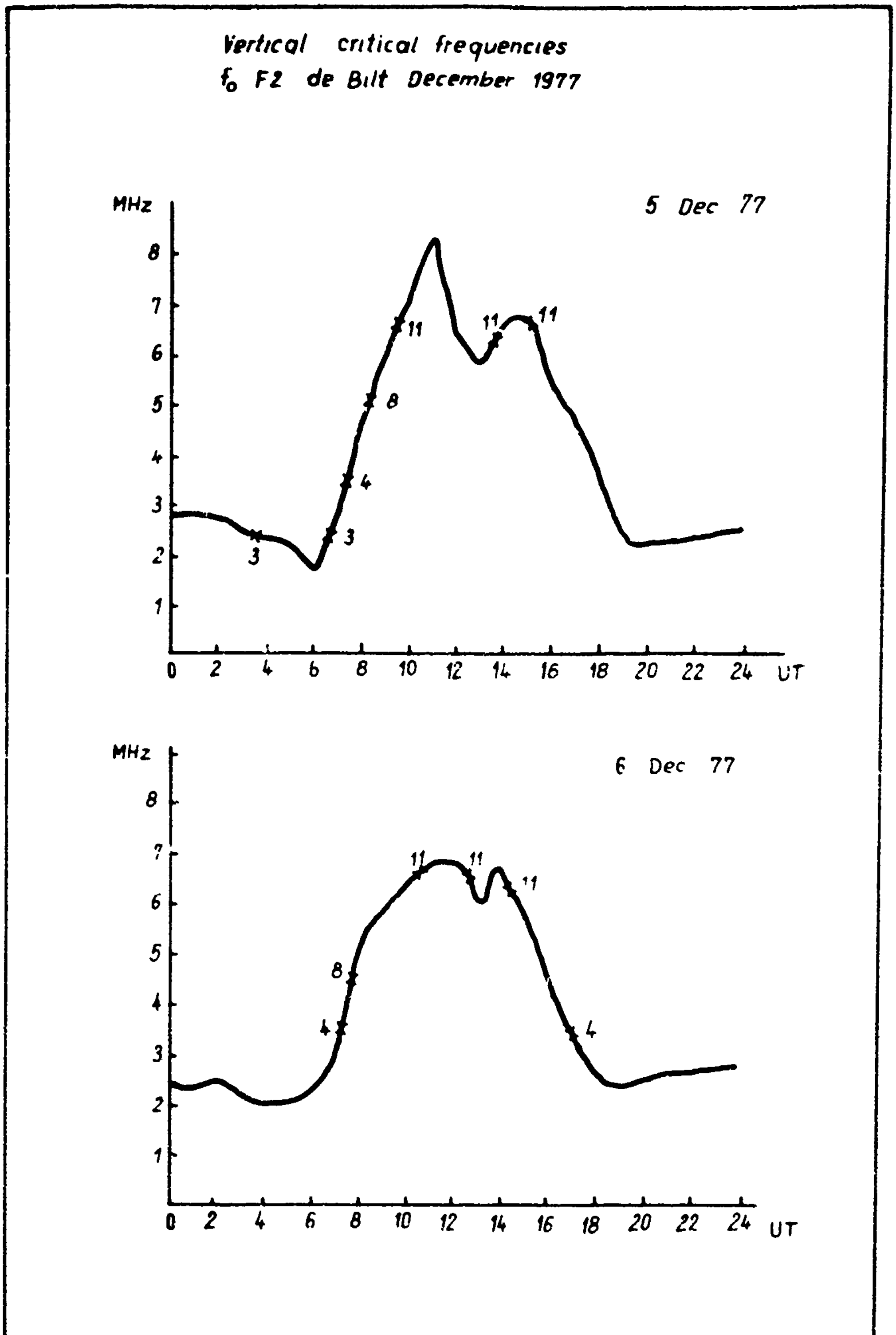
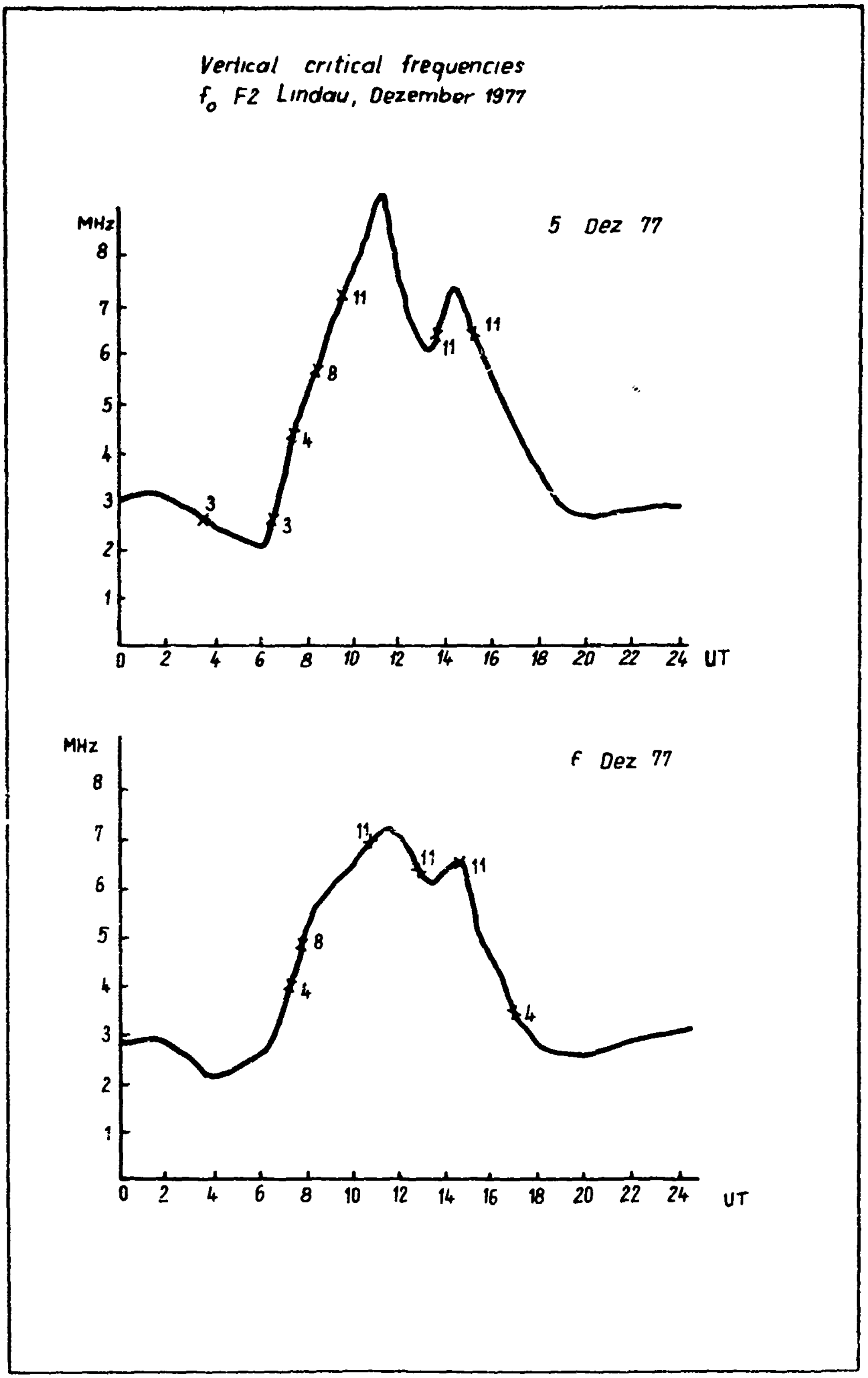


Fig. 5 Vertical critical frequencies measured during one day at de Bilt. The times of MUF transits are indicated by "x" together with the frequency where the MUF transit was observed



FTZ Nr 2 55

Fig. 6 Vertical critical frequencies measured during one day at Lindau. The times of MUF transits are indicated by "x" together with the frequency where the MUF transit was observed.

AUGMENTATION OF HF PROPAGATION

T. J. Barrett
Mission Research Corporation
Post Office Drawer 719
Santa Barbara, California 93102

SUMMARY

This paper describes a method to augment HF communications systems.

It appears that chemical releases might be employed to scatter communication signals for limited critical periods. This technique is an augmentation to be used in the event of an emergency, or an apparent emergency.

The technique, while short-lived, does provide communication coverage over vast areas (~3000 mile range). In a nuclear environment, the chemical ion cloud allows the use of VHF transmissions, rather than HF, thus greatly reducing blackout durations.

Assessing the nuclear environment provides information to avoid blackout caused by beta particle ionization and allows choice of optimum cloud location.

This technique does not replace any current communication system but does provide a way to maintain communications for a few minutes which might otherwise be blacked out for hours.

Potential applications of this technique include:

- Communication with NATO
- Communication with Fleets
- Communication with Subforces
- Communication with SAC
- Tactical Communications

1. INTRODUCTION

HF communications currently play an important role in U.S. national defense. Due to the characteristics of HF systems and the earth's ionosphere, such systems provide a unique method of communication between points on the earth's surface which cannot communicate via line-of-sight transmission. Unfortunately, such communications are dependent on the state of the ionosphere. Frequently, at night, there are large regions in the northern hemisphere which cannot support communications from the U.S. to Europe and other areas. Uncertainties in future geomagnetic activity and solar flux variations lead to unreliability over many HF links. Because of these uncertainties and the lack of coverage even on normal days/nights, the U.S. is currently considering satellite communications as a primary link between the United States and NATO forces in Europe.

Satellite communication systems may be susceptible to degradation when forced to operate in nuclear environments and may also be subject to degradation when propagation paths pass through natural aurora. The primary degradations are fluctuations in signal amplitude, phase, and angle-of-arrival caused by propagation through regions of ionization which are aligned with the geomagnetic field and which perturb the incident wave to cause severe fluctuations. Satellites also may be vulnerable to the radiations produced by high altitude nuclear detonations.

Data have shown that chemical releases produce ion clouds which allow the side, forward, and backscatter of HF and VHF waves with reflection coefficients near unity. Clouds of 5 km radius, or more, have been generated using barium vapor released at an altitude of about 200 km. Such clouds have scattering cross sections on the order of $10^8 m^2$. These cross sections allow the transmission of communications signals over ranges of 3000 miles for periods on the order of several minutes.

Communications using chemical release clouds does not replace any current system. They provide a backup system, a means of communication during emergencies. The feasibility and utility of this communication scheme is the subject of this paper.

2. HF GEOMETRY AND FREQUENCIES

Figure 1 shows a simple sketch of an HF propagation path from a transmitter (T), refraction of the ray caused by ionization in the E- and F-region, and subsequent bending of the ray back to the receiver (R) on the earth. Figure 2 shows the optimum frequency to be used for transmission of HF signals between Fort Huachuca, Arizona and Western Europe over 24-hour periods in June and December. Note that the maximum optimum frequency was 15 MHz, in June, and that in December, communication between 12:00 and 23:00 was not possible. Figure 3 shows the optimum frequency for communication between Fort Huachuca and Honolulu. Again, the optimum frequency does not exceed 15 MHz, however, communications can be maintained over the entire 24-hour period. Finally, Figure 4 shows transmission characteristics for a near vertical path—Maryland to Massachusetts. While transmission can occur over the 24-hour period, the maximum frequency is only about 7 MHz.

Figure 5 shows an oval-shaped region* located between Ireland, England, and Iceland which cannot be communicated with 95 percent of the time at night from an HF transmitter station located in the State of Maine.

---

* This figure is a simple artist concept.

In short, HF frequencies are generally limited to about 15 MHz for transmissions at long range and perhaps half that value for shorter range. There are large regions in the northern hemisphere to which Conus-based transmissions cannot propagate at night.

3.0 CHEMICAL RELEASE CHARACTERISTICS

Sodium, cesium, and barium have all been used in past chemical release studies. A great wealth of data was obtained during past tests. Barium was utilized in the daytime with both sunrise and sunset releases. Barium vapor is generated via a thermite reaction with copper or with barium nitrate. Typically, 10 percent of the available barium is vaporized. This vapor is readily photoionized by sunlight. The data base shows that the time history of ionization in the ion clouds behaves as

$$N_e \approx \frac{4 \times 10^7}{t^{1/2}} \text{ cm}^{-3} \tag{1}$$

where N_e is the electron concentration and t is the time after ionization expressed in seconds. Barium released under night conditions will not produce an ion cloud. The initial thermal ionization is rapidly quenched by reactions with O_2. That is, the barium ion is energetic, the oxidation reaction

$$Ba + O_2 \longrightarrow BaO^+ + O \tag{2}$$

occurs rapidly and is followed by the fast dissociative recombination reaction

$$BaO^+ + e \longrightarrow Ba + O \tag{3}$$

In the day, reaction (3) is followed by the competing reactions

$$\begin{aligned} Ba + h\nu &\longrightarrow Ba^+ + e \\ Ba + O_2 &\longrightarrow BaO + O \end{aligned} \tag{4}$$

The oxidizing reaction in Equation 4 is much slower than in Equation 2 because the ion is no longer highly energetic. Ion/electron concentrations described by Equation 1 result from ambipolar diffusion of the cloud. (Ion/electron radiative recombination is too slow to have any effect on N_e).

Barium releases have provided clouds which are large (5 to 30 km radii, depending on release altitude) and which have radar cross sections which approximate geometric values, $10^8 m^2$. Unfortunately, barium ion clouds are only useful as a scattering device while the clouds are sunlit.

Cesium and sodium occupy a different column in the chart of the elements. These elements have only one electron in the outer shell, compared to two electrons for the barium atom. As a result, an energetic ion such as Cs^+ does not readily react with O_2. The thermal ionization created via thermite reactions is not oxidized by O_2 (though it would be oxidized by O_3, and, therefore, is probably not useful at altitudes below about 80 km) and the time history of cesium is given by

$$N_e \approx \frac{N_o}{t^{1/2}} \text{ cm}^{-4}$$

where N_o is the initial concentration of cesium ions generated by thermal ionization. Data shows that N_o for small (1 kg) releases at about 105 kms is on the order of 2×10^7 cm^{-3}. Cesium (or sodium) thus provides ion clouds which possess large scattering coefficients during both day and night conditions.

The initial clouds are spherical. Dispersive spreading and magnetic containing forces gradually distort the initial spherical shapes into ellipsoids with the major axis aligned with the geomagnetic field.

Gradient drift instabilities cause the ion cloud to develop a sharp edge and eventually striate into field-aligned filaments of ionization. At this time, usually several minutes or more, the cloud becomes a diffuse scatterer and may no longer be useful in terms of communication augmentation.

4.0 SCATTERING CHARACTERISTICS AND GEOMETRY

The plasma frequency of an ion cloud can be expressed as

$$f_p = 8.9 \times 10^3 N_e^{1/2} \text{ Hz} \tag{6}$$

where N_e is the electron concentration (cm^{-3}). Using Equations 1 and 5, the plasma frequencies for barium and cesium ion clouds are

$$f_{PB} = \frac{5.7 \times 10^7}{t^{1/4}} \text{ Hz} \tag{7}$$

$$f_{PC} = \frac{8.9 \times 10^3 N_o^{1/2}}{t^{1/4}} \text{ Hz} \tag{8}$$

Equations 7 and 8 show the frequencies which will backscatter from the cloud. That is, radio waves of any frequency f such that $f \leq f_p$ will backscatter. At angles other than backscatter, an additional factor secant θ (θ is the angle between the incoming wave vector and the normal to the surface), multiplicatively increases the usable frequency. Taking $N_o = 10^8$ cm^{-3} in Equation 8 and a scattering angle of 45 degrees, the maximum usable frequency versus time for both barium and cesium ion clouds has been computed. The results are shown in Figure 6. Note that for the cesium cloud, a frequency of 30 MHz can be used for approximately five minutes.

Figure 7 shows the backscatter radar cross sections measured during a barium test. Note that the cross section exceeds 10^8 m^2 at 10 3 MHz for about 18 minutes after release.

Figure 8 shows a propagation link between a transmitter, T, an ion cloud located above the E-layer, and a receiver, R, over-the-horizon from the transmitter. During a test series an ion cloud was released over Florida at about 200 km altitude. Transmissions from White Sands, New Mexico were scattered off the cloud and received at Antigua in the Carribean Sea. The range from transmitter to receiver was about 3000 miles.

Figure 9 shows four circles with diameters of 3000 miles. Note on the lower circle that White Sands and Antigua lie on the periphery, with the ion cloud over Florida at the center of the circle. The other three circles show the geometric coverage which could be provided by ion clouds, for transmitters located at Seattle, Washington, Omaha, Nebraska; and Bangor, Maine. The entire U S Northern boundary and up to the North Pole can be covered with three ion clouds.

Figure 10 shows a 3000 mile diameter circle centered at Rome, Italy. Note that communications can be made from London, England to Istanbul, Turkey. The entire NATO community can communicate using a single cloud.

5.0 NUCLEAR ENVIRONMENTS

HF systems subjected to nuclear environments are extremely susceptible to blackout due to the absorption produced by X ray, gamma-ray, and beta-particle ionization. Figure 11 shows an artists concept of a large yield nuclear burst occurring at an altitude of 100 km directly above a transmitter which is propagating a wave up toward a chemical release cloud. The range from the burst to the region of ionization is about 200 km. Figure 12 shows the one-way vertical absorption* produced during the day by X-ray ionization.

Note that 10 dB absorption at 10 MHz persists for about 25 minutes while at 50 MHz, only for about one minute. Clearly, increasing the frequency of propagation produces a large reduction in the duration of blackout. For X-ray blackout, the duration of a given level of absorption scales about as f^{-2}, hence, five-fold increase in frequency results in a 25-fold diminution in duration.

Figure 13 shows the one-way vertical absorption produced by the delayed gamma-ray ionization from this burst. A system operating at 10 MHz would be blacked out for several hours after the detonation. At 50 MHz, the blackout (10 dB) persists for five or 10 minutes depending on ray path geometry.

Figure 14 shows the absorption produced by beta-particle ionization, arbitrarily spread over a 200 km radius. For ray paths which intersect the beta-patch, even systems operating at 100 MHz are blacked out for over one hour. Clearly, beta-patch regions must be avoided by any HF or VHF communication system.

We will now briefly discuss methods of assessing such environments and describe mitigation techniques that could be employed to minimize the effects of nuclear detonations.

6. ASSESSMENT AND MITIGATION

Past studies have shown that electronic sensors can be used to identify a nuclear detonation and locate the burst position with reasonable accuracy. Studies have also shown that electronic measurements can be used to obtain estimates of the weapon yield to within about a factor of two.

With these estimates as inputs, computer simulations can be used to provide estimates of beta patch location as a function of time. In addition, such a simulation can provide estimates of the absorption time histories for any ray path of interest.

---

* To obtain the two-way slant path absorption, multiply by 2 sec E, where E is the ray elevation angle.

REFERENCES

BARRETT, T.J., with others, 1972, "PROJECT SECEDE: A Rewview (U)," Santa Barbara, California, General Electric—TEMPO, DASIAC TN72-3 (Secret).

BARRETT, T.J., with others, 1972, "Structured Ionization Effects on Radar (U)," Santa Barbara, California, General Electric—TEMPO, 72TMP-29 (Secret-Restricted Data) (NoForn).

BARRETT, T.J., with others, 1974, "Bhangmeters (U)," Draft, Santa Barbara, California, General Electric Company—TEMPO, A8303 (Secret/Restricted Data).

LANE, G., 1977, "Output from ITSA-78," Fort Huachuca Headquarters, Arizona.

JORDANO, R.J., with others, 1975, "RANC IV, Modification 3 - Revised Nuclear Phenomenology and Radar Propagation Models for RANC IV (U)," Santa Barbara, California, General Electric Company—TEMPO, GE74TMP-13 (Secret/Formerly Restricted Data)

SLAGER, W.L., 1974, "Nuclear Burst Location and Yield Determination System (U)," Santa Barbara, California, General Electric Company—TEMPO, GE73TMP-11 (Secret/Restricted Data).

THOME, G.D., 1969, "Project SECEDE I - HF Radar Studies of Barium Clouds," Spencer Laboratory, Raytheon Company (Unclassified).

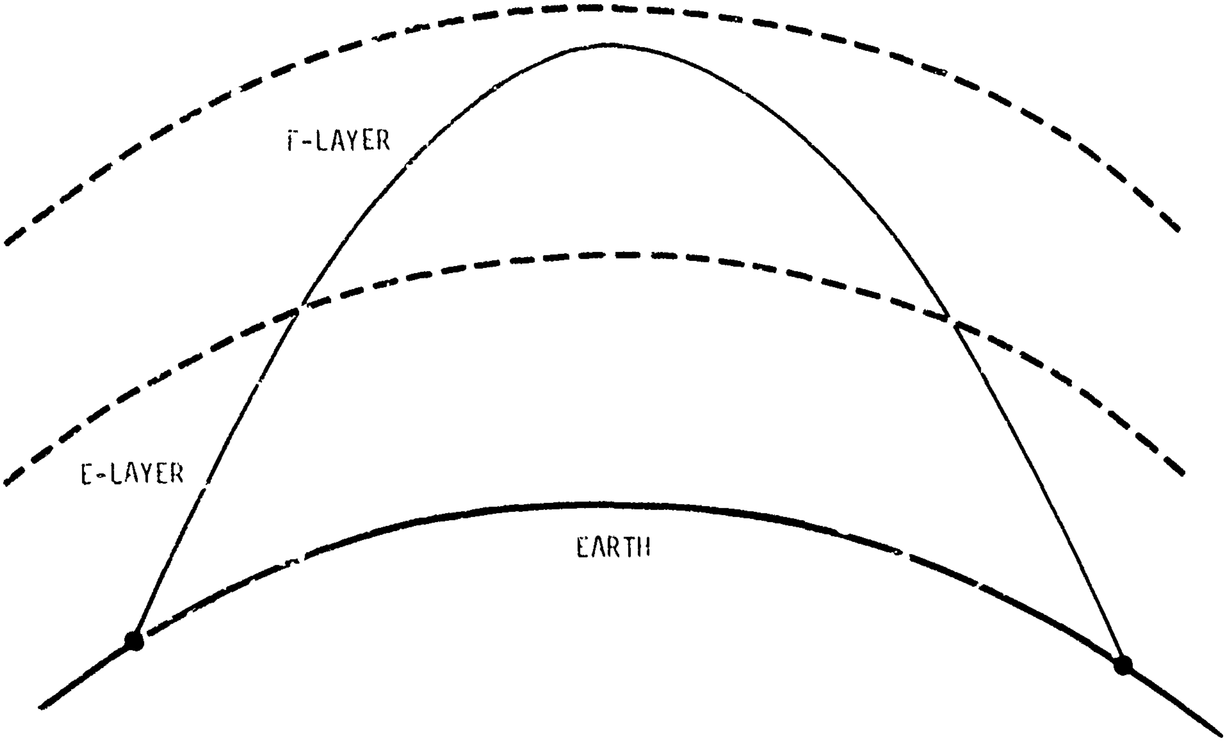

F-LAYER
E-LAYER
EARTH

Figure 1. HF propagation geometry.

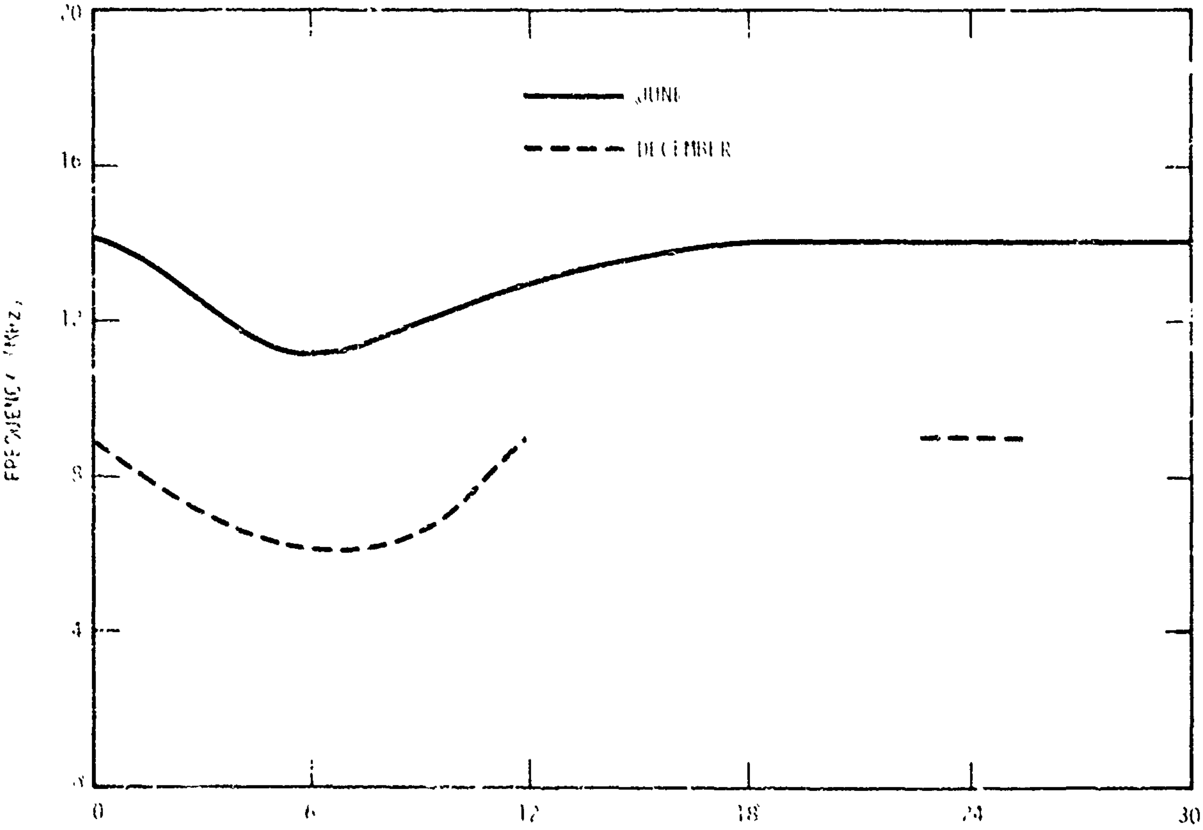


Figure 2. HF optimum frequency from Fort Huachuca to Western Europe.

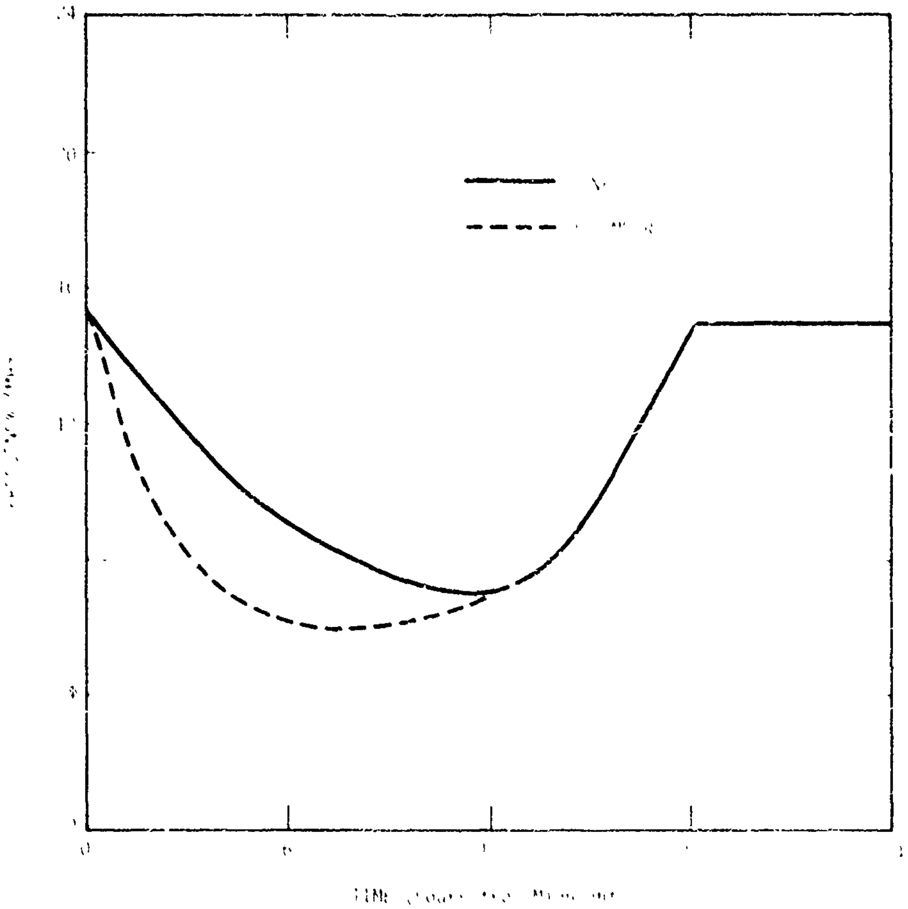


Figure 3. HF optimum frequency from Fort Huachuca to Honolulu

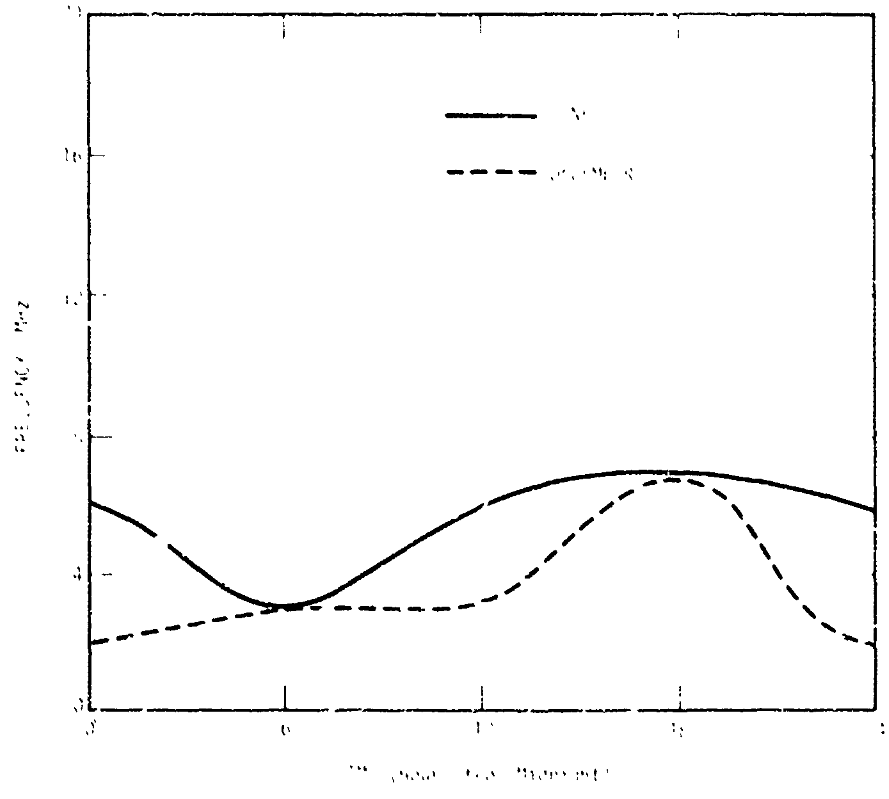


Figure 4. HF optimum frequency from Maryland to Massachusetts.

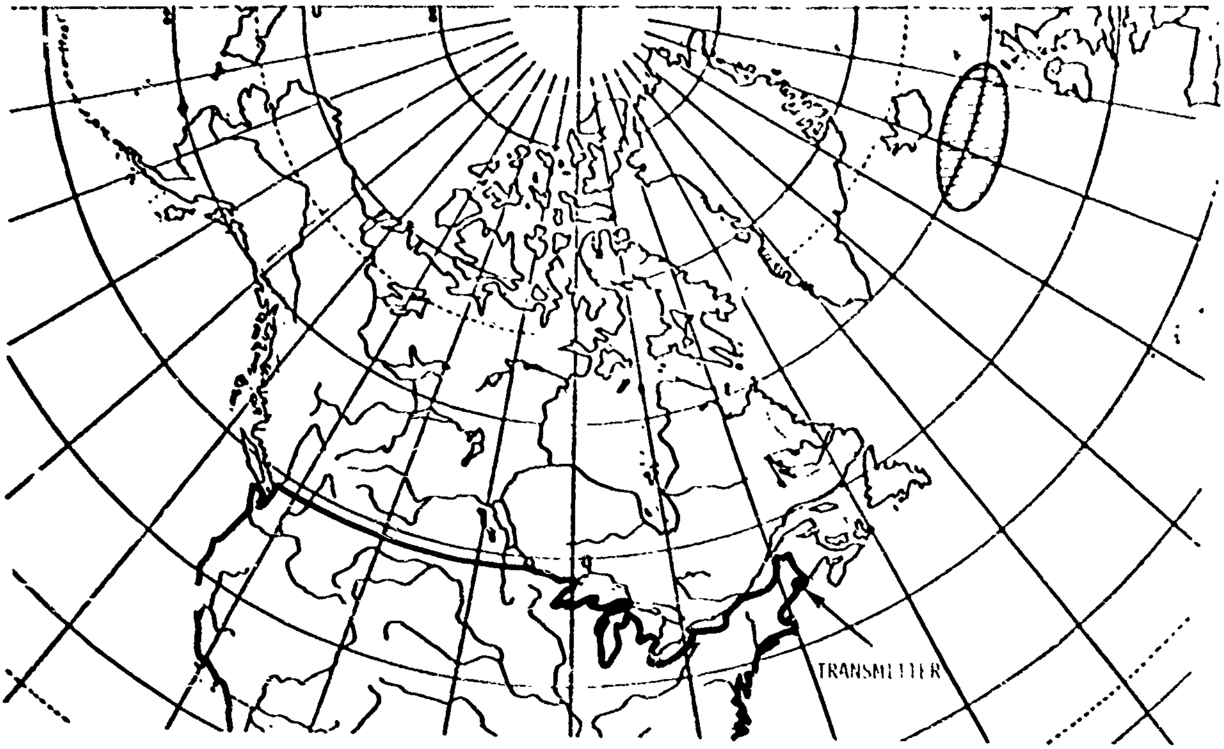

TRANSMITTER

Figure 5. HF outage area, T_x, in Maine.

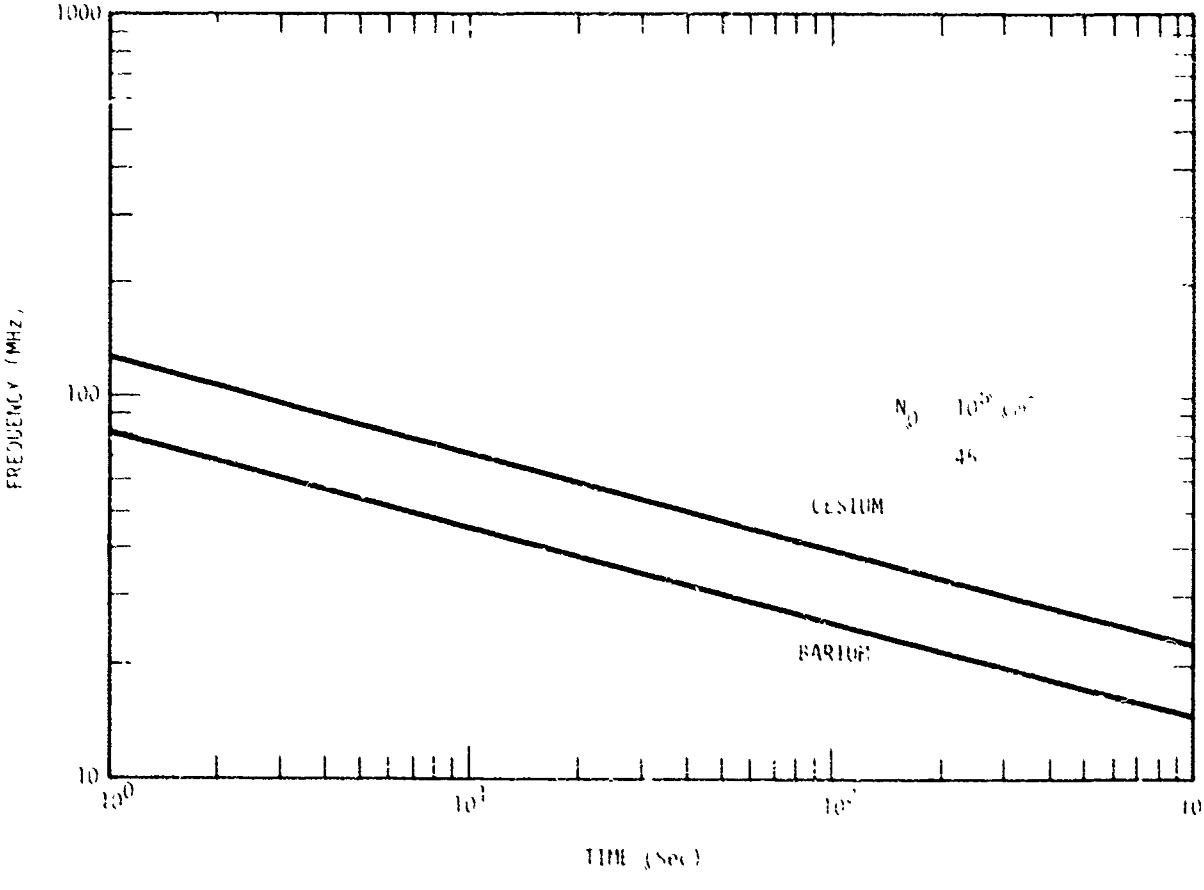


Figure 6. Maximum operating frequency versus time.

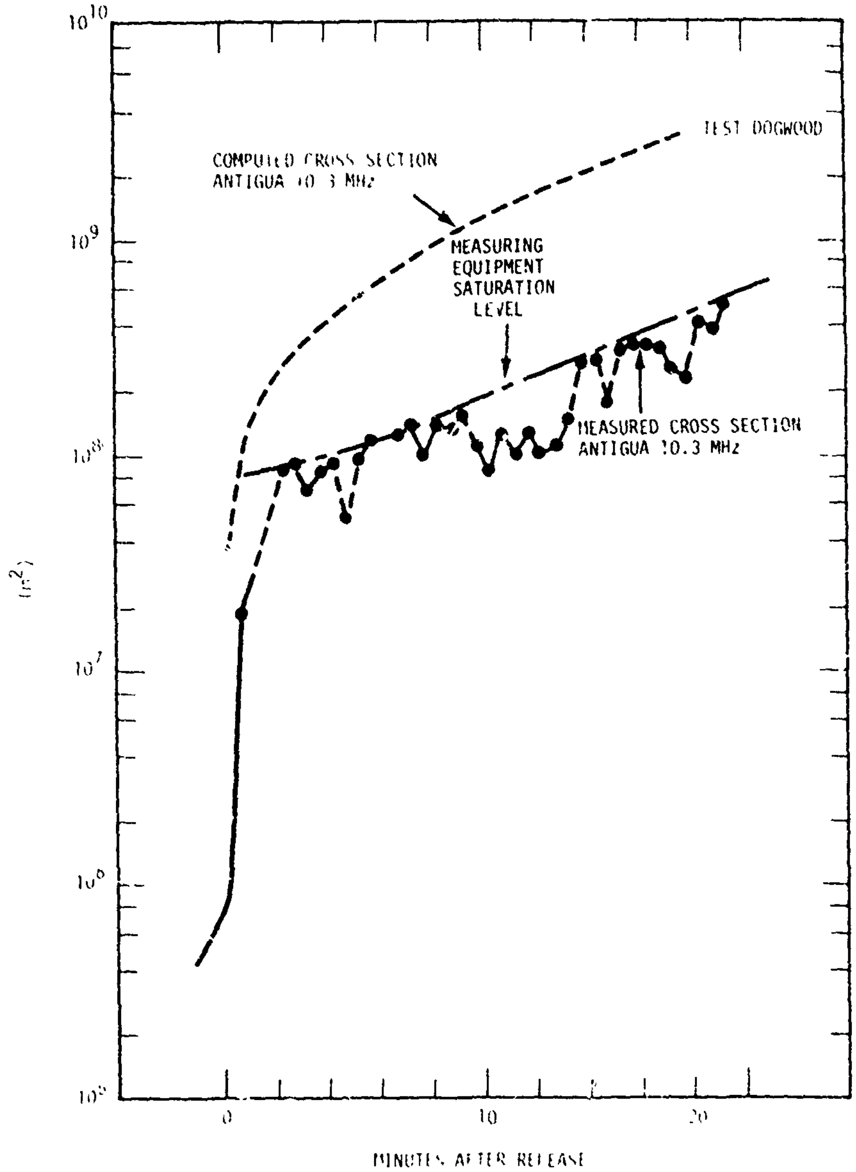


Figure 7. Comparison of experimental cross section with theoretical cross section of a time-varying ellipse.

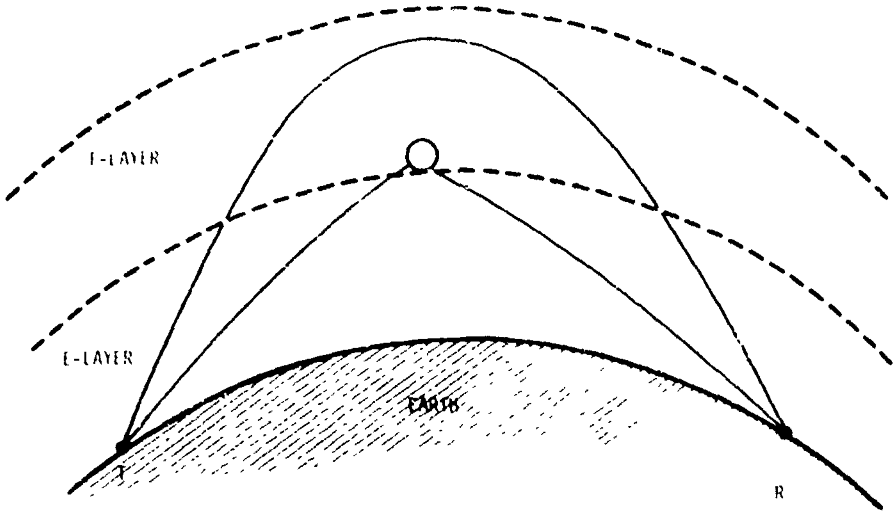

F-LAYER
E-LAYER
EARTH
T
R

Figure 8. HF propagation with chemical release.

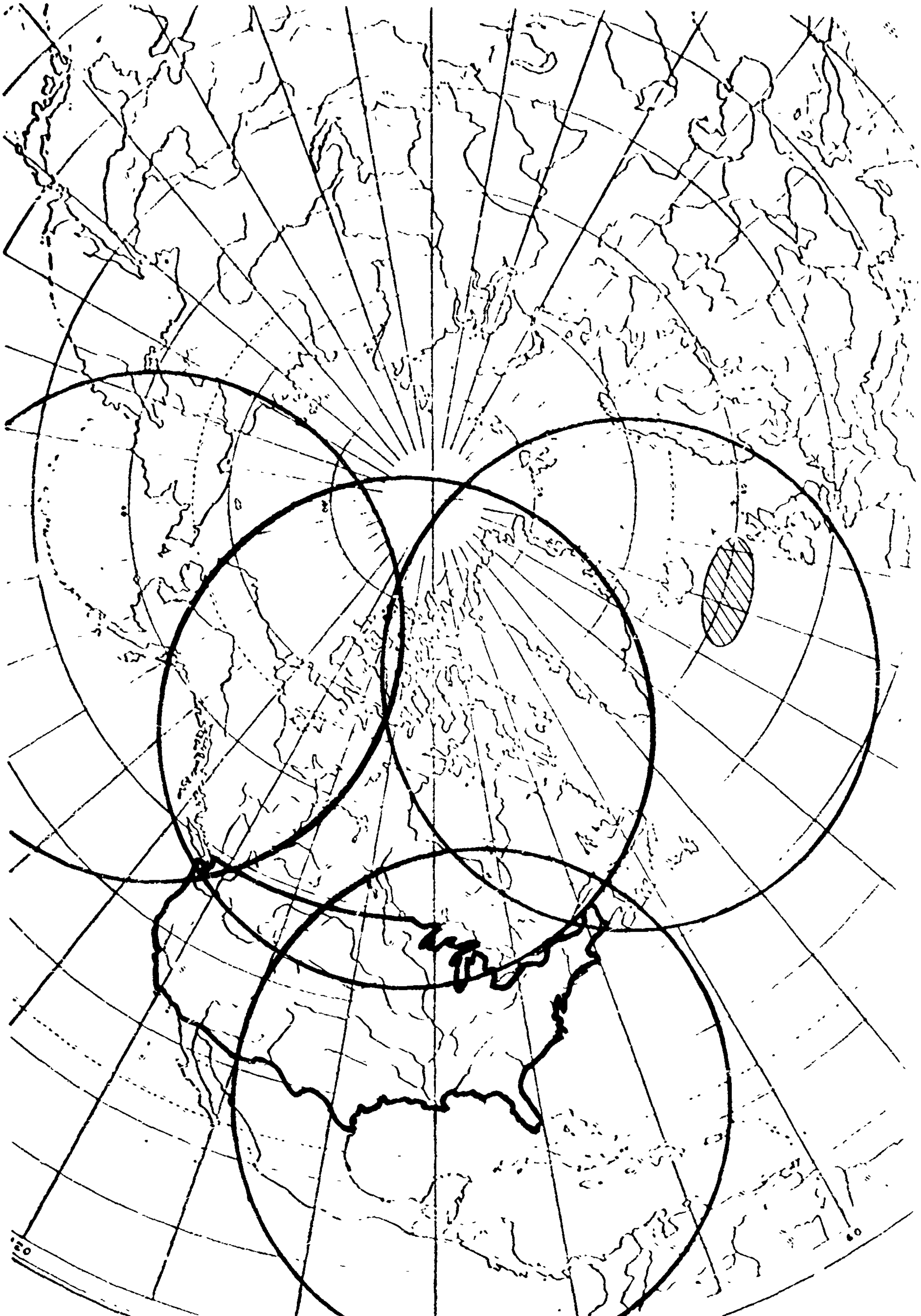

120
105
90
60

Figure [illegible] Contamination coverage using [illegible] release

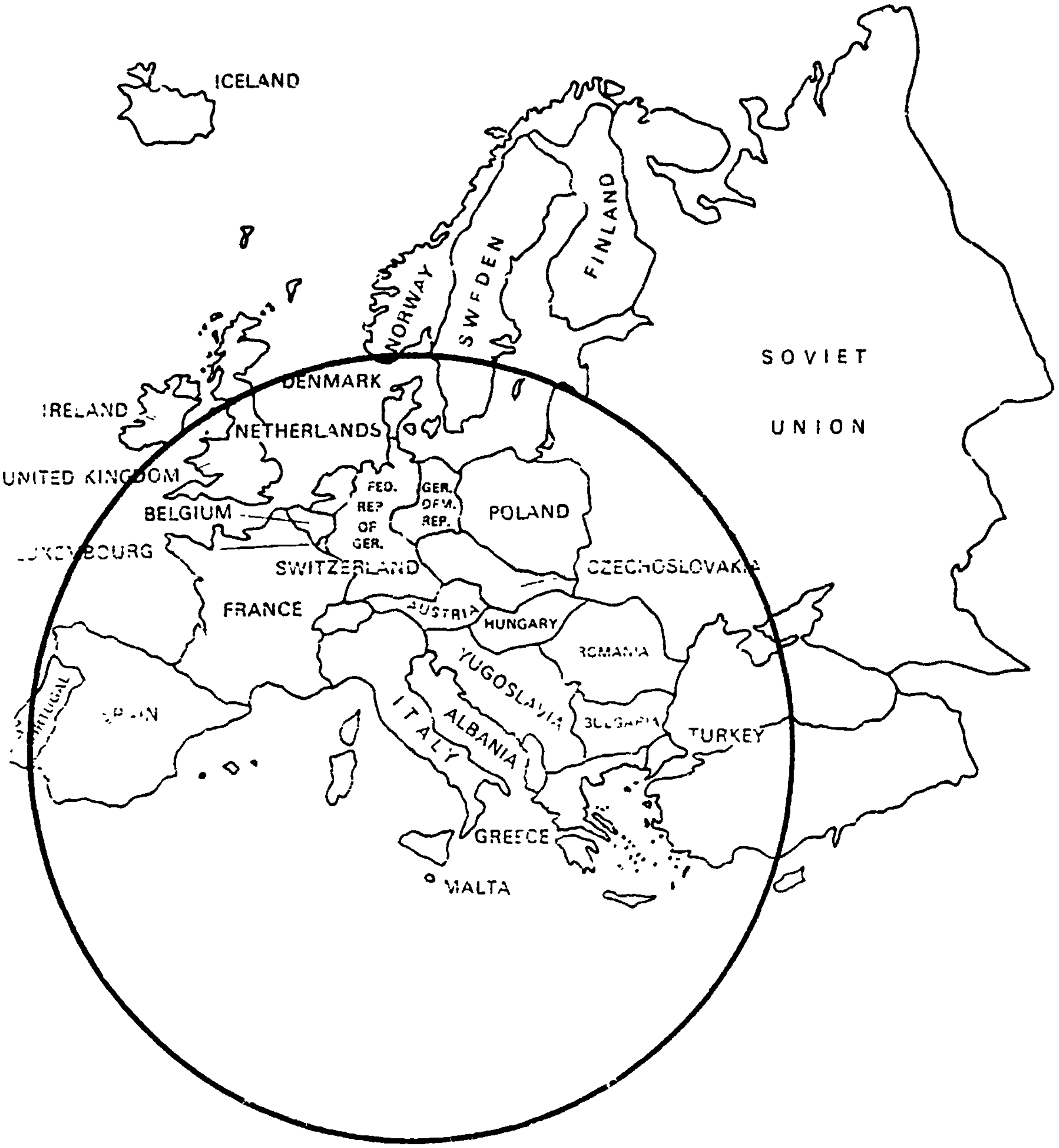

ICELAND
NORWAY
SWEDEN
FINLAND
SOVIET UNION
DENMARK
IRELAND
NETHERLANDS
UNITED KINGDOM
BELGIUM
LUXEMBOURG
FED. REP. OF GER.
GER. DEM. REP.
POLAND
SWITZERLAND
CZECHOSLOVAKIA
FRANCE
AUSTRIA
HUNGARY
ROMANIA
YUGOSLAVIA
PORTUGAL
SPAIN
ITALY
ALBANIA
BULGARIA
TURKEY
GREECE
MALTA

Figure 10. NATO barium cloud scattering footprint.

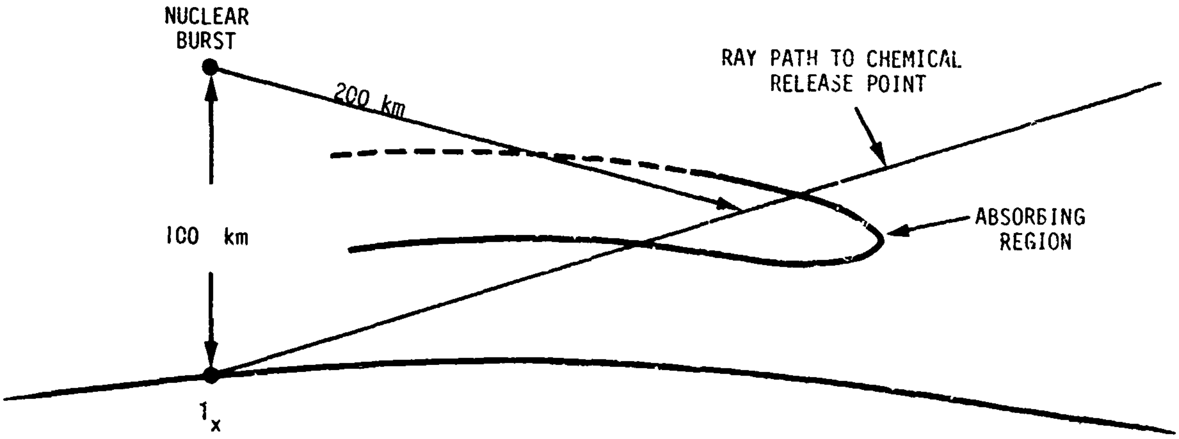

NUCLEAR BURST
200 km
RAY PATH TO CHEMICAL RELEASE POINT
100 km
ABSORBING REGION
1_x

Figure 11. Burst geometry.

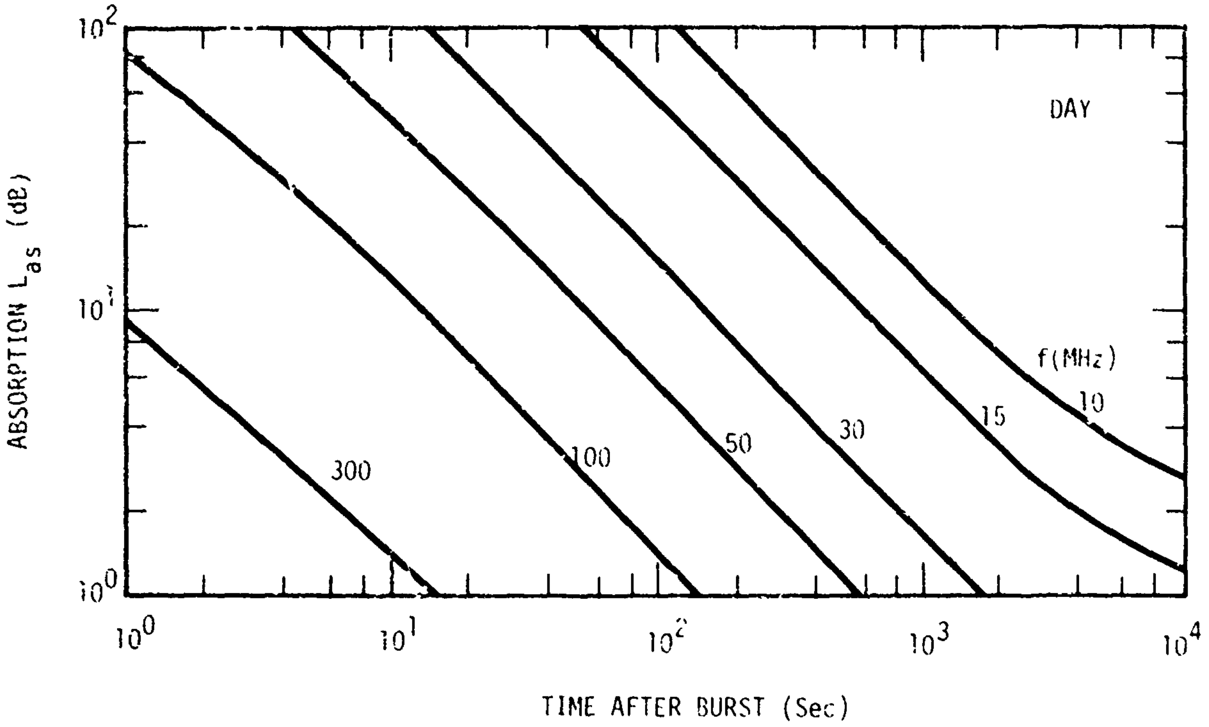


Figure 12. One-way vertical absorption due to saturation impulse of ionization.

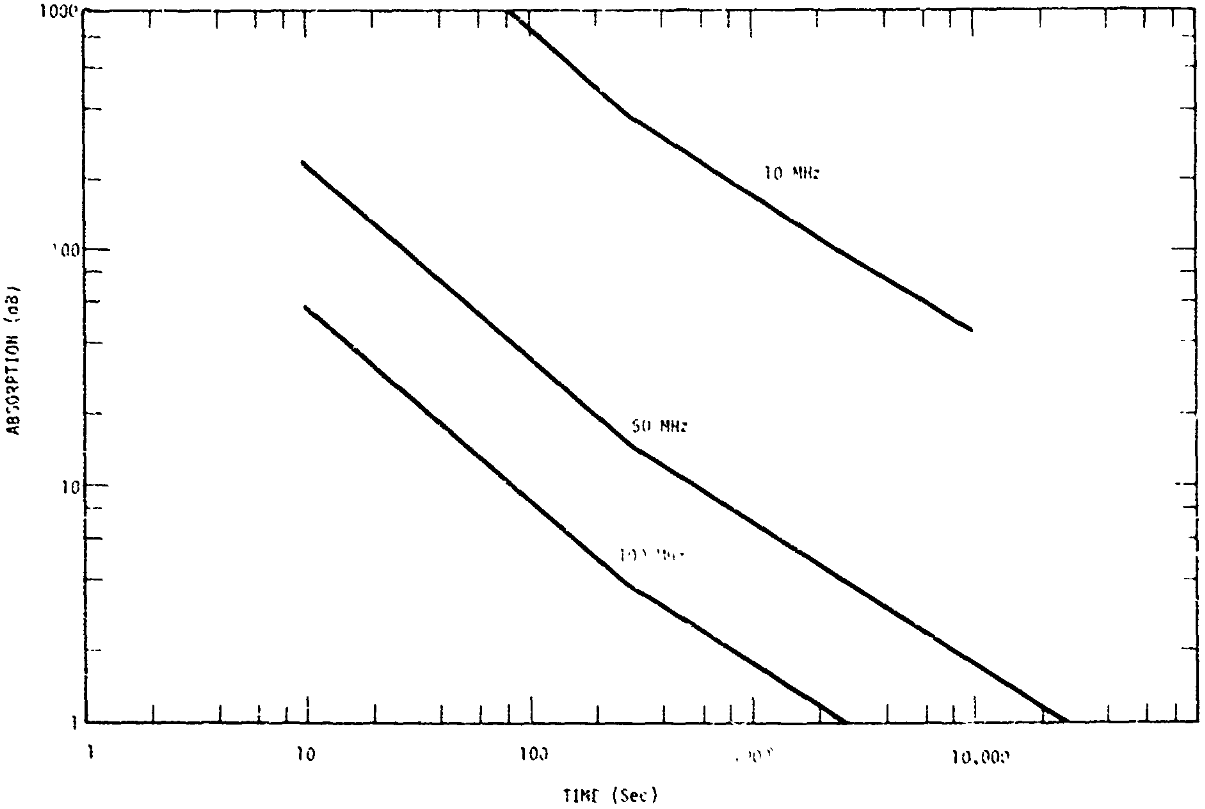


Figure 13. Gamma-ray absorption versus time.

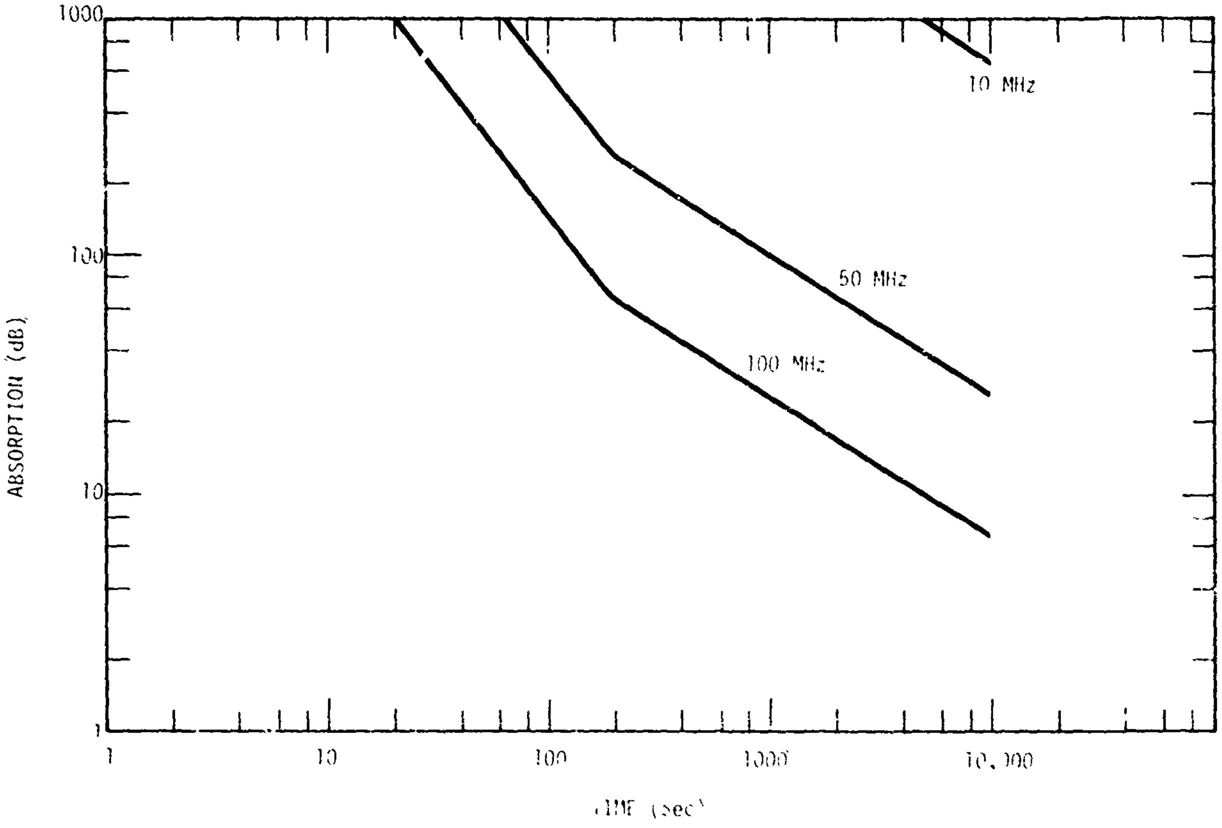


Figure 14. Beta-ray induced absorption versus time.

DISCUSSION

J. Rottger, Ge

Why does "isotropic s ccur at frequencies lower than the critical plasma frequency? Isn't "(total) reflection" from the sur e cloud?

Author's Reply

No, total reflection is too simple a concept here. Rather the incoming wave generates surface currents which produce isotropic scattering as from a metallic sphere.

T B. Jones, UK

Would you please comment on the possibility of the production of electrostatic images of the cloud in the E layer The E layer conjugations produced in this way can support additional modes of propagation which result in modal interference etc.

Author's Reply

Images can be generated, both in the local E region but also in the magnetic conjugation region over the equator. These images may support other propagation modes. This problem has not been addressed to date, but will be considered in future experimental planning

James R. Wait, US

Could you indicate what theoretical model was used to calculate the bi-static cross section of the cloud? In this connection, did you account for the radial density of the cloud density?

Author's Reply

Short pulse radar returns from the cloud showed no pulse spreading indicating the electron density gradient was short compared to a wavelength As a result computations of the cross section using geometrical optics approximations were employed.

Comment by N.C. Gerson US

If the ionization produced by the chemical releases becomes field aligned, several factors restrict the results presented. First, field aligned propagation implies an aspect sensitive condition, in which the transmitter and receiver must be in certain areas, respectively, for propagation to take place. Secondly, with field aligned propagation, both the transmitter and receiver must be located below a given geomagnetic latitude for the reflected ray to be returned to the earth

Author's Comment

I agree with this comment The results presented only apply prior to the ion cloud becoming field-aligned, approximately 5 minutes after release.

George H. Millman, US

You mentioned that transmission distances on the order of 3000 miles were attained during one of the barium releases. Could you amplify this statement in terms of transmission frequency, and reflection height?

Author's Reply

The transmission frequency was approximately 10Mhz. The height of reflection was about 215km, hence some refraction occurred to achieve the range of 3000 miles.

TROPOSPHERIC EFFECTS ON HF PROPAGATION

George H. Millman
General Electric Company
Syracuse, New York 13221, USA

SUMMARY

An evaluation is made of the effect of tropospheric refractive bending on the propagation of HF radio waves. The index of refraction in the troposphere is modeled in terms of the CRPL Reference Refractivity Atmosphere - 1958, while the index of refraction in the ionosphere is defined in terms of the transverse-ordinary mode of propagation, the electron density being represented by a Chapman model.

Ray tracings are performed based on the assumption that the propagation media are stratified into layers of constant refractive index. The ground scatter distance and the true and virtual reflection heights are calculated as a function of surface refractivity and various electron density profiles.

The relationship between radar range and target-ground distance is also examined. The existence of long range propagation paths beyond 4500 km for specific tropospheric and ionospheric conditions is discussed.

1. INTRODUCTION

When electromagnetic waves are propagated through the troposphere and ionosphere, they undergo a change in direction or refractive bending. This phenomenon which arises from the nonhomogeneous characteristics of the media introduces an angular error in radar measurement data (Millman, 1958).

The elevation angle deflection of radar waves in the troposphere is independent of frequency, and in the ionosphere it is frequency dependent, i.e., inversely proportional to frequency squared (Millman, 1958). However, in the case of the troposphere, the angular deviation is directly proportional to the surface refractivity, N_o, i.e., the deviation increasing with increasing N_o (Millman, 1971).

In predicting the range-coverage performance of an HF backscatter radar or the transmission-frequency requirements of a communications system, ionospheric propagation characteristics are only considered. The effects of the tropospheric refractive properties on HF propagation are in general not taken into account.

However, a preliminary examination of the HF propagation phenomena has revealed that the ground-scatter distance and the true and virtual reflection heights are modified when tropospheric refraction effects are taken into account (Millman, 1975).

In this paper, an evaluation is made of the effect of refractive bending in the troposphere on the propagation of HF radio waves.

2. THEORETICAL CONSIDERATIONS

2.1 Tropospheric Index of Refraction

The index of refraction, n_t, in the troposphere, can be expressed in terms of the functions

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

and

$$N = \frac{a}{T}\left(p + \frac{be}{T}\right) \tag{2}$$

where N is the refractivity, T is the air temperature (°K), p is the total pressure (mbar) and e is the partial pressure of water vapor (mbar). According to Smith and Weintraub (1953), the constants, a and b, are 77.6°K/mbar and 4810°K, respectively.

It should be noted that the above expression for the refractivity of air is independent of frequency in the 100- to 30,000-MHz range. The first term in Equation (2), ap/T, applies to both optical and radio frequencies, and is often referred to as the dry term. The second term, abe/T^2, which is the wet term, is the water vapor relationship required only at radio frequencies.

The tropospheric refractive index model employed in this analysis is the CRPL Reference Refractivity Atmosphere - 1958 (Bean and Dutton, 1966) which is described by

$$N(h) = N_o + (h - h_o)\,\Delta N \tag{3}$$

where N_o is the surface refractivity and h_o is the surface height above mean sea level. This expression is valid for $h_o \le h \le (h_o + 1)$ km. The parameter, ΔN, is defined by

$$\Delta N = -7.32 \exp (0.005577\, N_o) \tag{4}$$

For the region defined by $(h_o + 1) \le h \le 9$ km, the refractivity decays as

$$N(h) = N_1 \exp [-c(h - h_o - 1)] \tag{5}$$

where N_1 is the value of N(h) at 1 km above the surface and

$$c = \frac{1}{8 - h_o} \log_e \left(\frac{N_1}{105} \right) \tag{6}$$

Above 9 km, the exponential decay is of the form

$$N(h) = 105 \exp [-0.1424\,(h - 9)] \tag{7}$$

Surface refractivities of 320-N and 400-N units were only considered. The latter is representative of severe refraction conditions while the former of average conditions.

2.2 Ionospheric Index of Refraction

The index of refraction in the ionosphere can be expressed by the relationship

$$n_i = \left[1 - \left(\frac{\omega_N}{\omega} \right)^2 \right]^{1/2} \quad \left[1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} \right]^{1/2} \tag{8}$$

where ω_N is the angular plasma frequency of the medium (rad/s), N_e is the electron density (electrons/m^3), e is the electron charge (1.6 x 10^{-19} C), m_e is the electron mass (9.1 x 10^{-31} kg), ϵ_o is the electric permittivity of free space ($10^{-9}/36\pi$ F/m) and ω is the angular frequency of the incident wave (rad/s).

The distribution of electron density with height is assumed to follow the Chapman model of the form

$$N_e = N_m \exp \left\{ \frac{1}{2} \left[1 - \frac{h - h_m}{H_s} - \exp \left(\frac{-(h - h_m)}{H_s} \right) \right] \right\} \tag{9}$$

where H_s is the scale height of the neutral particles and N_m is the electron density at the level of maximum ionization, h_m.

The values of the parameters defining the daytime ionospheric electron density profiles and the equivalent plasma frequencies of the maximum ionization levels, used in the calculations, are presented in Table 1. Minimum electron density between the E- and F1-layer is attained at 128.01 km altitude and between the F1- and F2-layer at 213.55-km altitude.

It should be noted that the ionospheric refractive index, given by Equation (8), is also a function of both the electron collision frequency and the earth's magnetic field. For frequencies on the order of 10 MHz and above, and at altitudes greater than 80 km, the effect of the collision frequency term on the index of refraction is negligible (Davies, 1965).

The refractive index is slightly in error when the magnetic field is neglected. It can be shown that, when the magnetic field is taken in account, the refractive index is defined by the following (Ratcliffe, 195

1. Longitudinal-ordinary propagation mode

$$n_{Lo} = \left\{ 1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} \left[1 + \frac{eB}{m_e \omega} \cos \theta \right]^{-1} \right\}^{1/2} \tag{10}$$

2. Longitudinal-extraordinary propagation mode

$$n_{Lx} = \left\{ 1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} \left[1 - \frac{eB}{m_e \omega} \cos \theta \right]^{-1} \right\}^{1/2} \tag{11}$$

3. Transverse-ordinary propagation mode

$$n_{To} = \left[1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} \right]^{1/2} \tag{12}$$

4. Transverse-extraordinary propagation mode

$$n_{Tx} = \left\{ 1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} \left[1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} \right] \left[1 - \frac{N_e e^2}{\epsilon_o m_e \omega^2} - \left(\frac{eB}{m_e \omega} \sin \theta \right)^2 \right]^{-1} \right\}^{1/2} \tag{13}$$

where B is the magnetic induction (Wb/m) and θ is the pr e, i.e., the angle between the magnetic field vector and the direction of propagation.

When evaluating the errors with respect to the transverse-ordinary mode of propagation which corresponds to Equation (8), the nonmagnetic field case, it is found that the maximum error in the refractive index occurs the peak of the F2-layer and, at 30 MHz, is less than 0.4 percent. At 20 MHz, the error increases to slightly greater than 1 percent. The calculations are based on the electron density models described in Table 1 and on a magnetic field intensity of 0.5 G ($B = 0.5 \times 10^{-4}$ Wb/m^2) which is assumed to be invariant with altitude (Millman, 1977).

It should be noted that transverse propagation is applicable, to a first approximation, to transmissions originating in the midlatitudes and oriented towards the polar ionosphere. Longitudinal propagation, on the other hand, applies to midlatitudes transmissions directed equatorward.

2.3 Computational Procedure

In this analysis, it is assumed that (1) the troposphere is contained between the earth's surface and 30-km altitude, (2) the base of the ionosphere is located at an altitude of 80 km and (3) free space prevails in the region between the troposphere and the ionosphere.

In evaluating the effects of the tropospheric refraction phenomenon on HF propagation, the basic assumption employed is that both the troposphere and ionosphere can be considered to be stratified into spherical layers of constant refractive index. The stratified layer method, although approximate in nature, can be greatly improved by merely increasing the number of layers in the medium, i.e., decreasing the thickness of each individual layer element (Millman, 1961).

The slab sizes employed in the computations are as follows: 50 m in the altitude region from h = 0 to 30 km; 50 km from h = 30 to 80 km; and 1 km from h = 80 to 300 km.

3. DISCUSSION

Estimates of the effect of tropospheric refraction on HF propagation are presented in Tables 2 through 7 which list the ground-scatter distances and the true and virtual reflection heights for transmissions at 10, 20 and 30 MHz and at 1° and 3.5° elevation angles. The calculations are based on the nonmagnetic field, ionospheric index of refraction, i.e., transverse-ordinary mode of propagation, defined by Equation (8) and apply to the four electron density models described in Table 1.

The analytical expressions for the true reflection height, h_r, and the virtual reflection height, h'_r, are given by (Millman, 1977)

$$h_r = r_o \left\{ n_c \cos E_o \left[1 - \frac{N_e e^2}{4\pi^2 \epsilon_o m_e f^2} \right]^{-1/2} - 1 \right\} \tag{14}$$

and

$$h'_r = r_o \left\{ \cos E_o \left(\cos \left[E_o + \frac{S}{2r_o} \right] \right)^{-1} - 1 \right\} \tag{15}$$

where r_o is the earth's radius, n_o is the index of refraction at the earth's surface, E_c is the apparent ground elevation angle and S is the ground-scatter distance.

An examination of the data reveals that, for a given set of propagation conditions, i.e., tropospheric and ionospheric models, the true reflection height increases with increasing transmission frequency and elevation angle. It is seen that, for a given electron density model, the true reflection height decreases with increasing surface refractivity.

The ground-scatter distance and the virtual reflection height, on the other hand, increase with increasing surface refractivity except in the case of the 20-MHz, 1° elevation angle data (Table 4) where the reverse occurs for electron density model C. This reversal can be attributed to the fact that the true height of reflection of $N_c = 0$ takes place above the peak of the F1-layer, i.e., at an altitude of 214.17 km while the $N_o = 320$ and $N_o = 400$ reflections occur below the F1-layer peak, i.e., at 181.30- and 179.56-km altitude, respectively.

It is of interest to note that the decrease in the ground-scatter distance at a surface refractivity of 400-N units, electron density model A (Table 4), is mainly the result of the true reflection taking place at approximately the peak of the E-layer, i.e., 99.96-km altitude. This is in contrast with the $N_o = 0$ and $N_c = 320$ data which reflect at the lower portion of the F1-layer, i.e., 174.41 and 172.34 km altitude, respectively.

An interesting disclosure in the 20-MHz results, Tables 4 and 5, is the existence of long-range propagation paths for specific tropospheric and ionospheric conditions. It is noted that long-range propagation, i.e., ground-scatter distances greater than approximately 4500 km, generally tends to occur for rays undergoing reflection at altitudes on the order of 214 km which is slightly above the altitude where the F1- and F2-layer are joined together, the ionization valley being located at 213.55-km altitude.

Figures 1 through 3 disclose 20-MHz high-ray one-hop F1 mode propagation (Pederson ray) at distances up to 6200 km, with maximum range being attained at elevation angles between approximately 1° and 3.5°. As illustrated in Figure 4, a slight variation in the electron density profile can result in: (1) the maximum range being acquired at the lowest elevation angle and, (2) the disappearance of the F1 propagation mode.

Long-distance, one-hop F1 propagation over approximately a 4500-km path has been observed by Tveten (1961). Propagation by the one-hop F2 mode over a 5300-km path is possible according to the experimental measurements of Warren and Hagg (1958) and the theoretical calculations of Kift (1958).

Utilizing ray tracings techniques, Muldrew and Maliphant (1962) investigated the properties of long-distance, one-hop propagation. They found that long-distance propagation may occur via the F1- and as well as the F2-layer and that one-hop propagation may extend to ranges in excess of 7500 km in the temperate regions and 10,000 km in the equatorial region.

It should be apparent from Figures 1 through 4 which are visual representations of the data in Tables 4 and 5 that, at low elevation angles, the ground-scatter distance is controlled to some extent by the troposphere. As previously mentioned, when tropospheric refraction is severe (i.e., high surface refractivity) the ground scatter distance is generally a maximum. It is a minimum when the surface refractivity is not taken into account (i.e., $N_o = 0$). For example, the ground-scatter reflection point, at 2° elevation angle for ionospheric model D, Figure 4, evaluates to 5052.50 km for $N_o = 400$ and decreases to 4966.34 km for $N_o = 320$ and 4738.26 km for $N_o = 0$.

An additional feature of interest in Figures 1 through 3 is the indication of the possible existence of multiple rays incident at the same location on the earth's surface commencing at distances greater than about 3800 km. This phenomenon implies that there is a focusing of the rays which results in an apparent enhancement of the incident radiation at the long ranges.

An evaluation of HF radio focusing at maximum range caused by the ionization distribution between ionospheric layers has been made by Croft (1967).

Figures 5 and 6 are plots of the radar range, i.e., group path length, as a function of the ground distance at a frequency of 20 MHz for ionospheric models A and D, respectively. It is seen that, for a given set of propagation conditions, i.e., ionospheric model and high- or low-ray propagation mode, a linear relationship exists between the radar range and the ground distance which is independent of the tropospheric refraction conditions. It follows, therefore, that there is no need to take into account the effect of refractive bending in the troposphere when determining the location of an object in space by HF backscatter radar techniques. In other words, no special calculations have to be made to account for the effects of the troposphere on radar-target, coordinate-registration process.

On comparing Figure 5 with Figure 6, it is found that the high-ray (F2 mode) slopes are practically the same. An identical situation exists for the slopes of the low rays (F2 mode).

The 30-MHz radar-range data shown in Figures 7 and 8 also display similar slopes for the low rays and for the high rays. It is noted that the slope of the data points of the high rays is slightly greater than that of the low rays for both the 20- and 30-MHz computations.

The linear relationships between the radar ranges and the ground distances illustrated in Figures 5 through 8 are summarized in Table 8.

Additional items which should be mentioned with regard to the 30-MHz data are: (1) long-range propagation beyond 4500-km ground distance was not attainable for the four ionospheric models considered in this analysis, (2) the elevation angle versus ground-scatter distance plots were similar in appearance to the curves illustrated in Figure 4, and (3) the ground distance traversed by the high rays tended to approach that of the low rays.

4. CONCLUSIONS

The presence of tropospheric bending tends to have the following effects on HF propagation: to decrease the true height of reflection and, for the most part, to increase the ground distance over which the waves can be transmitted and to increase the virtual height of reflection.

Slight variations in the ionospheric electron density distribution could result in the ground scatter distance and virtual reflection height decreasing with increasing severity of tropospheric refraction.

Inasmuch as a linear relationship exists between radar range and ground distance, tropospheric refraction effects need not be taken into account when deducing the location of a target by means of HF backscatter radar techniques.

Long distance propagation beyond 4500 km appears to occur for rays: (1) having very low takeoff angles, i.e., less than about 4° and, (2) undergoing reflection in the altitude region where the F1 and F2 layers are joined together.

It is found that, for the electron density models considered in this analysis, the long-distance propagation mode is frequency sensitive in that the extended surface coverage occurred only at 20 MHz and not at 10 and 30 MHz.

5. REFERENCES

Bean, B.R. and E.J. Dutton, "Radio Meteorology," National Bureau of Standards Monograph 92, U.S. Government Printing Office, Washington, D.C., 1966.

Croft, T.A., "HF Radio Focusing Caused by the Electron Distribution Between Ionospheric Layers," Journal of Geophysical Research, Vol. 72, pp 2343-2355, May 1, 1967.

Davies, K., "Ionospheric Radio Propagation," National Bureau of Standards Monograph 80, U.S. Government Printing Office, Washington, D.C., 1965.

Kift, F., "Single-Hop Propagation of Radio Waves to a Distance of 5300 Km," Nature, Vol. 181, pp 1459-1460, May 24, 1958.

Millman, G.H., "Atmospheric Effects on VHF and UHF Propagation," Proceedings of IRE, Vol. 46, pp 1492-1501, August 1958.

Millman, G.H., "Atmospheric and Extraterrestrial Effects on Radio Wave Propagation," General Electric Technical Information Series Report No. R61EMH29, June 1961.

Millman, G.H., "Tropospheric Effects on Space Communications," AGARD Conference Proceedings No. 70 (Tropospheric Radio Wave Propagation), Part 1, pp (4-1) - (4-29), 1971.

Millman, G.H., "Refraction Effects on Magnetic Field Geometry and HF Propagation," Journal of Atmospheric and Terrestrial Physics, Vol. 37, pp 751-760, May 1975.

Millman, G.H., "The Effect of Tropospheric Refraction on HF Propagation," General Electric Technical Information Series Report No. R77EMH5, March 1977.

Muldrew, D.B., and R.G. Maliphant, "Long-Distance One-Hop Ionospheric Radio-Wave Propagation," Journal of Geophysical Research, Vol. 67, pp 1805-1815, May 1962.

Ratcliffe, J.A., "The Magneto-Ionic Theory and Its Application to the Ionosphere," Cambridge University Press, New York, 1959.

Smith, E.K., and S. Weintraub, "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies," Proceedings of IRE, Vol. 41, pp 1035-1037, August 1953.

Tveten, L.H., "Long-Distance One-Hop F_1 Propagation Through the Auroral Zone," Journal of Geophysical Research, Vol. 66, pp 1683-1684, June 1961.

Warren, E. and E.L. Hagg, "Single-Hop Propagation of Radio Waves to a Distance of 5300 Km," Nature, Vol. 181, pp 34-35, January 4, 1958.

TABLE 1

IONOSPHERIC ELECTRON DENSITY PROFILES

Model	Layer	Scale Height H_s (km)	Altitude (km)	Maximum Electron Density N_m (x 10^{11} electrons/m^3)	Plasma Frequency (MHz)
A	E	10	100	1.500	3.477
	F1	40	200	3.000	4.917
	F2	50	300	12.500	10.037
B	E	10	100	1.459	3.429
	F1	40	200	2.918	4.850
	F2	50	300	12.160	9.899
C	E	10	100	1.440	3.406
	F1	40	200	2.879	4.817
	F2	50	300	11.997	9.833
D	E	10	100	1.421	3.384
	F1	40	200	2.841	4.785
	F2	50	300	11.838	9.767

TABLE 2

GROUND-SCATTER DISTANCE AND TRUE AND VIRTUAL REFLECTION HEIGHTS FOR 10-MHz TRANSMISSIONS AT 1° ELEVATION ANGLE

Electron Density Model	Refractivity (N Units)	Ground-Scatter Distance (km)	True Reflection Height (km)	Virtual Reflection Height (km)
A	0	1863.96	81.66	85.38
	320	1949.20	81.58	92.67
	400	1984.80	81.56	95.80
B	0	1863.86	81.77	85.37
	320	1948.78	81.69	92.63
	400	1984.26	81.66	95.76
C	0	1863.90	81.82	85.38
	320	1948.70	81.74	92.63
	400	1984.16	81.71	95.75
D	0	1864.02	81.87	85.39
	320	1952.88	81.79	92.99
	400	1984.12	81.76	95.74

TABLE 3

GROUND-SCATTER DISTANCE AND TRUE AND VIRTUAL REFLECTION HEIGHTS FOR 10-MHz TRANSMISSIONS AT 3.5° ELEVATION ANGLE

Electron Density Model	Surface Refractivity (N Units)	Ground-Scatter Distance (km)	True Reflection Height (km)	Virtual Relection Height (km)
A	0	1478.28	82.15	89.16
	320	1534.18	82.07	94.30
	400	1559.27	82.05	96.65
B	0	1467.92	82.27	88.23
	320	1507.60	82.19	91.34
	400	1520.72	82.16	93.05
C	0	1485.18	82.33	87.98
	320	1502.46	82.24	91.37
	400	1514.52	82.22	92.48
D	0	1463.28	82.39	87.81
	320	1499.74	82.30	91.12
	400	1510.60	82.24	92.12

TABLE 4

GROUND-SCATTER DISTANCE AND TRUE AND VIRTUAL REFLECTION HEIGHTS FOR 20-MHz TRANSMISSIONS AT 1° ELEVATION ANGLES

Electron Density Model	Surface Refractivity (N Units)	Ground-Scatter Distance (km)	True Reflection Height (km)	Virtual Reflection Height (km)
A	0	4117.06	174.41	387.16
	20	4282.24	172.34	418.92
	400	2536.74	99.96	151.34
B	0	4176.54	179.66	398.43
	320	4251.74	177.47	420.79
	400	4297.90	176.92	422.00
C	0	5767.26	214.17	773.97
	320	4609.34	181.30	486.1[illegible]
	400	4360.96	179.56	434.56
D	0	5220.18	214.75	627.80
	320	5732.76	214.35	764.18
	400	5934.86	214.25	822.66

TABLE 5

GROUND-SCATTER DISTANCE AND TRUE AND VIRTUAL REFLECTION HEIGHTS FOR 20-MHz TRANSMISSIONS AT 3.5° ELEVATION ANGLE

Electron Density Model	Surface Refractivity (N Units)	Ground-Scatter Distance (km)	True Reflection Height (km)	Virtual Reflection Height (km)
A	0	4590.08	214.49	597.49
	320	4992.12	214.12	700.99
	400	5179.02	214.02	752.59
B	0	4203.46	215.71	507.14
	320	4393.24	215.33	530.40
	400	4453.82	215.23	564.65
C	0	4124.14	216.33	489.67
	320	4224.52	215.94	511.84
	400	4267.90	215.84	521.60
D	0	4007.94	216.95	464.72
	320	4145.30	216.55	494.30
	400	4185.52	216.45	503.16

TABLE 6

GROUND-SCATTER DISTANCE AND TRUE AND VIRTUAL REFLECTION HEIGHTS FOR 30-MHz TRANSMISSIONS AT 1° ELEVATION ANGLE

Electron Density Model	Surface Refractivity (N Units)	Ground-Scatter Distance (km)	True Reflection Height (km)	Virtual Reflection Height (km)
A	0	3944.84	246.67	355.56
	320	4093.26	246.16	382.70
	400	4135.32	246.03	400.11
B	0	3973.60	248.48	360.73
	320	4045.00	247.95	373.75
	400	4089.60	247.62	382.01
C	0	3989.24	249.39	363.56
	320	4057.56	248.86	376.07
	400	4102.84	248.72	384.49
D	0	4011.54	250.38	367.61
	320	4074.66	249.77	379.23
	400	4116.58	249.64	387.07

TABLE 7

GROUND-SCATTER DISTANCE AND TRUE AND VIRTUAL REFLECTION HEIGHTS FOR 30-MHz TRANSMISSIONS AT 3.5° ELEVATION ANGLE

Electron Density Model	Surface Refractivity (N Units)	Ground-Scatter Distance (km)	True Reflection Height (km)	Virtual Reflection Height (km)
A	0	3472.76	248.44	359.21
	320	3487.48	248.94	361.91
	400	3505.98	248.81	365.32
B	0	3501.88	251.62	364.57
	320	3519.48	250.99	367.82
	400	3538.14	250.83	371.29
C	0	3510.42	252.78	366.14
	320	3550.22	252.14	373.55
	400	3611.44	251.98	385.10
D	0	3518.88	253.95	367.71
	320	3603.22	253.30	383.53
	400	3642.38	253.13	391.01

TABLE 8

LINEAR RELATIONSHIP BETWEEN RADAR RANGE AND GROUND DISTANCE

Frequency (MHz)	Electron Density Model	Low Ray (F1 Mode)		High Ray (F1 Mode) and Low Ray (F2 Mode)		Low Ray (F2 Mode)		High Ray (F2 Mode)	
		a	b (km)	a	b (km)	a	b (km)	a	b (km)
20	A	1.06	-96.47	1.02	110.83	-	-	1.21	-103.03
	D	-	-	-	-	1.01	127.80	1.28	-240.37
30	A	-	-	-	-	0.99	105.22	1.15	277.39
	D	-	-	-	-	0.97	245.83	1.13	-238.71

$R = aS + b$

R = Radar Range (km)

S = Ground Distance (km)

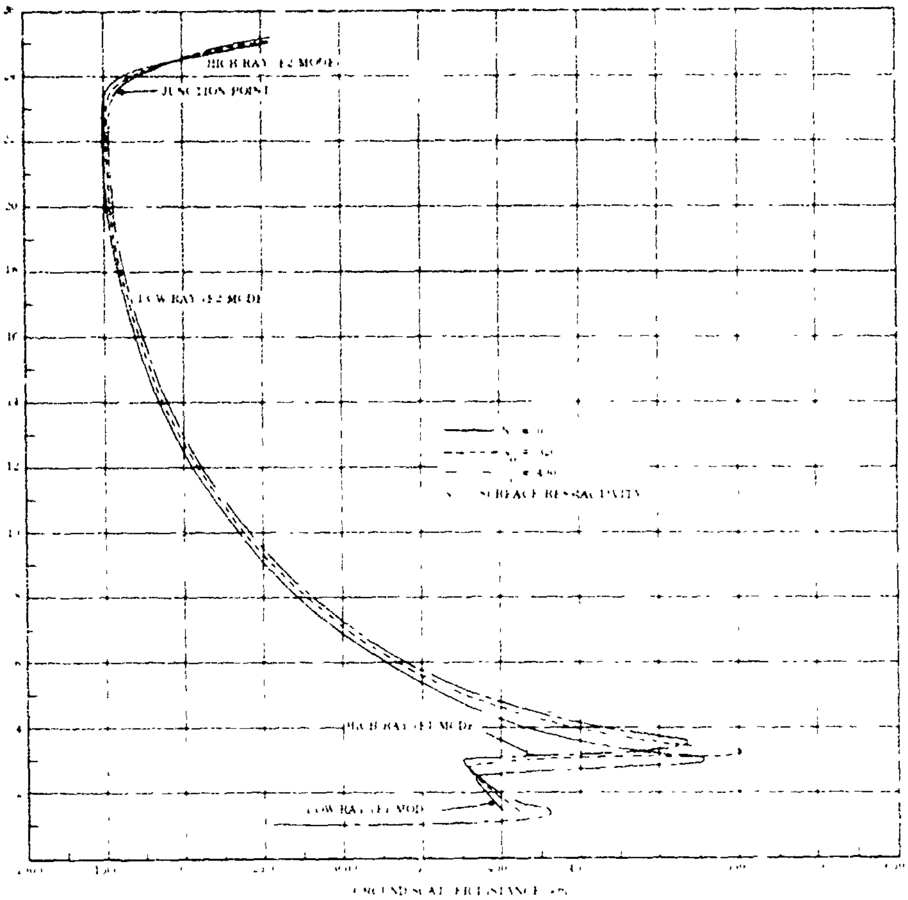


Figure 1. Ground-Scatter Distance As A Function of Elevation Angle For Transmissions at 20 MHz, Ionospheric Electron Density Model A

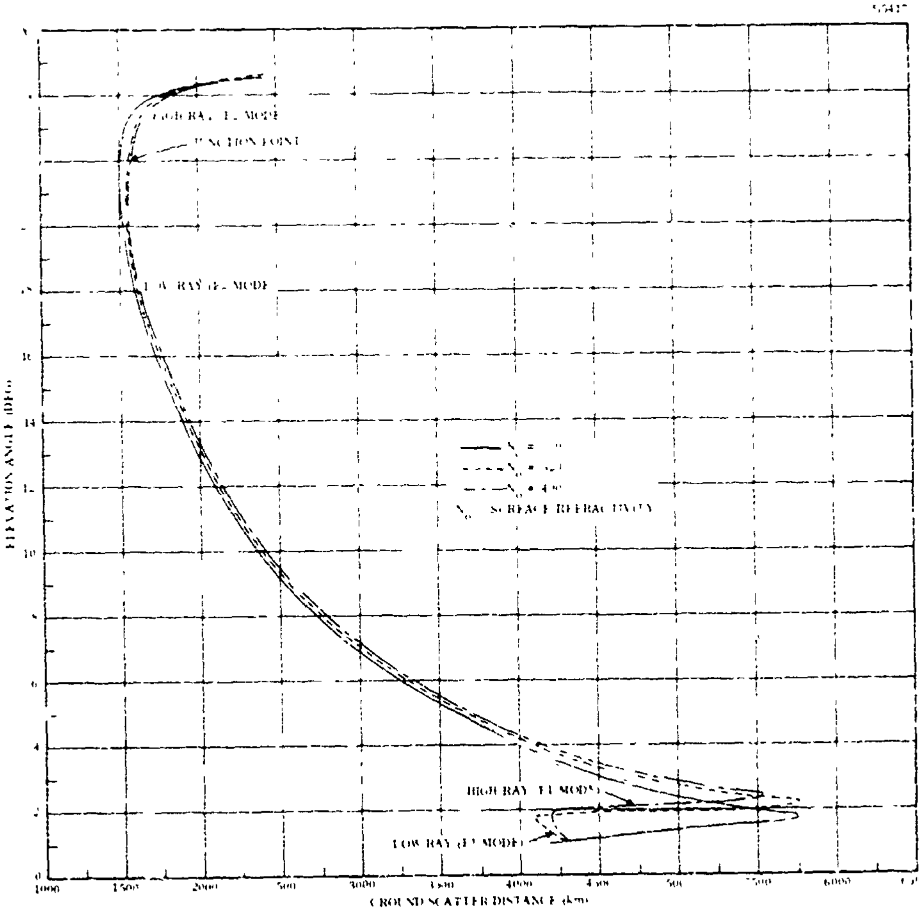


Figure 2. Ground-Scatter Distance As A Function of Elevation Angle For Transmissions at 20 MHz, Ionospheric Electron Density Model B

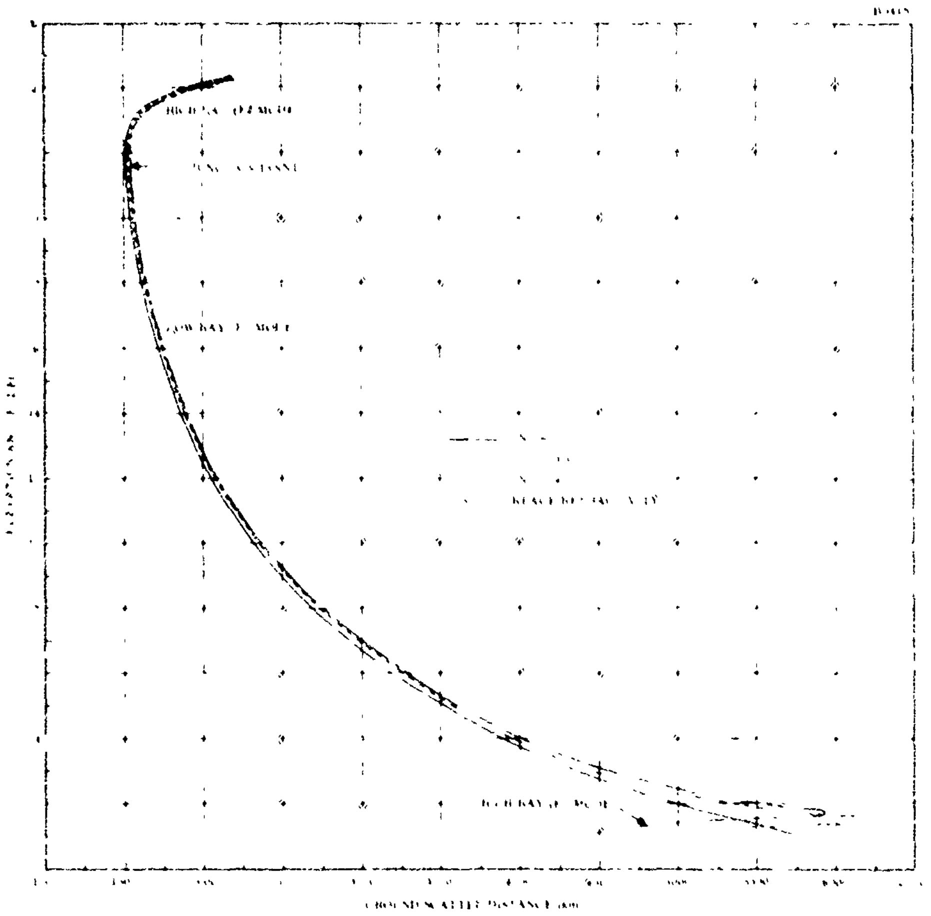


Figure 3. Ground-Scatter Distance As A Function of Elevation Angle For Transmissions at 20 MHz, Ionospheric Electron Density Model C

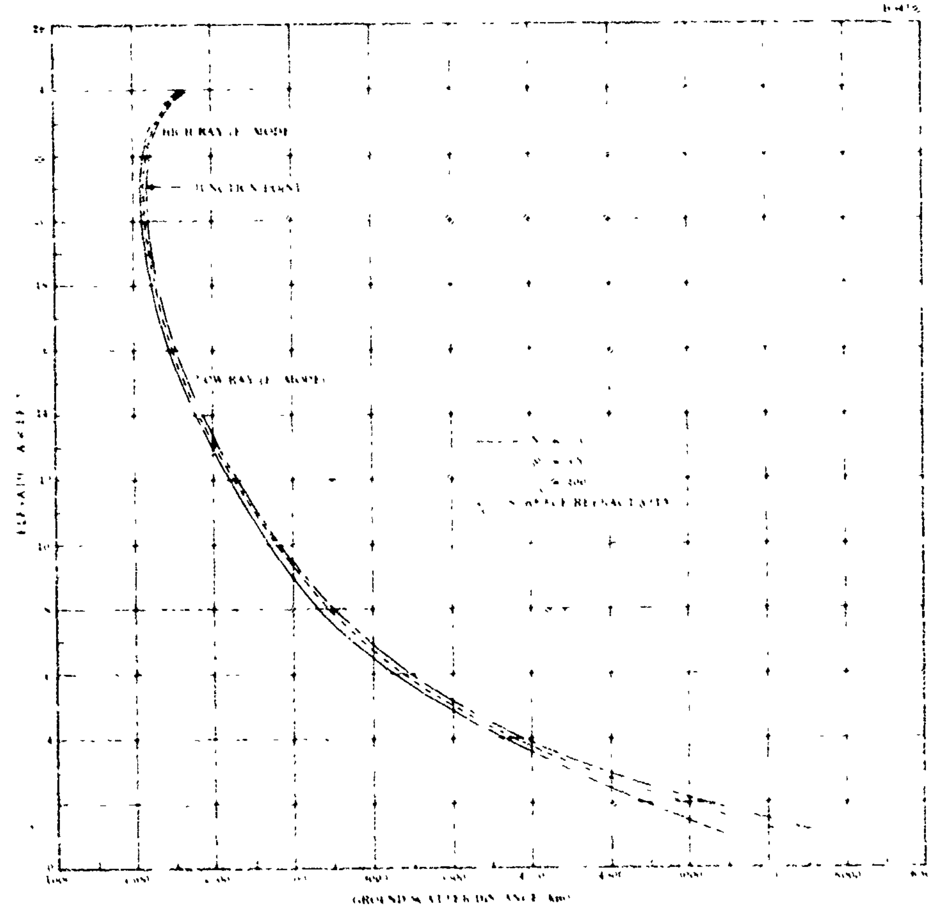


Figure 4. Ground-Scatter Distance As A Function of Elevation Angle For Transmissions at 20 MHz, Ionospheric Electron Density Model D

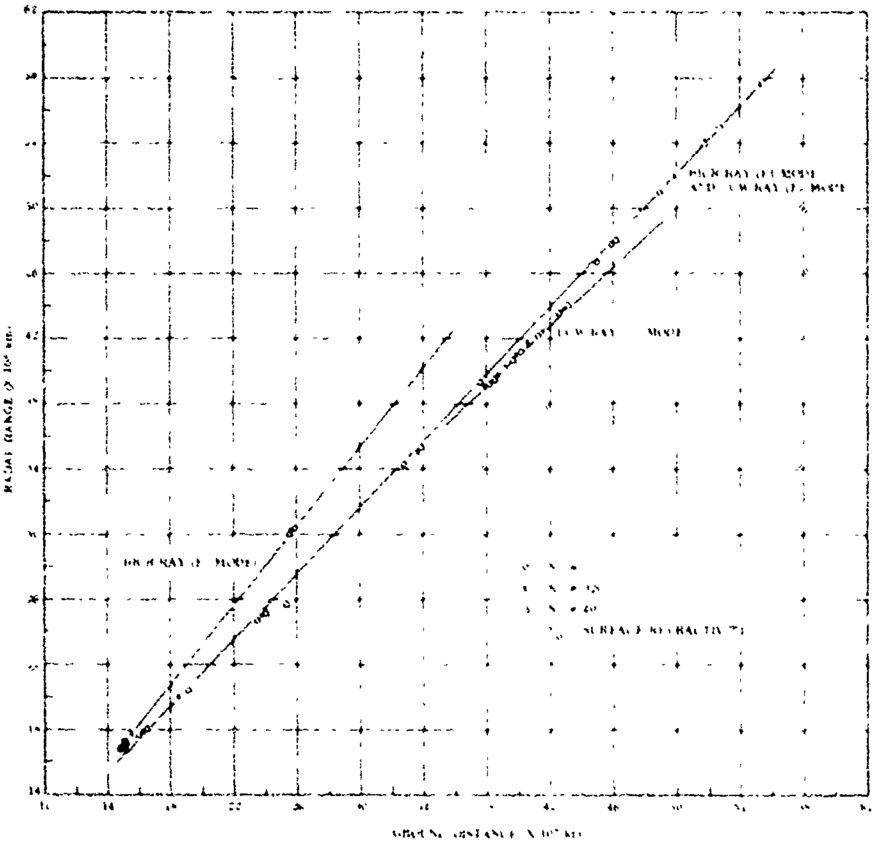


Figure 5. Radar Range As A Function of Ground Distance At 20 MHz For Ionospheric Electron Density Model A

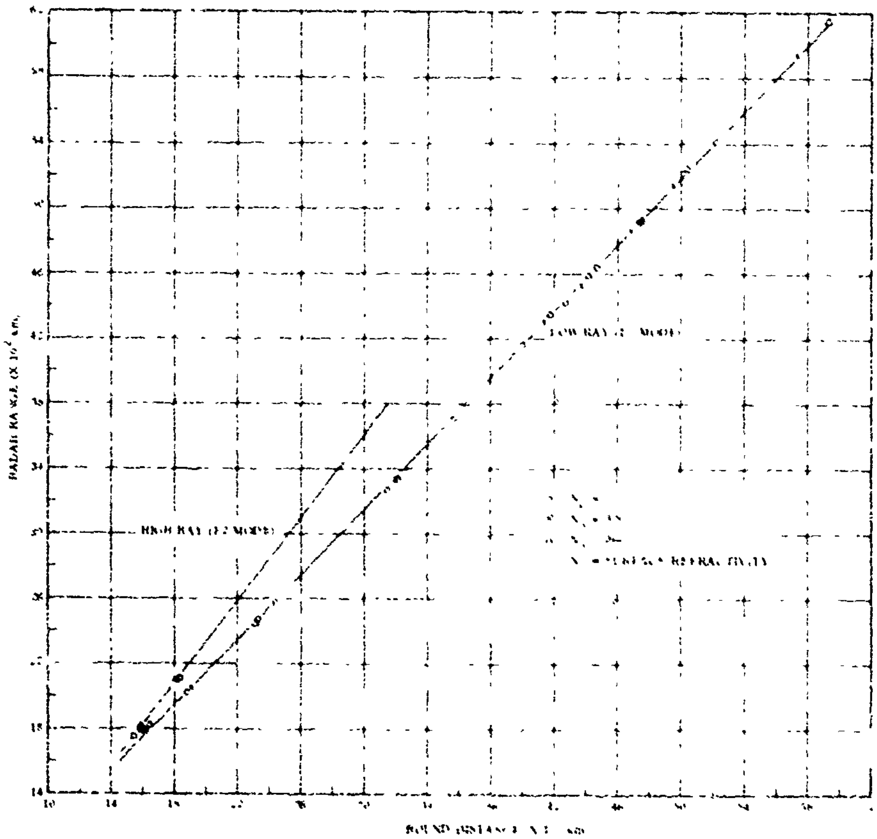


Figure 6. Radar Range As A Function of Ground Distance At 20 MHz For Ionospheric Electron Density Model D

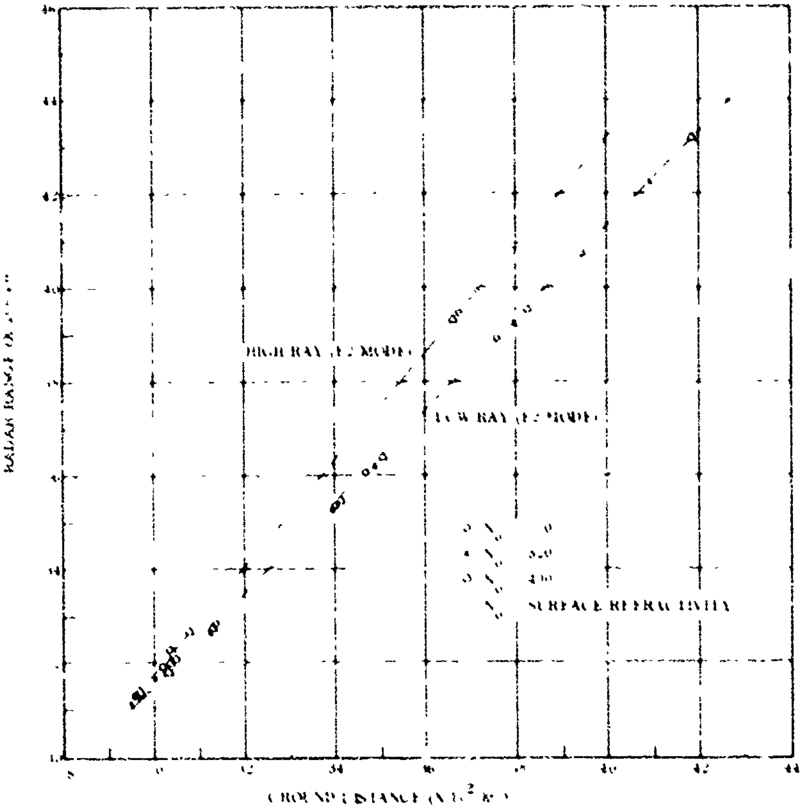


Figure 7. Radar Range As A Function of Ground Distance At 30 MHz For Ionospheric Electron Density Model A

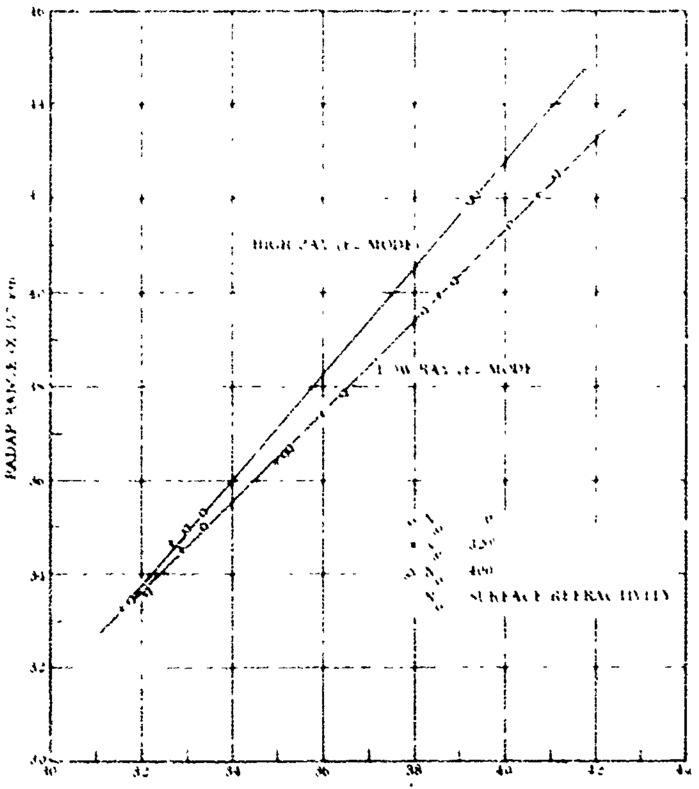


Figure 8. Radar Range As A Function of Ground Distance At 30 MHz For Ionospheric Electron Density Model D

DISCUSSION

Comment by N.C. Gerson US

It should be noted that modified refractive indices of 400 occur primarily in tropical oceanic areas, over India during the monsoon period and several other similar geographic areas Range changes of the type calculated could be expected to occur primarily in these areas At middle latitudes, modified refractive indices of 400 or more would occur primarily during summer and over large bodies of water At high latitudes such refractivity are rarely attained if at all

Author's Comment

It is true that surface refractivities of 400 N units most likely can not be attained in the high latitudes This value merely results in the maximum effects that could be encountered in HF propagation Surface refractivity of 320-N units which I have also considered in this study is applicable for average conditions

Comment by J.R Wait, US

I feel that tropospheric ducting could play a role for grazing type rays at HF Such duct occurence could modify some of these conclusions

Author's Reply

As I mentioned, in this study I only examined the effects of the standard atmospheric model It is quite possible that tropospheric ducting could alter the results presented

SUMMARY OF SESSION II

MODELING OF IONOSPHERIC PARAMETERS

The session entitled, "Modeling of Ionospheric Parameters" dealt with recent developments in the area of ionospheric modeling for the purposes of improving the specification of radio propagation conditions. The papers presented in the session and the ensuing discussions were concerned with improvements in our understanding of the dynamics of the ionosphere, improvements in our ability to observe significant ionospheric changes, and improved methods for translating ionospheric information into radio progagation data (and vice versa).

The paper by Wright and Paul discussed various aspects of a proposed new ionospheric sounding network. This new network is composed of modern, digital ionosondes that are fully computer-interactive. These ionosondes could provide vertical incidence and oblique incidence ionograms, signal amplitude, direction of arrival information and polarization. The authors point out the advantages of such a network of sophisticated ionosondes and call for deliberate planning in the upgrade of the existing ionosonde network.

The paper by **Cho, Migliora** and **Felsen** addressed a new method of analyzing radio propagation using a combination of ray and mode theory. Using as an illustration the propagation of radio waves in an elevated tropospheric duct, the presentation given by **Felsen** stressed how the hybrid ray-mode approach could be readily applied to a number of different propagation related problems across the frequency spectrum.

The paper by **Fleury** and **Gourvez** discussed the importance of various ionospheric parameters in the determination of HF long-distance predictions. The authors pointed out that the critical frequency of the F2 region, foF2 is the most important parameter in the determination of HF propagation conditions and they discussed how uncertainties in foF2 are manifested in uncertainties in HF propagation predictions.

The paper by **DuBroff, Rao** and **Yeh** addressed the problem of determining ionospheric conditions from swept-frequency oblique backscatter ionograms. The presentation by **Yeh** emphasized the methods developed to obtain an estimate of the equivalent electron density profile that results from the inversion of backscatter measurements.

Goodman discussed the morphological behavior of HF radio waves that propagate through the ionosphere and are observed at satellite levels. Performing calculations using realistic ionospheric models the author described the types of information about the ionospheric structure that could be deduced from satellite measurements of ground-based HF signals or HF noise sources.

The paper by **Prölss** provided a discussion of how classical ground-based ionospheric data coupled with satellite observations can help unravel the problems associated with understanding the dynamics of the ionosphere during a geomagnetic storm. The author pointed out the extemely important fact that the earth's ionozed atmosphere can not be understood and predicted in the absence of comparable knowledge about the neutral atmosphere.

The paper by **Harnischmacher** and **Rawer** addressed the day-to-day fluctuations in the ionospheric parameters and how they appear tied to lunar influences. In addition, the authors discussed apparent planetary influences on the observed behavior of the F2 region critical frequency.

Hanbaba described efforts to determine operationally useful methods of predicting the maximum usable frequency (MUF). The technique developed permits the electron density model given by **Bradley** and **Duderey** to be subjected to analytical expressions in order to determine HF propagation parameters.

TOWARD GLOBAL MONITORING OF THE IONOSPHERE IN REAL TIME BY A BOTTOMSIDE SOUNDING NETWORK: THE GEOPHYSICAL REQUIREMENTS AND THE TECHNOLOGICAL OPPORTUNITY

J.W. Wright and A.K. Paul
Space Environment Laboratory
NOAA Environmental Research Laboratories
Boulder, Colorado 80303

SUMMARY

We take as granted that improved knowledge of the upper atmosphere is necessary for scientific, social, commercial, and defensive purposes; that even as such knowledge advances, practical applications will require continuous monitoring of easily-observable atmospheric properties; that of these, the concentration of ionization between 60 and 400 Km is at once the most informative and easily observable property; and finally, that HF radio sounding retains an important--and perhaps predominant--role in this field of environmental monitoring.

Current developments in ionospheric sounders demonstrate the advantages of fully computer-interactive systems and digital data acquisition, in which all aspects of system control and real-time data processing are software-defined. Priority is at present given to "vertical incidence" capabilities, and particularly to automatic estimation of the electron density distribution (and its time variation) overhead. Two or more such instruments can obtain measurements bistatically to sample ionospheric properties at intermediate locations. Measureable properties, whether by vertical or oblique sounding, include the time and frequency dependence of group path and phase, the signal amplitude, the echo direction of arrival and polarization. Deducible ionospheric properties again include the vertical electron density distribution, but now extend to its spatial and temporal variations over the region. Reduced data may be exchanged among the systems using the same facilities as used for bistatic soundings. Additional information on ionospheric structure is available to each system by ground backscatter and by passive observation of nearby (100-2000 Km) transmitters of known location, frequency, and transmission schedule. Sounder transponders, using telephone lines to return information, can provide additional detail. A global network using ionosonde systems of this kind is virtually certain to evolve naturally out of present interests and capabilities. It will replace the 150-odd obsolete ionosondes now in use with a minimum of about 90 widely-spaced digital ionosonde centers having the capabilities described above, 90 such instruments being sufficient for global bistatic soundings over 2500 Km distances. The "natural" evolution of this program might consume 25 years, and be rather haphazard. Deliberate planning today can reduce this time and improve the resulting system.

1.0 INTRODUCTION

1.1 The Present Ionosonde Network

There are about 150 ionosondes operating on the globe today, at locations identified by the various parts of Figure 1. Only a handfull of these are reasonably modern instruments, the great majority following design concepts which were developed in the 1940's and 1950's. They commit their measurement information to analog film in the form of "ionograms". Even a few manually tuned instruments are still in regular use, requiring full-time operators. Only a small fraction (usually 25%) of the ionograms are reduced to numerical data. Reckoned in terms of directly meaningful geophysical quantities this amounts to some half-dozen parameters manually "digitized" for each hour. Long multi-solar cycle time series of only a few of these parameters are available and from only a few stations. Conversion of ionograms to electron density profiles (N(z)) is only done irregularly, with a few exceptions (Wright and Smith, 1967; IAG-54, 1976, p. 27-29.

This is an unnecessarily wasteful situation, and one unworthy of carrying on through the 1980's. Many of the traditional installations of Figure 1 should be considered by their governing administrations, not as "better than nothing" contributions to an international "network", but as actual and active obstacles to the evolution of a serviceable network. An unfortunately prevalent example of the counterproductivity of the traditional methods is their preservation (usually under government sponsorship) at educational institutions. This becomes more evident when the modern alternatives are understood. Despite the immediate impression which this paragraph might give, it is not intended as an incentive to the abrupt and arbitrary termination of ionosonde stations. Instead, we wish to encourage a planned redistribution of the present resources, leading to a smaller number of modern stations which, individually and collectively, will provide much more information (and do so more rapidly) than can the present network.

It must be emphasized that the stations of Figure 1 function in a largely uncoordinated manner; it is usually only in a long-term (monthly-median) sense that this degree of "global coverage" can even be approached. To a contrasting extreme, Europe, parts of the Soviet Union, India, and the Far East are saturated with traditional instruments. There exist several regional programs (USA, USSR, France, India) for representing and forecasting large-scale ionospheric structure. A "global" program (Thomson and Secan, 1979) is maintained by the AFGWC, based in part upon data from 44 widely-spaced ionosondes (some providing near-real-time parameters). However, this comparatively ambitious effort is considered "data-starved" by its authors.

Apart from such needs for more data is there a need for a modern alternative? We feel that this question is too easily begged by the demonstration that only a fraction of the information obtained from the present network is actually used. (What is so demonstrated, of course, is that most of this 'information' is not useable). The practical need for a modern ionospheric monitoring network is best presented by those who would or could use the information in telecommunications, environmental, defense, and commercial applications. As geophysicists, we can assert with confidence that data from a modern network will certainly be used in geophysics. The data would be in computer-accessible form, and expressed in geophysical parameters, at the point of measurement. Although the geophysical community probably cannot bring about a modern network for "their" goals alone, we believe that these goals serve, and subsume, all of the others.

1.2 The Role of the Ionosonde in Atmospheric Physics

"The ionosphere occupies a key position in solar terrestrial physics because, extending up to the magnetosphere and down to the domain of meteorology, it impinges on both", (Rishbeth and Kohl, 1976). Quite analogously, radio sounding by total reflection occupies a key position in ionospheric measurement: the plasma is a sensitive tracer of virtually all atmospheric processes underway in the ionosphere, and radio wave reflection provides a plasma probe of unparalleled sensitivity. Some of the measurements are useful directly, without conversion to geophysical parameters, as in the representation of long-distance radio communication conditions. They may also be inverted to a variety of atmospheric information, much of it available in no other practical way. The methods can provide information on all scales of the temporal spectrum, extending from the beginning of our awareness of the ionosphere a half-century ago, down through those (of solar or terrestrial origin) occupying fractions of a second. Information describing a wide range of spatial scales is accessible from a single observing location, typically starting with those comparable to the radio wavelengths used (.01 - .3 Km), and extending far beyond the first Fresnel zone scale (typically 3 Km) to horizontal scales at least as large as the height of the ionosphere, 100 - 300 Km.

These features are not obtained without some compromise, of course. Total reflection itself means that the topside ionosphere (and, occasionally, everything above an intense Sporadic E layer) is beyond observation. Historically more significant is the fact that for purposes other than simple radio propagation applications, the wanted geophysical information is only indirectly available from total reflection measurements. Numerical inversion procedures, some of them rather difficult or uncertain, are usually required. To attain the full benefit of total reflection sounding under these circumstances, there are three inescapable requirements:

1. The measurement system must have complete agility in time and frequency sounding patterns, and it must obtain a full description of each radio echo. It follows that the system must be programable and digital.

2. The measurements must be fully reduced to the geophysical information they contain; otherwise, when reduction is incomplete, the interpretation of each part suffers unnecessary ambiguity.

3. Global (or at least regional) and continuous monitoring is necessary, first for obvious reasons identical to those for meteorology, and also because--as suggested by this paper--temporal and spatial continuity of measurement proves the necessary information by which each instrument can adapt its data acquisition pattern to prevailing conditions in the ionosphere.

In the following sections we shall summarize the demonstrated measurements capabilities of the "Dynasonde" class of ionospheric measurement systems (Wright, 1969; Wright and Pitteway, 1979 a, b), which we believe satisfy requirement 1, above. Some natural extensions of these capabilities to provide oblique incidence, backscatter, and passive observations of the ionosphere, each contributing to increased lateral converage, are also described. Examples of available data processing procedures are given to illustrate the present position with regard to requirement 2; we also identify some further developments which are needed. Finally, we attempt to suggest the size, deployment, modes of operation, costs and benefits of a global ionospheric monitoring network satisfying requirement 3.

Two additional introductory comments seem worthwhile. It is frequently proposed that ionospheric predictions and morphology can (eventually, at least, and to some useful approximation) be derived from theoretical modeling, instead of from empirical observation. Without wishing to risk judgement of this question for the distant future, we suggest that at present this view inverts the role which theoretical models can and should play. Furthermore, this view purports to conclude that an ionospheric monitoring network is not needed. Again, the analogy with meteorology is instructive. Ionospheric "weather" is immediately and easily observable by radio sounding, and the problems of (short-term) predictions (i.e., extrapolation) and morphology can be reduced to those of data processing and communications, provided that an adequate observing program is maintained. The proper use of theoretical modeling, we feel, is to extract from empirical observation of the ionosphere those properties of the atmosphere which are difficult to observe directly. We shall suggest some examples later in this paper. Secondly, we wish to emphasize the close relationship to be anticipated between monitoring and research, should a network of modern ionosondes be established. The research worker requires background (or monitoring) data in nearly every investigation, and most observational research programs can augment the "monitoring" data base. An important feature of the Dynasonde class of digital ionosondes is their ability (through suitably programmed operating systems) to maintain scheduled monitoring activities while performing research-oriented measurements as well. Further significant improvement in the practical applications of ionospheric knowledge (e.g. for communications, environmental concerns, detection of unnatural disturbances, etc.) is surely to be gained through improved understanding. And improved understanding demands, we believe a more comprehensive monitoring effort of the kind proposed here.

2.0 DYNASONDE MEASUREMENT CAPABILITIES

2.1 Hardware Aspects of a Modern Digital Ionosonde

The necessary hardware features of Dynasonde systems have been adequately described elsewhere (Wright, 1975, 1977) and need not be reviewed here. It is sufficient to show, with Figure 2, the functional layout of the latest such system as designed and assembled at NOAA-SEL by R. N. Grubb. The general-purpose computer, supported by the usual digital peripherals and by two microcomputers dedicated respectively to real-time signal processing and system timing management, are the essential features by which requirement 1 of our Introduction are achieved. The minicomputer, disk memory and tape provide a large potential for real-time and deferred data processing, in addition to complete flexibility for defining data acquisition patterns. Realization of the full potential of these instruments will require a considerable amount of further software development, and field tests. However, much of this potential has been demonstrated in a prototype Dynasonde (Wright, 1969, 1975, 1977). Much more on the matter of oblique incidence radio measurements has been accomplished elsewhere quite independently of the Dynasonde, but which may be incorporated into the new instruments through software (and perhaps some hardware) extensions.

2.2 Geophysical Functions

The geophysical functions of a digital ionosonde are suggested by Figure 3, where we distinguish among:

Observables including the "dependent variables" which are measured for each ionospheric echo--effectively the group path and phase, and echo amplitude. These (or equivalent) quantities must be obtained as functions of the independent variables for space, time, and radio frequency. The frequency converts directly to plasma density, but the spatial localization of the echo is an involved inversion problem.

Derivables are quantities obtained by calculation from the observables using established theory and (if appropriate) statistically-informative estimation procedures. The status of some of these data inversion procedures was discussed by Wright, 1975, and need not be repeated here. They include methods for obtaining the electron density distribution from multifrequency vertical, oblique, and backscatter soundings, methods for locating and tracking discrete "targets" of medium scale, and methods for deriving ionospheric velocities.

Deducibles include almost all of the parameters, properities and processes of the higher neutral atmosphere and ionosphere. Some of these (e.g. winds, conductivity) are almost directly related to observables, and might be included in the "Derivable" category. Others (e.g. eddy mixing, electric fields) are hardly observable by any direct means, and must be deduced from their effects.

Comparables are quantities or processes which are measurable by means entirely independent of radio sounding. They are often important for validation of data acquisition and inversion methods, and of course both "sides" may benefit in such comparisons.

2.3 The Adaptive N(z,t) Mode

By far the most fundamental item of information to be derived from radio sounding is the electron density profile, N(z). An automatic procedure which optimizes input data selection has been demonstrated (Wright, et. al., 1972), and a simplified but accurate inversion routine for minicomputers is available (Paul, 1977). Awaiting further development in the new digital ionosondes is the adaptive real-time profile measurement system (A-mode) illustrated in Figure 4 for N(z,t). Although there may be about 1000 measurement frequencies in a typical ionogram, less than 50 of these are usually selected (subjectively) for group range input to the inversion procedure. The AUTONH process automates and optimizes this selection, based upon knowledge of the just-proceeding profile. The adaptive procedure of Figure 4 uses the specification of optimum frequencies in a further step, to control the sounding activity itself. By thus separating the functions of "ionogram" sounding (which is still useful for its familiar display of ionospheric structure) from the N(z,t) profile function, much better temporal resolution and continuity in the profile may be achieved. Furthermore, it seems reasonable that the same frequencies as are selected dynamically in A-mode (plus a few others to make the procedure more robust) are also the optimum frequencies for drift, angle-of-arrival, temperature and other estimation procedures.

A few examples are worthwhile to support some of the 'Deducibles' of Figure 3 which are less widely recognized as available by total-reflection sounding.

Ionospheric Temperatures: The plasma temperature (the mean of the electron, ion temperature) controls the F2 layer thickness apart from transient departures due to winds, waves, and electric fields. The scale height and plasma temperature are thus available (after averaging away transitoria) from the curvature of the N(z) profile at the layer peak (Rishbeth and Garriott, 1969, 4.41). That this works in practice is suggested by a number of synoptic studies (Wright, 1963, 1964; Becker and Stubbe, 1963; Rohrbaugh, et. al., 1973) and by the small sample of direct-comparison data in Figure 5. An entirely independent method for obtaining temperatures is based upon the deviative absorption suffered by radio waves reflected near E and F layer peaks (Ganguly, 1974; Sabra and Verkatachari, 1975). An example (from Danilkin, et. al., 1978) is shown in Figure 6; it compares values of collision frequency made near the E peak (96-110 Km) with values computed using gas-kinetic theory and an atmospheric model. The collision frequency and (neutral) temperature are closely related at E region altitudes, but comparative data are desirable to validate the deduction of temperature.

Neutral Composition: The F2 layer electron density depends strongly on the composition ratio $[O]/[N_2]$, and several authors have used this fact to deduce the ratio (and related neutral composition information) from N(z) profiles (Prolss, et. al., 1974, Antoniadias, 1976). The example of Figure 7 (from Stubbe, 1973) compares reitoria) from the curvature of the N(z) profile at the layer with measurements by the OGO-6 satellite and by 6300 airglow. The practical use of such deductions may certainly be questioned, at least until more calibration and modeling have been carried out, but these examples show the proper role of theoretical modeling and the importance of identifying self-consistency in the system illustrated by Figure 3.

Neutral Winds: It is now well established that analysis of the moving radio diffraction pattern, measured at the ground, yields a good estimate of the neutral air motion at the radio reflection level (Wright, et. al., 1975 a,b). This makes the measurement of winds virtually direct and "Derivable", although the statistical analysis procedure is rather lengthy (Fedor, 1967). [A simplified procedure has been developed by Paul (197?: unpublished) for real-time use in a digital ionosonde, and is being tested.] The reason that radio diffraction from the plasma can convey information on the neutral air motion, even when the bulk motions of the two gases may differ, is not fully understood; presumably irregularities in the two gases are closely coupled, even where their motions are not.

Numerous other examples of parameters and phenomena deducible from ionospheric radio sounding may be mentioned:

Sunstorm electric fields (Ruster, 1965)
Energetic particle precipitation (Wright, 1975)
Large-scale atmospheric waves (Thome, 1964,
  Morgan, et. al., 1978; Harper and Gowhill, 1974)
Jet-stream induced waves (Vidal-Madjar, et. al., 1978)
Tsunami-warning (Najita, et. al., 1974)
Atmospheric explosions (Broche, 1977)
Aircraft-induced disturbances (Marcos, 1966)

2.4 The Need for Coordinated Measurements

Returning our attention now to Figure 3, two similar points deserve separate emphasis. Each of them involve the lines among boxes in that diagram rather than the boxes themselves:

The relationships among "Derivable" and "Deducible" properties are identical to the subject matter often termed "theoretical modeling" of the ionosphere and high atmosphere. Although particular links between boxes must often be developed in isolation, it is the degree of success of the entire pattern as a self-consistent system which expresses our state of knowledge. If important links are omitted in the modeling, or are inaccessible because of measurement system limitations, confidence in the entire structure suffers. An important advantage of HF sounding is the fact that it can "stand alone", if need be, in providing the essential input to most of Figure 3.

Secondly, and specifically regarding total reflection sounding, we must emphasize the interdependent character of the 'observables' in their relation to the derivables. Earlier sounding systems which (for example) measured ionospheric absorption, or doppler, or direction-of-arrival, or drifts, at one frequency only, were exploratory and developmental efforts of pioneering value, but it is only through doing all of these in concert that the modern digital ionosonde can be expected to perform adequately for the geophysical functions of Figure 3.

An extrapolation of this second point provides the main theme for the present paper. A very small number of digital ionosondes, no matter how sophisticated, can of course not satisfy requirements for global monitoring of propagation, ionospheric, or geophysical conditions. But a threshold exists at which a moderate number of competent instruments, operating with real-time coordination, can do so. In addition to the "vertical-incidence" measurement capabilities just described, the same hardware and data processing facilities are applicable to measurement of lateral variability in the ionosphere with relatively little modification, as discussed in the next section.

3.0 METHODS FOR LATERAL OBSERVATIONS

Ionosondes with typical power, sensitivity, and antenna systems have a "field of view" roughly comparable to the radius of the station labels in Figure 1. This is about 500 km for the F region, and proportionately less for the lower ionosphere. A horizontal component of the local ionization gradient displaces the condition for total reflection from "overhead", and variations of the gradient, if large enough, can cause multiple specular points to occur. These are abundant and self-evident in routine ionograms, but in such recordings, lacking any information concerning echo direction of arrival, the lateral and vertical structure is usually confounded. This is another example of the counterproductivity of the standard ionosonde.

In modern instruments (Figure 1) accurate measurement of echo complex amplitude at a closely-spaced array of receiving antennas permits calculation of the direction of arrival of each echo. (Paul, et. al., 1974; Wright and Pitteway, 1979b). If, in addition to the "vertical-incidence" transmitting antennas, there are available broad-band antennas producing low-elevation radiation in several azimuths, the same ionosonde can obtain a much greater lateral view by a variety of well-established or potentially valuable methods. Figure 8 suggests the relationships among these, and gives them pseudo-acronyms for the sake of brevity in identification.

3.1 DYNASND

Measurement of the zenith and azimuth angles of arrival during ionogram or adaptive soundings provides information on the nature and magnitude of ionosphere tilts and horizontal gradients. If the tilt changes slowly with radio frequency (height) or time, a large-scale structure is implied, small-scale structures produce correspondingly smaller time and height scales. Qualitatively different effects occur for intense structure of small-scale: more than one normal-incidence reflection point may occur, and looping ray paths (not involving normal incidence) are found occasionally (Paul, Smith and Wright, 1968). Doppler information may provide an extra dimension of discrimination when multiple ray paths are known or suspected (Pfister 1971, Brownlie, et. al., 1973).

Methods for converting this lateral information into a quantitative 3-dimensional description of the local ionosphere are not fully developed. Distinct targets (e.g. Cs$^+$ clouds, Wright, 1974) can be located and tracked by "echolocation" calculations from spaced antenna data, and the same techniques are clearly applicable to electron clouds produced by energetic particle precipitation and to other discrete "sporadic E" patches. Irregularities with "soft" boundaries embedded in the higher ionosphere are more difficult to observe and to interpret They must be divided according to whether their electron density exceeds the background density, or is exceeded by it. Depletions tend to be more observable than ionization enhancements since the far side of a hole provides a concave focussing surface. Conversely, enhancements create defocussing surfaces, and unless these are below about 150 Km altitude (where small-scale irregularities give them a rough scattering surface) they may escape detection by ionosondes (Wright and Paul, unpublished manuscript describing Ba$^+$ cloud experiment, at various altitudes). Even when abundant echoes from an embedded disturbance are available, no direct inversion procedure is available to correct the data to a description of the irregularity; iterative ray-tracing, to "home" on several distinctive properties of the data (group range, angle-of-arrival, frequency-dependence, observing location) is necessary (Paul, Smith, Wright 1978). For much larger gradient scale lengths, a method is needed to combine electron density profile inversion with echo-location. A first-order correction for layer tilts is easy. the echo-location calculation yields the "group path vector" (Paul, Wright and Fedor, 1974) and the vertical component of the vector can simply be used for h' in the profile inversion. While this approach may somewhat improve the profile estimates (in the sense of yielding profiles which are more consistently "vertical") it is only an approximation, and it yields little direct information on the lateral gradients except when the large scale of a tilt is self-evident. The analysis can be carried further by ray tracing at the measured angle-of-arrival into the "vertical' profile introducing enough "tilt" to obtain normal incidence, reflection at the critical ionization contour. The logic of such a procedure is suggested by Figure 6. The analysis then requires iteration. These methods can be expected to provide an acceptably accurate description of lateral variability within a 100 Km radius.

3.2 OPORTUNE:

Passive observations of "transmitters of opportunity" might provide useful supplementary information concerning lateral gradients and traveling disturbances, at only the "cost" of data processing. For remote transmitters of known schedule, location and frequency stability, the digital ionosonde can measure the presence, field strength, Doppler, and angle-of-arrival of the signal. A minimum of about 0.5 seconds observation per frequency should be sufficient for this, divided equally between the same two orthogonal pairs of receiving antennas as are used for 'vertical' sounding. However, it should be noted that Baulch and Butcher (1978), who have performed CW angle-of-arrival experiments of this kind, suggest that much longer observation times (> 10 sec.) are desirable to reduce effects of mixed 1, 2 hop signals. OPORTUNE might in principle be applied to signals originating out to the 1 hop, limit (2000-4000 Km), but as this implies low arrival angles, short wavelengths, and vertical polarization, special antennas would be needed It is probably better to advocate the method for shorter distances, say between 200 and 1000 Km ground range, until its usefulness and limitations are better established. If, in each of six azimuths, five transmitters at ranges of about 200, 400, 600, 800 and 1000 Km were available, OPORTUNE would require 15 seconds for data acquisition. The sampling of each azimuth should be followed by its 180° complement, since the information from such paired directions is most easily combined with the "vertical" sounding results.

3.3 BAKSCTR

With transmitting antennas which favor a particular azimuth and an intermediate zenith angle, oblique backscatter at frequencies exceeding the overhead critical frequency may be observed. About all that is seen by systems of ordinary power is the so called "leading edge backscatter"--the envelope of skip-distance or minimum-time-delay scatter, and even that is seldom available beyond half-paths of about 1500 Km. Within the observable range, however, the trace provides nearly direct information on the variation of maximum density with distance from the observing location. Computer simulation and the development of backscatter data analysis is being pursued actively by the University of Illinois (DuBroff, et. al., 1978); backscatter measurements have been used to map foF2 within a radius of 1500 Km, by Hatfield (1970). The analysis is again iterative; starting from the local vertical distribution, it attempts to match group range by ray tracing in a model with adjustable gradients of peak density, height, and thickness. DuBroff, et. al. (1978) show that these can be recovered with acceptable accuracy from the data. However, the methods require considerable computer time, even in a large machine, and uniqueness suffers in the presence of quadratic gradients and underlying ionization.

For tracking sporadic E clouds these refraction problems do not arise, and backscatter can be a powerful tool for detecting such clouds, estimating their diameters, velocities, and sizes (Tanaka, 1979).

3.4 BISTAT:

Bistatic oblique sounding over one-hop paths provides a means for probing the ionosphere at the path midpoint. If transmission and reception in both directions is performed, the absolute group path delay is the mean value of the two observed echo delays measured from the local transmission times. The only timing accuracy required is that which is sufficient to assure frequency synchronization, and is of the order of the repetition interval. Numerous experimental programs of oblique sounding have been carried out (Davies, 1965, § 4.41), and some specific propagation paths are routinely monitored by the chirp version of the hardware (Barry & Fenwick, 1969). Inversion of oblique ionograms to midpoint N(z) profiles is not fundamentally different from the vertical sounding problems, but some quantitative distinctions are important:

1. For the same electron density at "reflection", a higher radio frquency is required, in proportion to the secant of the take-off zenith angle.

2. Magnetoionic effects, important for separate consideration of ordinary" and "extraordinary" propagation modes at vertical incidence, become steadily less important as the distance of oblique propagation increases. Partly, this results from the use of higher frequencies in relation to the electron gyrofrequency. But more important is the averaging effect of the larger scattering region, which tends to blur the O/X resolution in the measurements (Barry and Fenwick, 1969); since O and X are affected oppositely--if not equally--from the no-field case, neglect of the magnetic field (in ray tracing, for example) offers a valuable simplification.

3. "Screening" of a higher part of the ionosphere by a lower part (or by the horizon itself) can be troublesome for profile inversion. The problem is similar to valley or sporadic E blanketing in the vertical-incidence case, but can be more prevalent at oblique incidence because the layer density and height effects are combined in determining the screening height.

4. Horizontal gradients may cause significant displacements of the 'reflection points' from the path midpoint, expecially for the longer one-hop paths contemplated here. It is not known whether measurements of angle-of-arrival will provide useful additional information by which these gradients can be estimated.

3.5 TRNSPND

Simplified systems (receive and digitize only) can be located at distances of 50 - 500 Km from each ionosonde location, on various azimuths--geography permitting. These could provide higher spatial resolution in certain regions.

4.0 CONSIDERATIONS AFFECTING NETWORK DEPLOYMENT

4.1 Idealizations

A purely geometric approach to station deployment, while obviously impractical to realize, provides a few useful insights. The ratio of the earth's surface area (radius R) to that observable from a single site (radius r) is 4 R^2/r^2); if r is taken to be 1000 Km, 162 sites are required, 113 for r = 1200 Km, and 72 if r = 1500 Km. On the other hand, the Platonic solids provide relationships among the number of vertices, edges meeting at vertices, etc. The maximum possible number of equidistant vertices touching the surface of a sphere is 20 (the dodecahedron), but only three edges (directions for bistatic sounding) meet at each vertex and obviously the interstation distances are far too great for one-hop propagation. The icosahedron (12 vertices) affords six equidistant directions, which seems ideal, but the distances are still greater.

A satisfactory plan may be based upon the icosahedron by dividing each edge by 3 and adding one vertex near the center of each equilateral face. This produces isoceles triangles in each face, as shown by the sketch at lower right in Figure 9. If the earth were divided in this way, 92 station locations would result, with three characteristic interstation distances. 2351, 2598, and 2722 Km. These distances are all comfortable for bistatic oblique sounding of the F region. In addition to the 92 observation sites for "vertical" observations, the contiguous-pair midpoints provide 260 additional measurement locations. Each of the 12 icosahedral vertices "sees" five interpolated sites at equal azimuths of 72°, while each of the interpolated sites sees six sites at 60° azimuths. The locations are represented in Mercator projection in the main part of Figure 9.

It has appeared to us less than obvious that a global network of this size could conduct scheduled bistatic soundings (in a minimum cycle of six azimuths each) between pairs of stations without contention for the same observing period in the cycle. Each sounder must select the appropriate transmit/receive antenna for the azimuth of its neighbor, without being required to serve another azimuth during the same period. The small numerals of Figure 9, at the midpoints of each station pair, suggest the extent to which this may be accomplished. Contention appears at five of the 260 midpoints (wherever the number of midpoints in a closed loop is odd) but otherwise an efficient and compatible schedule for bistatic and backscatter soundings seems attainable.

4.2 The Geophysical Requirements

There are various ways of weighting the requirements for an ionospheric monitoring system, not the least important of which is that the system be realistic. Thus, we give weight to global monitoring of the F region with good fidelity, while acknowledging that an equally faithful description of global sporadic E structures is probably unrealistic. For these we must be satisfied with the statistical data from regional sampling.

The correlation distance for foF_2 (noontime hourly data), for North-South and East West directions, and for three seasons, is shown in Figure 10, from Rush and Edwards (1976). A 50% improvement of prediction of foF2 at an unmeasured location, based upon data from a measured location, requires a correlation of 0.87 between the two locations (Gautier and Zacharisen, 1965). From Figure 10, station separations should lie in the range 1000 - 2000 Km to achieve this. An improvement of prediction confidence can be accomplished at greater station separations, however, by increasing the correlation between them. We believe that this can be accomplished by estimation of the local foF2 gradient at each location, together with profile estimation at each midpoint.

Some recent results suggest a more definitive basis for deciding the geophysical requirements for station separation. Using Anharmonic Frequency Analysis (Paul, 1972), the temporal variations (diurnal and seasonal) of foF2 can be described by a small number of spectral components (Paul, 1978). About 9 to 11 spectral lines account for about 90% of the monthly median foF2 variation; they have a geographic variation which is more smooth than foF2 itself, but which still show a strong dependence on solar activity. Figure 11 represents the latitude profile of the strongest component (the constant or zero-frequency average value) at two longitudes and two sunspot numbers. Figures 12 and 13 represent the second-strongest component (the 24 hour or diurnal line) at the two longitudes respectively, and for the same two levels of solar activity. The smallest spatial scales evident here are in the variation with latitude near the "equatorial anomaly" and are of the order of 10° of latitude or 1100 Km. These features move in latitude, with sunspot number, a distance comparable to their width.

The latitudinal gradient (dominated by the constant component, Figure 11) is about twice as large as the longitudinal gradient, which is dominated by the diurnal variation (Figures 12, 13), but the equatorial anomaly makes its own contribution to the longitudinal gradients in all components as they depend upon whether the magnetic equator is north (longitude 130°) or south (295°) of the geographic equator.

We conclude that a station separation of about 2500 Km in latitude is adequate for even the equatorial anomaly, assuming that bistatic sounding permits interpolation at the station midpoints, and that a somewhat larger separation is permissible in longitude except where the two equators are crossing at a steep angle (e.g. South America).

5.0 THE DEPLOYMENT OF PRACTICAL MODERN NETWORK

Figure 14 provides a global view of an "optimum" network of modern ionosonde instruments, based upon the requirements and technological opportunities reviewed above. At about 89 locations, we envision instruments capable of providing all of the detailed "vertical incidence" sounding information of the advanced digital ionosonde, plus additional information on lateral structure and gradients near each site, plus bistatic soundings between nearby pairs of sites.

Typical interstation distances in Figure 14 are about 2500 Km, although many (e.g. Dublin-Prague) are less than 2000 Km, and some, still useable for oblique soundings, are between 3000 and 4000 Km. (Seattle-Hawaii, Aricebo-Azores).

In the deployment of Figure 14 about 24 locations are "mandatory", being isolated islands the absence of which would leave unnecessary and serious gaps in global coverage. A few such gaps are unavoidable anyway. Table 1 lists these with their approximate dimensions (they are all oceanic), and lists for comparison the large gaps of the present network (Figure 1).

TABLE 1

Gaps in Global Ionospheric Monitoring

Present Network (Fig. 1)	An Optimum Network (Fig. 14)
S. Pacific Ocean (8000 x 12000 Km)	S. Pacific Basin (5500 x 9000 Km)
E. Equatorial Pacific (8000 x 10000)	E. Equatorial Pacific (5000 x 7000)
Indian Ocean (6500 x 9000)	S. Australian Basin (4000 x 5000)
S. Atlantic Ocean (6500 x 6500)	Mid Atlantic Basin (4000 x 5000)
N. Polar Cap (2000 x 6000)	N.W. Pacific Basin (3500 x 5000)
Africa (4000 x 6000)	
S. America (3000 x 6000)	
China (3500 x 4000)	
N. America (2000 x 4000)	

A list of the station locations of Figures 1 and 14 are given in Table 2. The 150 traditional stations, and 89 locations of the modern plan, have 41 places in common (of the 24 mandatory island locations, five are presently instrumented). Eleven more of the 89 were once occupied by ionosondes and have become inactive. This overlap of the present network with our plan is intentional,

TABLE 2
PRESENTLY ACTIVE IONOSONDE STATIONS AND A SUGGESTED OPTIMUM NETWORK

Present stations		Suggested Stations
Aberystwyth (Wales)		
Accra (Ghana)		
	P,M	Adak (USA)
		Adelaide (Aust.)
Ahmedabad (India)		
Akita (Japan)		
		Alexandrovsk (USSR)
Alma Ata (USSR)		
		Angola
		Ankara
Aricebo (USA)		Aricebo
Argentine Isl.	M	Argentine Isl.
	M	Ascension Isl.
Ashkabad		Ashkhabad
Athens		
Auckland		Auckland
	M	Azores
Bangkok		Bangkok
Baudouin (Ant)		Baudouin (Ant)
Belgrano (Ant)		
Beograd		
Billerica (USA)		
	P	Bogata
Bombay		
Boulder		Boulder
	D,M	Bouvet Isl.
Bribie Isl.		
Brisbane		Brisbane
Budapest		
Buenos Aires		Buenos Aires
Calcutta		
Campbell Isl.		
Canberra		
Cape Parry		
Cape Schmidt		
		Capetown
Cape Zevgari		
		Carnarvon
		Chad
Christchurch		
		Chokurdakh
	P,D,M	Christmas Isl.
Churchill		Churchill
		Cocos Isl.
College		College
Concepcion		Concepcion
Dakar		Dakar
		Dar es Salam
		Darwin
Delhi		Delhi
DeBilt		
	D,M	Diego Garcia Isl.
Dixon		
Djibouti		Djibouti
Dourbes		
		Dublin
Dushanbe		
	D,M	Easter Isl.
	P,M	Fiji
Forteleza		Forteleza
Fort Monmouth		
Freiburg		
	D,M	Galapagos
Garchy		
Godhavn		
Goose Bay		
Gorky		
Grahmstown		
Gratz		
	P,M	Guam Isl.
Halley Bay		Halley Bay
Hanover N.H.		
Heiss Isl.		
Hermanus		
Highgate Springs		
Hobart		
Hong Kong		
Huancayo		
Ibadan		Ibadan
	D,M	Camp Heurtin (Ile Amsterdam)
Irkutsk		Irkutsk
Istanbul		
Jicamarca		Jicamarca
Johannesburg		
Juliusruh		
Kaliningrad		
Karaganda		
Kerguelen	M	Kerguelen
Khabarovsk		
		Khaybar (Saudi Arabia)
Kiev		
Kiruna		
Kodiakanal		
		Kouru
	P,M	Kwajalein Isl.
Lancaster		
	P	Lanchow
Lannion		
		LaPaz, Mex.
Leicester		
Leningrad		
Lindau		
		Lisbon
Lycksele		
	P,M	Macquarie Isl.
Madras		
Magadan		Magadan
Manila		
	P,M	Marion Isl.
	M	Marquesas Isl
	D	Mato Grosso, Brazil
Maui	M	Maui
	M	Mauritius
Mawson		
		Merida, Yucatan
Mexico City		
	M	Midway
Miedzeszyn		
Millstone Hill		
Mirny		Mirny
Moscow		Moscow
Mundaring		
Murmansk		
	D,M	Muscat
Nairobi		
Narssarssuaq		Narssarssuaq
Norfolk, Isl.		
Norilsk		Norilsk
Novokazalinsk		
Novosibirsk		
Nurmijarvi		
		Omsk
Okinawa		
Ottawa		
Oagadougou		
		Palermo, Sicily
	P	Paramaribo
Patrick AFB		
	P,D,M	Pitcairn
Point Arguello		
Poitiers		
		Prague
Port Stanley		Port Stanley
		Post Maurice Cartier (Algeria)
Prumenice		
Raratonga	M	Raratonga
Resolute Bay		

Present stations		Suggested Stations
		Rio de Janerio
Rome		
Rostov		
Sachs Harbor		
Salerkhard		
Salisbury	P	Salisbury
Sanae		
	D,M	San Ambrosio Isl. (Chile)
San Jose		
San Juan (Arg.)		
Scott Base		
Sondrestromfiord		
	P	Seattle
Seoul		
Slough		
Sodyankyla		
Sofia		
South Georgia		
South Uist	P	South Pole
		Stanleyville
Sverdlovsk		
Syowa		Syowa
Tahiti		
Taipai		Taipai
Tangerang		Tangerang
Tashkent		
Tbilisi		
Terre Adelie		Terre Adelie
Thule		Thule
Thumba		
Tiruchirapalli		
Tixie Bay		
Tokyo		Tokyo
Tomsk		
Tortosa		
Townesville		

Present stations		Suggested Stations
Trelew		
	D,M	Tristan de Cunha
Trivandrum		
Tromso		Tromso
Tsumeb		
Tucuman		
Tunguska		
Uppsala		
Ushaia		
Vanimo		
Vostok		
Wakkani		
Wallops		Wallops
White Sands		
Yakutsk		
Yamagawa		

Key:

P	Previously instrumented
D	Difficult logistically
M	Manditory location

reflecting our presumption that an existing station has a motivated administration behind it, has experienced personnel, and has a practical site. Stations marked "D" in our list are judged "difficult" logistically, but several of these (e.g. Pitcairn Island, Cocos Island) have once hosted an ionosonde. Many of our choices of location are quite arbitrary--and they are almost entirely so, within a 500 Km radius of the nominal station name, or about the radius of the station label shown in Figure 14. However, a significant relocation of any one station in our plan must be accomodated by smaller relocations of adjacent stations, or perhaps by adding or deleting a location and thus improving or compromising regional coverage. There are numerous locations at which stations could be added to improve local coverage. Typical additional locations (given those of Table 2, Figure 14) might be: Central France, Western Egypt, Western China, Mahi Island, Central Australia, Iceland. Where existing long time-series of observations, (e.g. Moscow, Washington, or a permanently important geophysical observatory (Aricebo, Tromso, Jicamarca) compel a choice of location, freedom to select surrounding locations is further restricted. With these points clear, we should not wish our suggestion of optimum locations to express any deeper political considerations, and we hope that it is clearly understood that any rearrangement of locations which preserve approximately the suggested station density, is equally satisfactory.

6.0 CONCLUSION

What steps are needed to bring about a plan of this kind? Setting aside practicalities temporarily, here are some significant technical problems worthy of study, approximately in order of decreasing priority:

1. The local-lateral-structure capabilities of the digital ionosonde require development and demonstration. The use of echolocation to describe tilts (with, or without, ray-tracing refinements) must be combined with existing programs for deducing the "vertical" structure (profile inversion). The methods should be tested against a satellite plasma probe; if the satellite is within a few scale heights of the (topside) F2 peak, the local, horizontal, log-gradient there should be equal to that at and near the peak.

2. A-mode, the adaptive, real-time (N(z,t) capability of the Dynasonde (Figure 4) should be demonstrated.

3. The suitability of A-mode sampling frequencies (which will change dynamically, as the ionosphere itself changes) for defining other height and time-dependent derivables, must be tested and demonstrated. Such derivables are tilts, drifts, $\Delta N/N$

4. A sampling method for deducing drifts (Paul, unpublished) requires further tests against the well-tested "Kinesonde" mode, so that the lengthy time-series (and time-series analysis) of the latter method are obviated.

5. With pairs of digital ionosondes, and preferably with at least a standard ionosonde at the midpoint, the oblique-incidence (BISTAT, BAKSCTR) modes of measurement, and data reduction, need development and demonstration. This is a large subject, but fortunately there have been a number of vigorous efforts to develop the necessary data inversion methods. (Croft, 1972; DuBroff, et. al., 1978; Nielson and Watt, 1972, Rao, et. al., 1976; Rao and Hoover, 1975, Smith, 1970; Tanaka, 1979.)

Some of these studies (1, 2, 3, 4) are essential to the research application of the new digital ionosondes individually, and will be pursued at NOAA and elsewhere in the natural course of development of Dynasonde software. The oblique incidence capabilities of item 5 (and other items of lower priority such as OPRTUNE) present less scientific than technical interest and may require specific administrative encouragement as part of a determined intent to modernize an administration's ionospheric monitoring program.

Turning to practicalities, the main questions are costs for hardware, installation and operations. The NOAA instrument (Figure 2) represents a hardware investment of about $120,000. A commercial equivalent might cost twice that, but the system design and its capability are probably somewhat overelaborate, at least for (say) half the stations of our plan. That design was intended from the beginning to provide a test-bed for refinement of ionosonde methods. We believe that as microprocessors continue to decrease in cost, while approaching the larger minicomputers in speed, memory size, and processing capability, the cost for the basic "vertical" incidence instrument should be kept below $100,000 in commercial or large-scale production.

We must emphasize that there is no need for absolute uniformity among the digital ionosonde instruments comprising our proposed network. Stations offering occasional or frequent research facilities (or an ionosonde research center, in the manner of incoherent scatter centers such as Aricebo) will of course need relatively elaborate systems. Very isolated stations (gap-fillers) will need only a minimal system, to which additional hardware could easily be interfaced temporarily if the station becomes central to a geophysical campaign. The minimal requirements for any station are easily stated:

a) Programmability, since we must assume, especially for modern instruments entering the network soon, that we really do not yet know the "best way" to make a given measurement, or compute a wanted parameter; software programmability is our best insurance policy.

b) Synthesized, and stable, radio frequencies.

c) Digitization of all echo parameters, of course, and with adequate resolution.

d) Modularity, affording easy maintenance, long MTBF and low cost.

Some particular comments are applicable to educational and government administrations which have become accustomed to the operation of their "own" ionosonde, and who would find it not needed in our plan.

a) Clearly an economy of radio sounding activity is desirable on grounds of efficient use of the radio spectrum and minimizing radio interference; these considerations alone may eventually make our network suggestions imperative.

b) National administrations in Europe realized at the beginning of large-scale space research, that a proliferation of individual national centers could be a weakening policy, as compared with the cooperative establishment of ESRO. A similar approach is worth consideration for ionospheric sounding centers.

c) Modern communications permit large numbers of widely separated users to enjoy the use of one, central, large, computing facility, almost as if it were their's alone. This can apply as well in multiple access to an ionospheric sounding facility, to obtain the latest data, or a summary, or to take active control of a share of the ionosonde's measurement potential. Surely, these possibilities, incorporating the advanced measurement ability of the center itself, must outweigh present advantages of operating one's own analog ionosonde.

d) At an intermediate level we consider the educational institution which can "fill a gap" and can identify ionospheric research as an important specialization. To such institutions the modern ionosonde is a small but complete data processing center, in addition to its primary sounding functions.

The most essential ingredient of our plan is again administrative, and as geophysicists rather than administrators we are unable to provide it. We refer to the planning, funding, and coordination required to bring about a network of the kind and capability described in this paper, on a global or regional scale in a reasonable time. Such a network is virtually certain to evolve eventually anyway, say within the next 20 years, because of the progressive obsolescence of present ionosondes and the continuing need for monitoring ionospheric "weather". But we suggest that if a few national or multinational efforts (e.g. by NATO) were coordinated internationally by URSI, a regional demonstration network could be realized within five years. The merits of expanding or duplicating the demonstration network could then be readily assessed.

REFERENCES

Antoniadias, D.A. (1976), Determination of Thermospheric quantities from simple ionospheric observations using numerical simulation. J. Atmos. Terr. Phys. 38, p. 187.

Argo, Paul E. (1979), Prophet: an application of propagation forecasting principles. Proc. Intern. Solar-Terrest. prediction workshop, R.F. Donnelly, Ed. (in press).

Barry, G.H. and R.B Fenwick (1969), Oblique Chirp Sounding. In. AGARD Conf. Proc. #13, Oblique Ionospheric Radiowave Propagation, p. 487.

Baulch, R. N. E. and L. C. Butcher (1978), Direction of arrival of radio waves reflected from the E region. J. Atmo. Terr. Phys., 40, p. 1235.

Becker, W. and P. Stubbe (1963), Temperature variations in the F region during the last sunspot cycle. Proc Intern. Conf on the Ionosphere, Inst. of Physics and the Physical Society, London.

Brochue, P. (197), Propagation des ondes acoustico-gravitationelles excitèes par des explosions. Ann Geophys. 33, p. 281.

Brownlie, D. D., L. G. Dryburgh and J. D. Whitehead (1973), J. Atmos. Terr. Phys. 35, p. 2147.

Budden, K. (1961), Radio waves in the ionosphere, Cambridge Univ. Press.

Chernov, Yu. A. (1971), 'Backscatter Ionospheric Sounding", Svyaz Press, Moscow, JPRS 54893, January 1972.

Chuang, S. L. and K. C. Yeh (1977), A method for inverting oblique sounding data in the ionosphere. Rad. Sci. 12, p. 135.

Croft, T. A. (1972), Skywave backscatter: A means for observing our environment at great distances. Rev. Geophys. and Space Physics 10, p. 73.

Danilkin, N.P., P.F. Denysenko, V.I. Vodolazkin, S.M. Sushchiy and Y.U.N. Fayer (1978), Effective Collision Frequency in the E Region of The Ionosphere. Geomag & Aeronomy 18, p. 97.

Davies, K. and J. Jones (1971), Three-dimensional observations of traveling ionospheric disturbances, J. Atmos. Terr. Phys., 33, 39-46.

DuBroff, R E. , N. N. Rao, K. C. Yeh (1978), Backscatter ionogram inversion by ray-tracing methods. RADC-TR-78-36.

Fedor (1967), A statistical approach to the determination of three-dimensional ionospheric drifts. J. Geophys. Res. 72, 5401-5415.

Ganguly, S. (1974), Estimation of electron collision frequency in the lower thermosphere. J. Geophys. Res. 79, p. 3235.

Gautier, T. N. and D H. Zacharisen (1965), Use of Space and Time Correlation in Short-term Ionospheric Predicionts. Conf. Record, First Annual IEEE Communications Convention.

Harper, R. M. and S. A. Bowhill (1974), Digital-ionosonde studies of F-region waves. U. Ill. Aeronomy report No. 60.

Hatfield, V. E. (1970), Derivation of ionospheric parameters from backscatter data. In: Ionospheric Forecasting, AGARD Conf: Proc. #49.

King, G. A. M. (1969), Re validity of the 'overlay' technique for analysing the F1, F2 transition. J. Atmos. Terr. Phys. 31, p. 515.

Marcos, F. A. (1966), Aircraft-induced ionospheric disturbances. Air Force Surveys in Geophysics No. 175, AFCRL (now AFGL).

Morgan, M. G., H. J. Calderon and K. A. Ballard (1978), Techniques for the study of TID's with multi-station rapid-run ionosondes. Radio Science 13, p. 729.

Najita, K., P. F. Weaver and P. C. Yuen (1974), A tsunami warning system using an ionospheric technique. Proc. IEEE 62, p. 563.

Nielson, D. L. and T. M. Watt (1972), Ionospheric profile inversion using oblique-incidence ionograms. In: Mathematics of Profile Inversion, NASA Tech. Memorandum TM X-62, 150 L. Colin, editor, pp. 4-61.

Paul, A. K., G. H. Smith, and J. W. Wright (1968). Ray-tracing synthesis of ionogram observations of a large local disturbance in the ionosphere. Radio Science, 3, pp. 15-25.

Paul, A. K. (1973), Radar ionospheric refraction corrections by a digital ionosonde, Radio Sci 8, 1119-1124.

Paul, A. K., J. W. Wright and L. S. Fedor (1973), The interpretation of ionospheric radio drift measurements--VI. Angle-of-arrival and group path [echolocation] measurements from digitized ionospheric soundings: The group path vector, J. Atmos. Terr. Phys. 36, 193-214.

Paul, A. K. (1977). A simplified inversion procedure for calculating electron density profiles from ionograms for use with minicomputers. Radio Science, 12, pp. 119-122.

Pfister, W. (1971). The wave-like nature of inhomogeneities in the E-region. J. Atmos. Terr. Phys., 33, pp. 999-1025.

Prolss, G. M., K. H. Fricke and U. von Zahn (1974), Observations during an ionospheric storm in late October, 1973. COSPAR, Sao Paulo, Brazil, 1974.

Rao, N. N. and K. E. Hoover (1975), Derivation of ionospheric layer parameters from electron angles or arrival of HF radio waves. J. Atmos. Terr. Phys. 37, p. 1167.

Rao, N. N., K. C. Yeh, M. Y. Youakim, K. E. Hoover, P. Parnam and R. E. DuBroff (1976), Techniques for determining ionospheric structure from oblique radio propagation measurements. RADC-TR-76-401.

Rishbeth, H. and O. K. Garriott (1969), Introduction to Ionospheric Physics, Academic Press, New York and London.

Rishbeth, H. and H. Kohl (1976), Topical questions of ionospheric physics: A working group report. J. Atmos. Terr. Phys. 38, p. 775.

Ronrbaugh, J. L., W. E. Swartz and J. S. Nisbet (1973), "Comparison of incoherent scatter and Ionosonde Measurements of Temperature with Calcium Plage and 2800 MHz intensities." J. Geophys. Res. 78, p. 281.

Rush, C. M. (1976), An ionospheric observation network for use in short-term propagation predictions. Telecom J. 43, p. 544

Rush, C. M. and W. R. Edmunds, Jr. (1976), An automated mapping technique for representing the hourly behavior of the ionosphere. Rad. Sci. 11, p. 931.

Ruster, R. (1965), Ionospheric parameters during two successive polar magnetic substorms. J. Atmos. Terr. Phys. 27, p. 1229.

Saha, A. J. and K. Venkatachari (1975), On extraction of collision frequency and temperature in F layer from absorption measurements. Ind. J. of Radio and Space Physics 4, p. 310.

Smith, M. S. (1970), The calculation of ionosphere profiles from data given on oblique incidence ionograms. J. Atmos. Terr. Phys. 32, p. 1047.

Stubbe, P. (1973), Atmospheric Parameters from Ionospheric Measurements. J. Geophys. Res. 79, p. 1043.

Tanaka, T. (1979), Skywave backscatter observations of Sporadic E over Japan. J. Atmos. Terr. Phys 41, p. 203.

Thome, G. (1964), Incoherent Scatter Observations of traveling ionospheric disturbances. J. Geophys. Res. 69, p. 4047.

Thomson, R. L. and J. A. Secan (1979), "Geophysical Forecasting at AFCWC". Proc. Intern. Solar-Terres. Predictions Workshop, R. F. Donnelly, Ed. (in press).

UAG-54 (1976), Catalog of ionospheric Vertical sounding data. World Data Center A, Solar-Terr. Physics NOAA, Boulder, Colorado.

Vidal-Madjar, D., F. Bertin and J. Testud (1978), Sur le Jet-stream de la tropopause en tant que source des ondes de gravitè observee dans la Thermosphere. Ann. Geophys. 34, p. 1.

Wright, J. W. (1963), Temperature control of the structure and variations of Re quiet ionosphere. Proc Intern. Conf. on the Ionosphere, Imperial College, Lon. July 1962. Publ, Inst. Phys. and the Physical Society, Lon.

Wright, J. W. (1964), On the implication of diurnal, seasonal, and geographical variations in composition of the high atmosphere, from F-region measurements. Electron Density Distribution in Ionosphere and Exosphere, Proc. NATO Adv. Study Inst. North Holland Publishing Co., Amsterdam.

Wright, J. W., and G. H. Smith (1967). Review of current methods for obtaining electron density profiles from ionograms. Radio Science 2, pp 1119-1125.

Wright, J. W. (1969), Some Current Developments in radio systems for sounding ionospheric structure and motions, Proc. of the IEEE 57, 481-486.

Wright, J. W., A. R. Laird, D. Obitts, E. J. Violette, and D. McKinnis (1972), Automatic N(h,t) profiles of the ionosphere with a digital ionosonde, Radio Sci. 7, 1033-1043.

Wright, J. W. (1974), Kinesonde studies of Cesium ion Clouds in the E-region. In: Methods of measurements and results of lower ionosphere structure, COSPAR, Akad. Verl., Berlin.

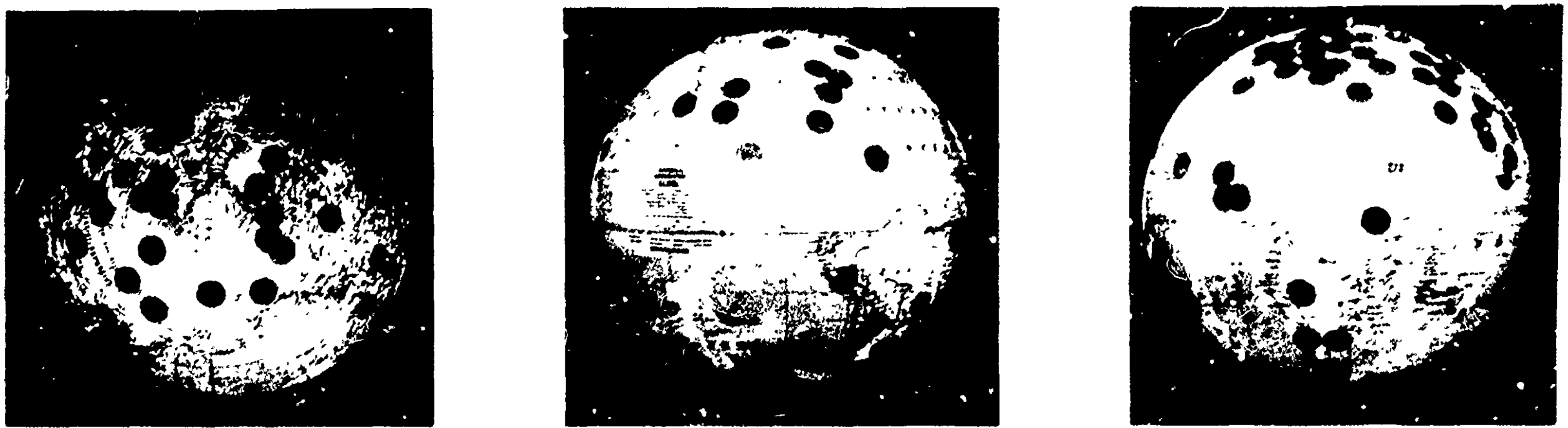
Wright, J. W. (1975), Development of Systems for Remote Sensing of Ionospheric Structure and Dynamics: Functional Characteristics and Applications of the 'Dynasonde', NOAA ERL SEL 206, U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration Environmental Research Laboratories.

Wright, J. W., M. Glass and A. Spizzichino (1975b), The interpretation of ionospheric radio drift measurements--VIII. Direct comparisons of meteor radar winds and kinesonde measurements: Mean and random motions. J. Atmos. Terr. Phys., 38, 721-730.

Wright, Vasseur and Amayenc (1975c), Direct comparisons between field-aligned ion drifts by incoherent scatter and kinesonde spaced-antenna measurements, (in preparation) to be submitted to J. Atmos. Terrs. Phys., 38, 731-738.

Wright, J. W. (1975), Evidence for precipitation of Energetic Particles by Ionospheric "Heating" transmissions, J. Geophys. Res. 80 4383-4386.

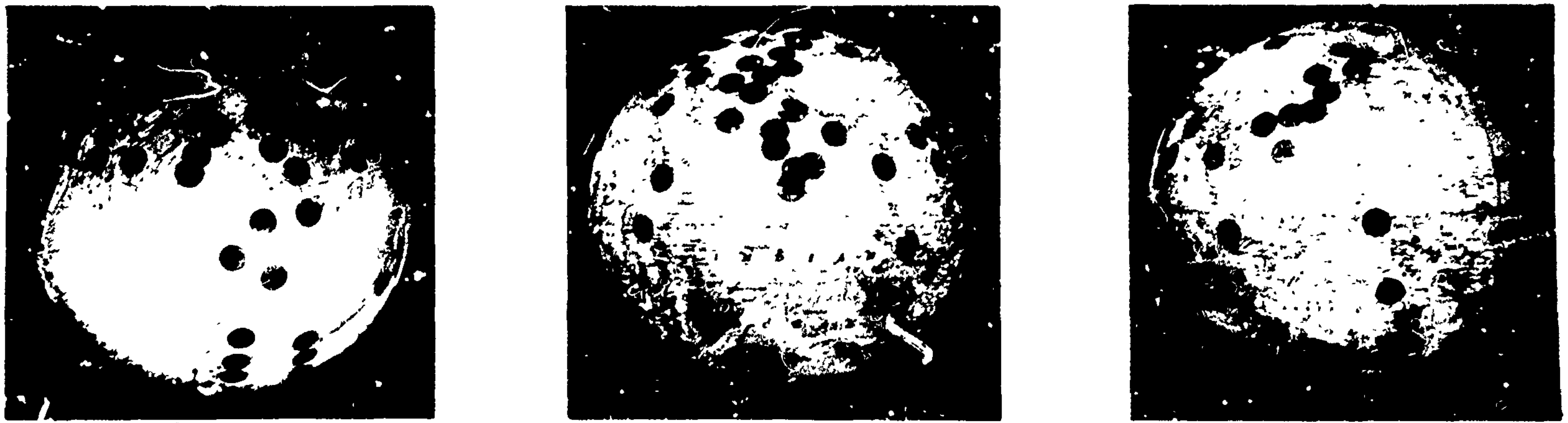
Wright, J. W. (1977), Development of Systems for Remote Sensing of Ionospheric Structure and Dynamics: The Dynasonde Data Acquisition and Dynamic Display System. NOAA Special Report, U.S. Department of Commerce, National Oceanic and Atmospheric Administration Environmental Research Laboratory.



North Pole

American Continents

Africa, Europe



South Pole

India

Asia

Figure 1 The presently active ionosonde network, or about 150 conventional instruments. Stations are listed in Table 2. Station labels are of about 500 km radius, and suggest the "field of view" afforded by usual antenna systems and instrument sensitivity, however, most conventional ionosondes cannot determine echolocation and thus confuse horizontal and vertical structure.

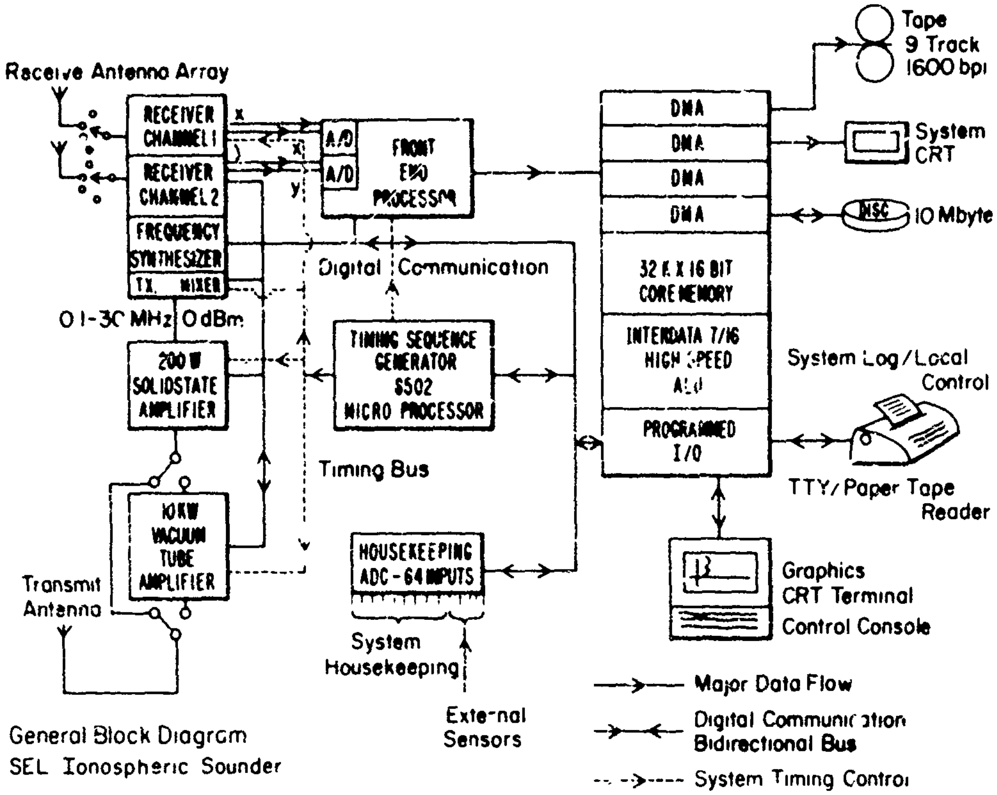

Receive Antenna Array
RECEIVER CHANNEL 1
RECEIVER CHANNEL 2
FREQUENCY SYNTHESIZER
TX MIXER
x
y
A/D
A/D
FRONT END PROCESSOR
0.1-30 MHz
0 dBm
Digital Communication
200 W SOLIDSTATE AMPLIFIER
TIMING SEQUENCE GENERATOR 6502 MICRO PROCESSOR
Timing Bus
10 KW VACUUM TUBE AMPLIFIER
Transmit Antenna
HOUSEKEEPING ADC - 64 INPUTS
System Housekeeping
External Sensors
DMA
DMA
DMA
DMA
32 K X 16 BIT CORE MEMORY
INTERDATA 7/16 HIGH SPEED ALU
PROGRAMMED I/O
Tape 9 Track 1600 bpi
System CRT
DISC 10 Mbyte
System Log / Local Control
TTY / Paper Tape Reader
Graphics CRT Terminal
Control Console
Major Data Flow
Digital Communication Bidirectional Bus
System Timing Control
General Block Diagram
SEL Ionospheric Sounder

Figure 2 Functional and hardware layout of the digital ionosonde currently assembled at NOAA, Space Environment Laboratory. Note the microprocessors dedicated to real-time digital signal processing from two parallel quadrature receivers, and to system timing. Full system control for data acquisition, data processing, and user interaction is achieved through software in the general-purpose minicomputer.

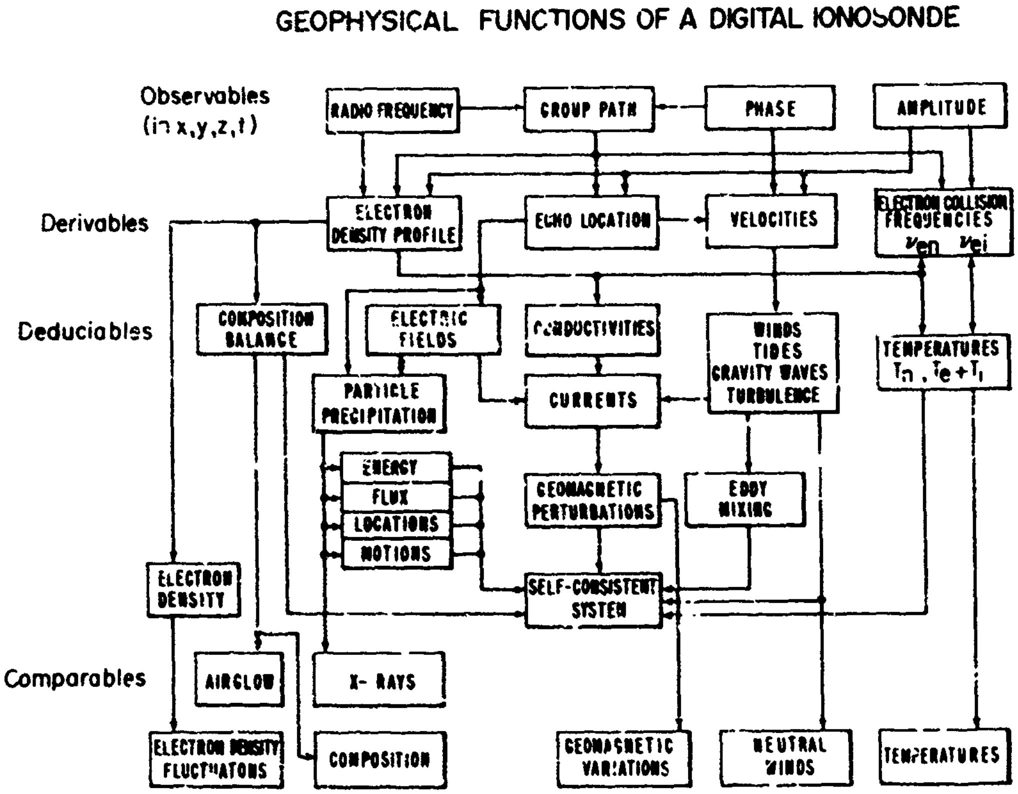

GEOPHYSICAL FUNCTIONS OF A DIGITAL IONOSONDE
Observables (in x,y,z,t)
RADIO FREQUENCY
GROUP PATH
PHASE
AMPLITUDE
Derivables
ELECTRON DENSITY PROFILE
ECHO LOCATION
VELOCITIES
ELECTRON COLLISION FREQUENCIES νen νei
Deducibles
COMPOSITION BALANCE
ELECTRIC FIELDS
CONDUCTIVITIES
WINDS TIDES GRAVITY WAVES TURBULENCE
TEMPERATURES Tn, Te+Ti
PARTICLE PRECIPITATION
CURRENTS
ENERGY
FLUX
LOCATIONS
MOTIONS
GEOMAGNETIC PERTURBATIONS
EDDY MIXING
ELECTRON DENSITY
SELF-CONSISTENT SYSTEM
Comparables
AIRGLOW
X- RAYS
ELECTRON DENSITY FLUCTUATIONS
COMPOSITION
GEOMAGNETIC VARIATIONS
NEUTRAL WINDS
TEMPERATURES

Figure 3 Relationships among the measureable quantities (observables) in ionospheric total-reflection radio sounding, and the geophysical properties of the ionospher and atmosphere. Note the interdependence of the derivable and deducible properties on all of the observables.

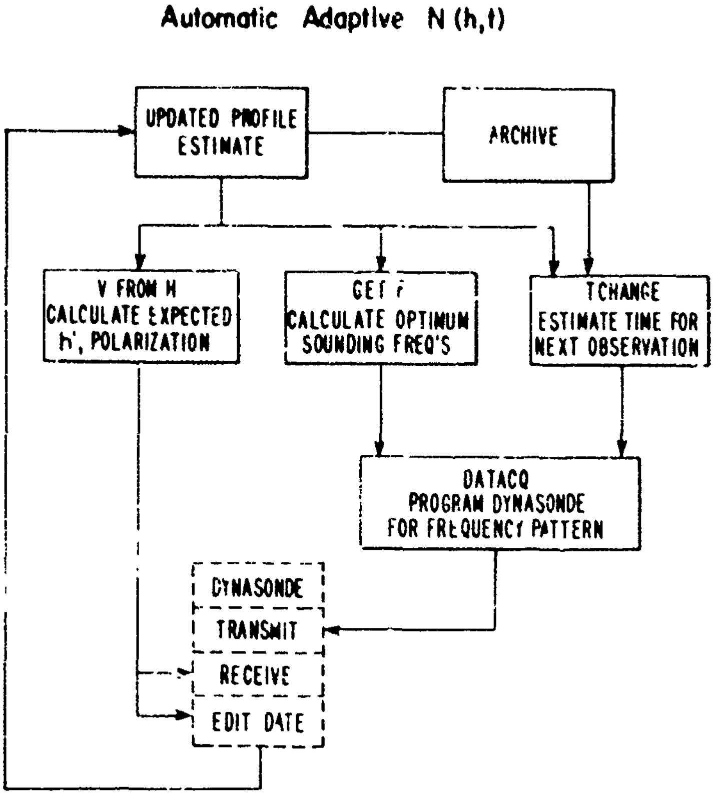

Automatic Adaptive N (h,t)
UPDATED PROFILE ESTIMATE
ARCHIVE
V FROM H CALCULATE EXPECTED h', POLARIZATION
GET f CALCULATE OPTIMUM SOUNDING FREQ'S
TCHANGE ESTIMATE TIME FOR NEXT OBSERVATION
DATACQ PROGRAM DYNASONDE FOR FREQUENCY PATTERN
DYNASONDE
TRANSMIT
RECEIVE
EDIT DATE

Figure 4 Automatic, adaptive N(z,t) inversion. A profile estimate N(z) is used by GETF to decide the optimum sounding frequencies; VFROMH predicts the expected group heights. DATACQ programs the Dynasonde to observe updated R'(f), from which the next profile N(z) is computed. Meanwhile, TCHANGE monitors the time rate of change of profile parameters, to influence the sounding schedule.

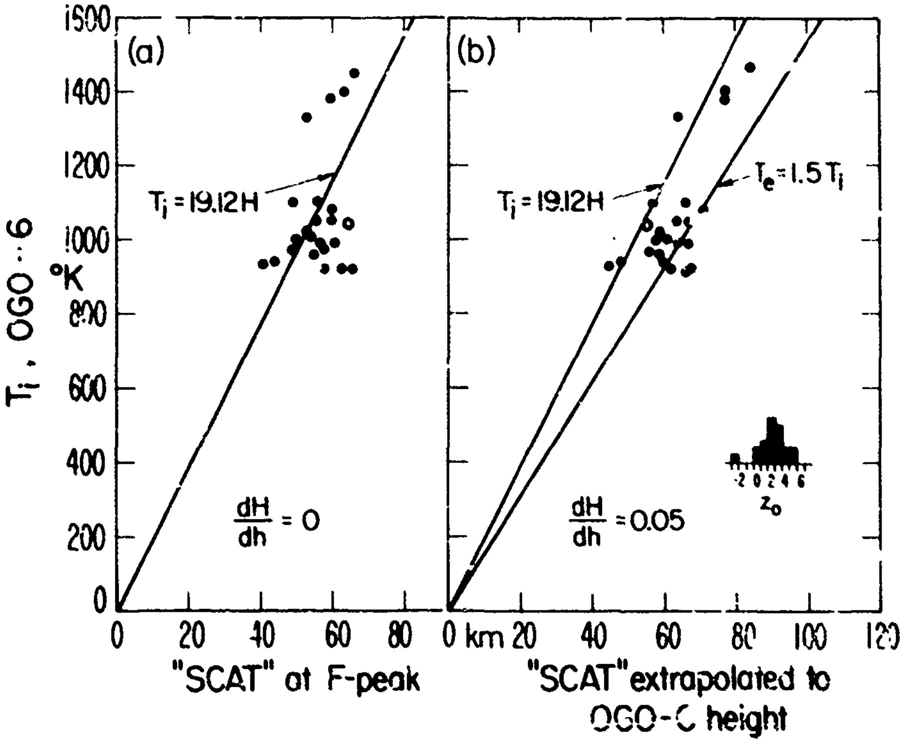


Figure 5 From a small sample of OGO-6 satellite passes over ionosonde stations, measurements of [illegible] temperature at satellite altitude (0-5 scale heights above h max F2, as shown [illegible] the Zo inset) [illegible] compared here with the curvature (expressed as a scale height) at the F2 [illegible] peak from [illegible] N(Z) profiles. Right and left panels show the effects of unequal Te, Ti and [illegible] height gradient.

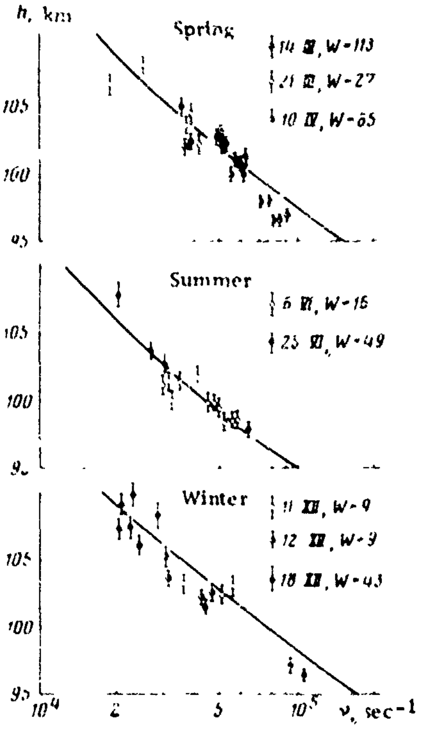
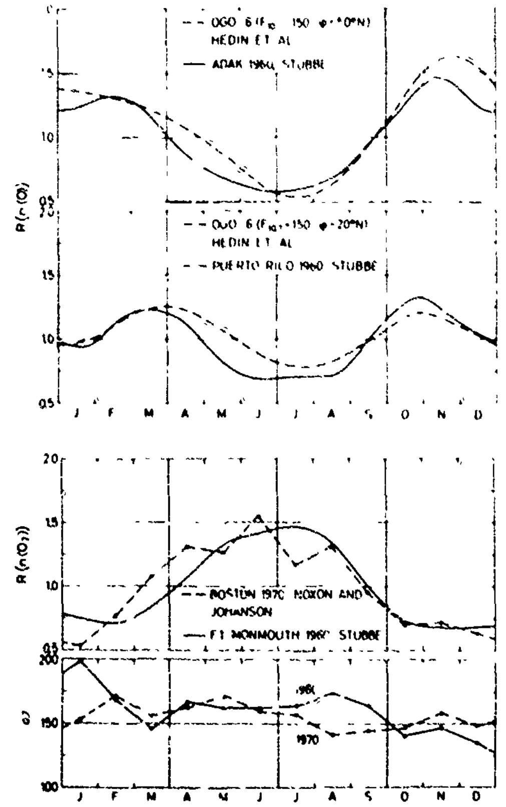


Figure 6 Height variations of electron collision frequency near the E-layer peak deduced from radio wave deviature absorption (plotted points) compared with gas-kinetic calculations using an atmospheric model (from Danilkin, et. al., 1978).



Relative seasonal variations of O density from ionograms (Stubbe) and from OGO-6 (Hedin). Of O_2 density from ionograms (Stubbe) and from 6300A airglow (Noxon and Johanson).

FROM: P. Stubbe, "Atmospheric Parameters from Ionospheric Measurements", Z. *Geophys.*, 1973, 39, 104[illegible]54.

Figure 7 Relative Seasonal Variations of O Density from Ionograms (Stubbe) and from OGO-6 (Hedin). Of O_2 Density from Ionograms (Stubbe) and from 6300A Airglow (Noxon and Johanson).

Coordinated Vertical and Oblique Measurements

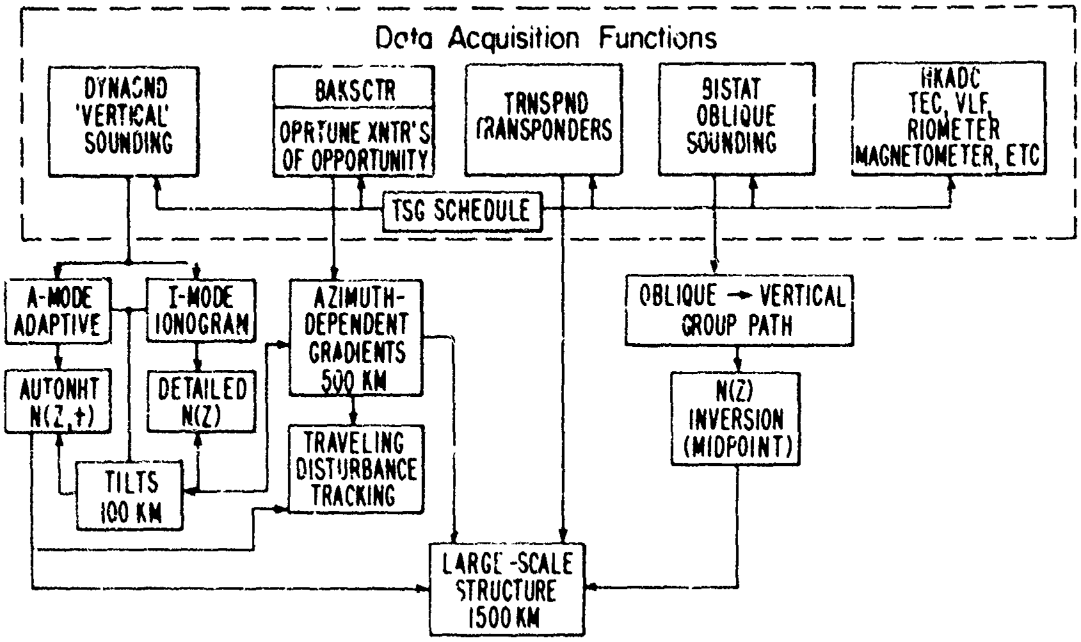

Data Acquisition Functions
DYNASND 'VERTICAL' SOUNDING
BAKSCTR
OPRTUNE XNTR'S OF OPPORTUNITY
TRNSPND TRANSPONDERS
BISTAT OBLIQUE SOUNDING
HKADC TEC, VLF, RIOMETER MAGNETOMETER, ETC
TSG SCHEDULE
A-MODE ADAPTIVE
I-MODE IONOGRAM
AZIMUTH-DEPENDENT GRADIENTS 500 KM
OBLIQUE → VERTICAL GROUP PATH
AUTONHT N(Z,t)
DETAILED N(Z)
TRAVELING DISTURBANCE TRACKING
N(Z) INVERSION (MIDPOINT)
TILTS 100 KM
LARGE-SCALE STRUCTURE 1500 KM

Figure 8 Relationships among the principal data acquisition modes of the Dynasonde, and the data processing methods needed to support and interrelate them.

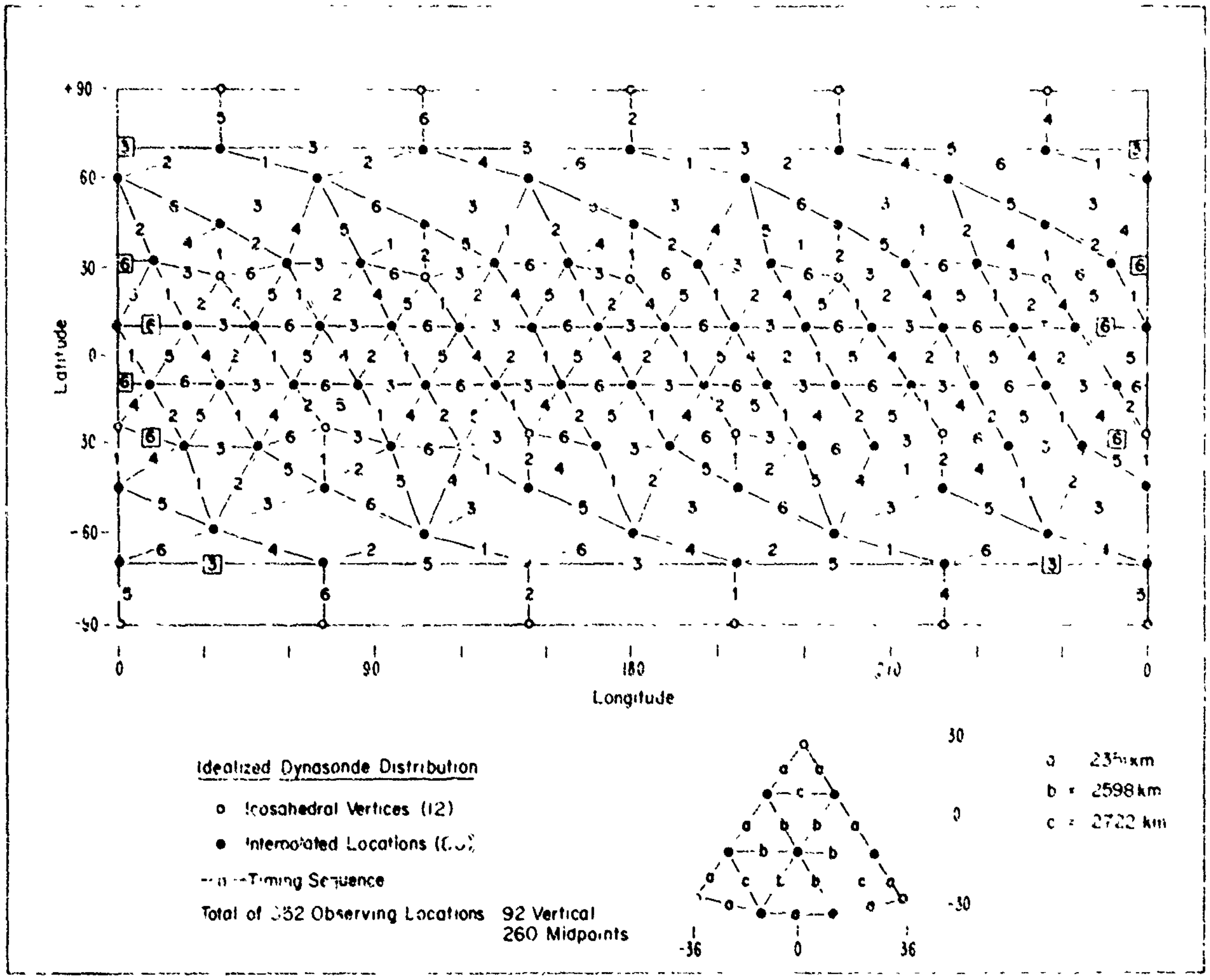

+90
60
30
0
30
-60
-90
Latitude
0
90
180
270
0
Longitude
Idealized Dynasonde Distribution
o Icosahedral Vertices (12)
● Interpolated Locations (80)
-- n -- Timing Sequence
Total of 352 Observing Locations 92 Vertical 260 Midpoints
30
0
-30
-36
0
36
a 23[illegible] km
b = 2598 km
c = 2722 km

Figure 9 An idealized station deployment, based upon the twelve vertices of an icosahedron (o), each of the 20 faces of which are further divided (●) into nine nearly equal isoceles triangles. Small numbers between sites suggest a possible schedule of bistatic soundings among adjacent station pairs in a cycle of six.

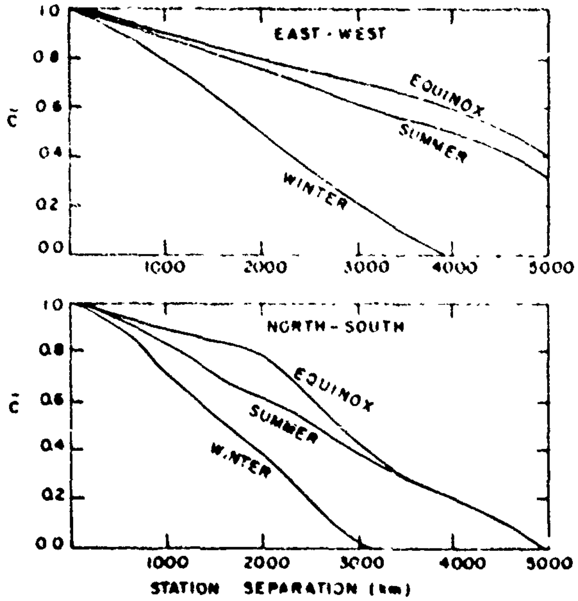


Figure 10 Correlation distance of foF2, noon values, from Rush and Edwards (1976), in longitude (top) and latitude (bottom) for three seasons.

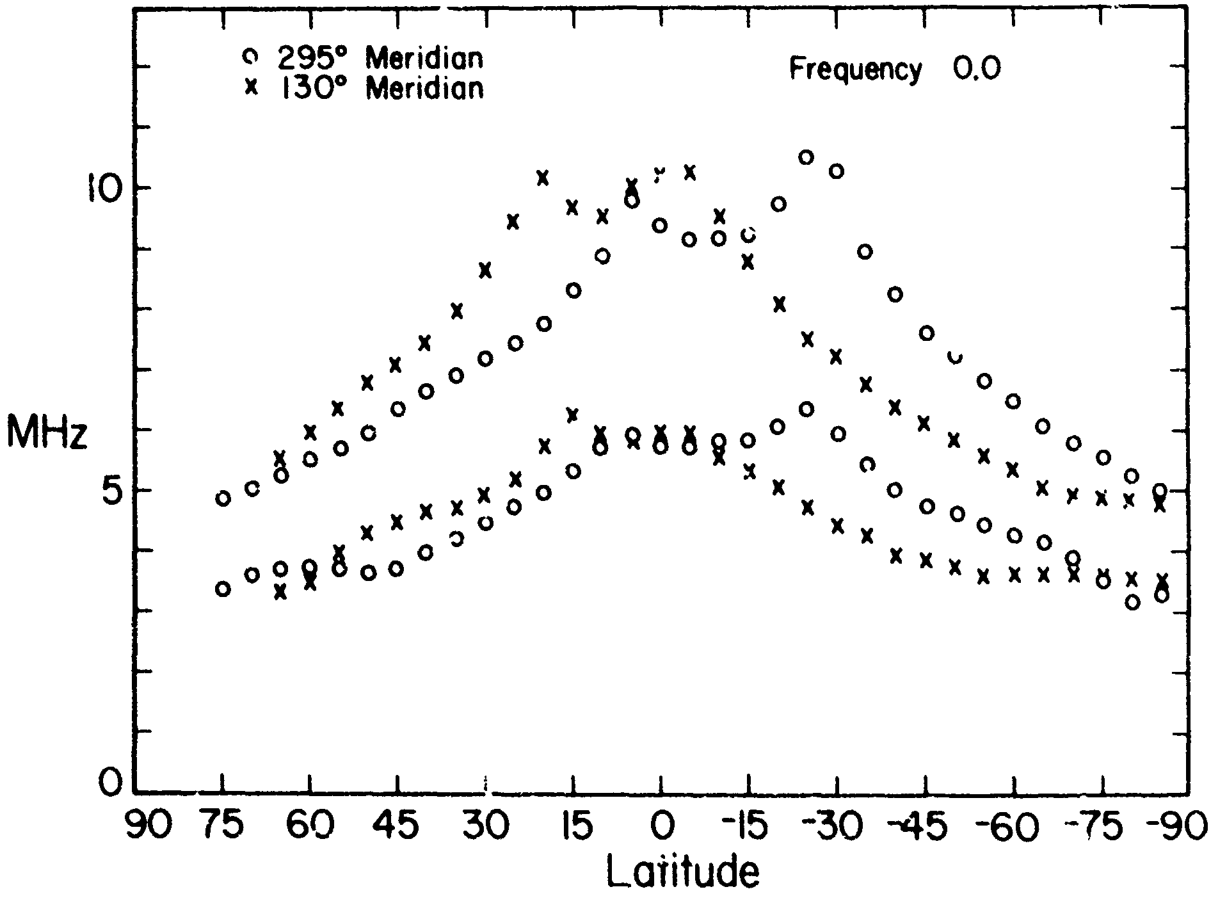


Figure 11 The constant component of foF2 from Anharmonic Frequency Analysis (AFA), versus latitude along two meridians and at two levels of solar activity, variation with latitude of SSN = 0, 100.

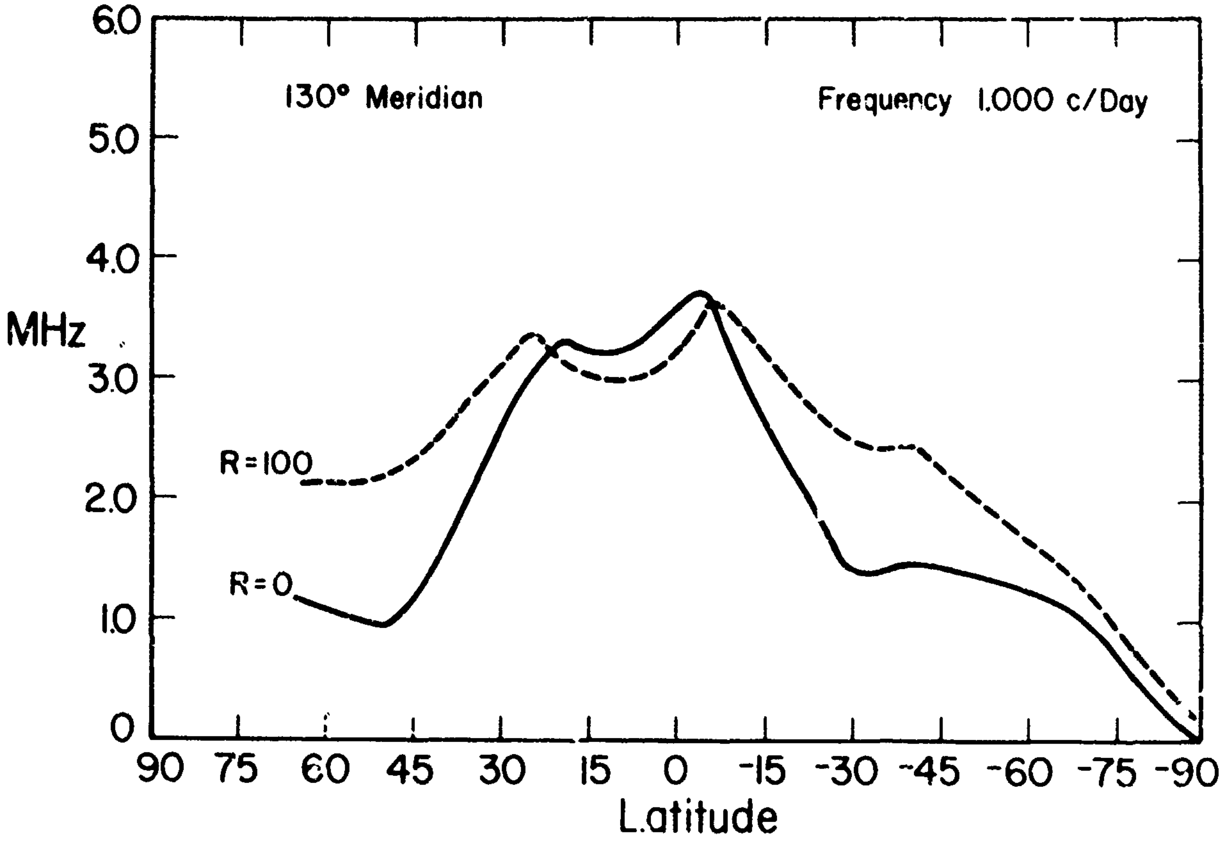


Figure 12 The Diurnal component of foF2 from AFA at two levels of solar activity, along the 130° meridian.

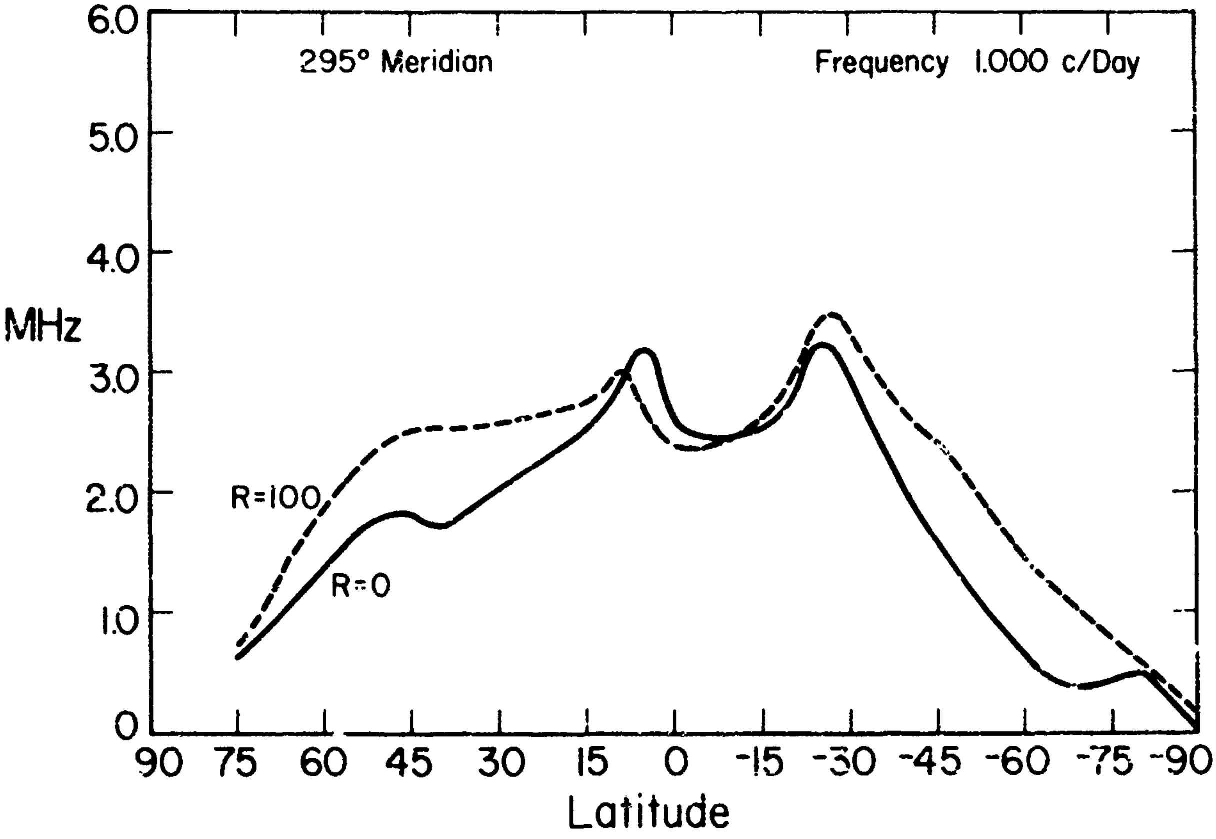
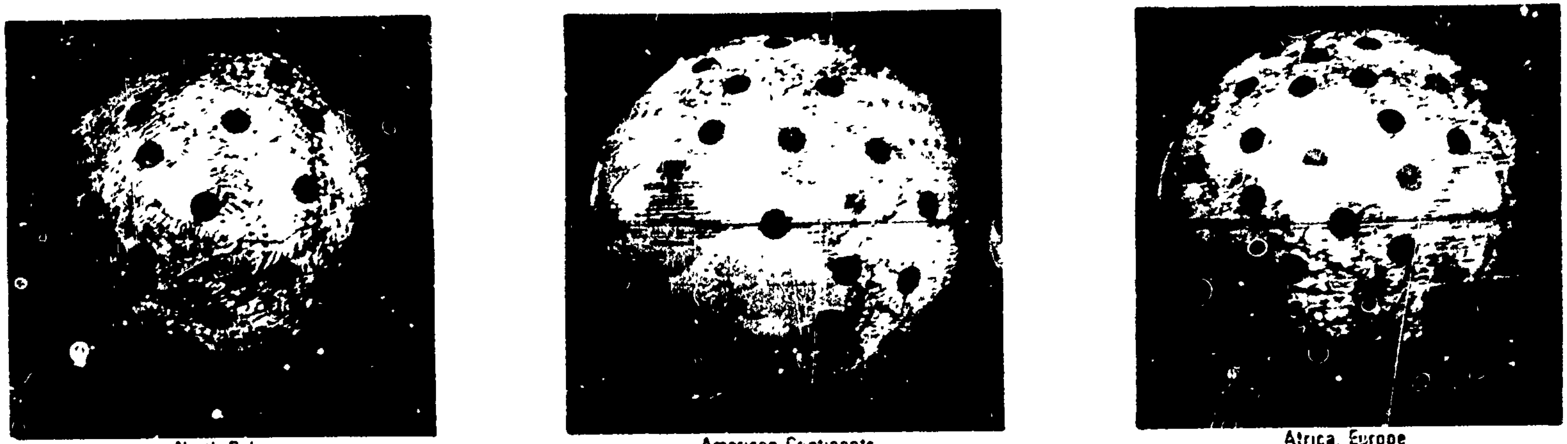


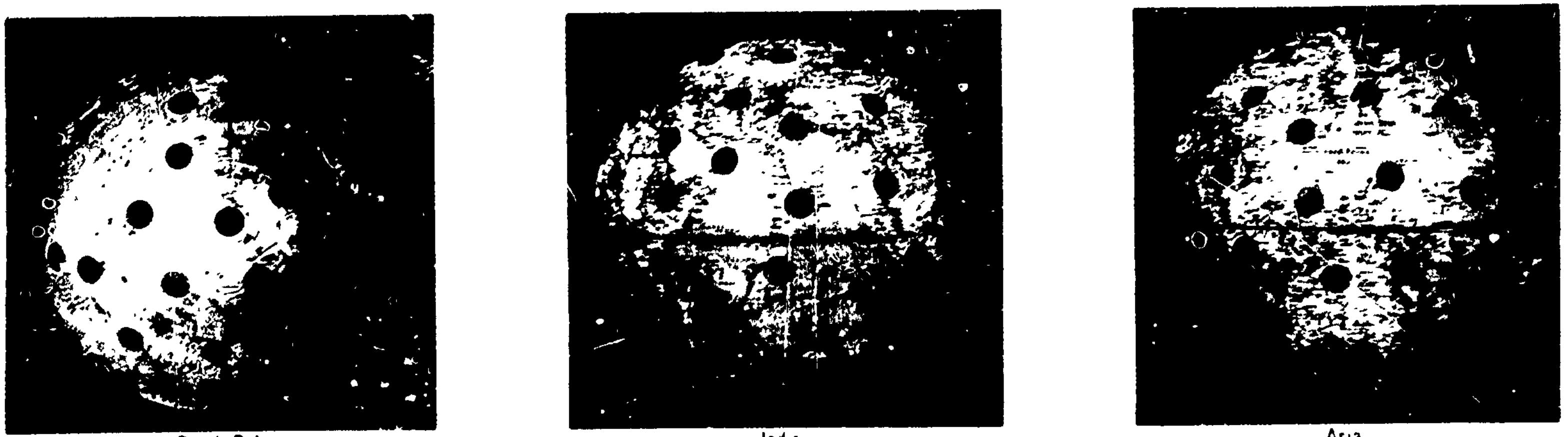
Figure 13 Variation with latitude of the diurnal component of foF2 from AFA at two levels of solar activity, along the 295° meridian.



North Pole

American Continents

Africa, Europe



South Pole

India

Asia

Figure 14 Suggested deployment of an optimum network, of about 90 modern digital ionosondes, also listed in Table 2. Each station would maintain an accurate three dimensional local model, applicable to a vertical cone of about the label radius (500 km) at the F2 peak. Backscatter, bistatic oblique sounding and other methods would maintain observations at about 250 intermediate locations.

HYBRID RAY-MODE FORMULATION OF TROPOSPHERIC PROPAGATION

S.H. Cho, C.G. Migliora and L.B. Felsen
Department of Electrical Engineering
Polytechnic Institute of New York
Farmingdale, New York 11735

SUMMARY

High frequency propagation in an elevated tropospheric duct is analyzed here by a new method involving an appropriate mixture of ray-optical fields and whispering gallery mode fields guided along the concave side of the duct boundary. In this formulation, ray fields may be regarded as expressing the remainder field when a guided mode series is truncated or, alternatively, a certain number of guided modes accounts for omitted higher order reflected rays. To explore the theory on a simple example, the duct is modeled by a single circular boundary separating an interior higher refractive index region from an exterior region with lower index. Source and observation points are assumed to be widely separated, and both are located on the duct boundary. The electric line source Green's function is first derived in terms of eigenmodes involving whispering gallery plus continuous spectrum contributions, and is then converted to the hybrid ray-mode alternative form. It is shown that the number of modes and rays required in the hybrid formulation is far less than when the field is expressed solely either in terms of modes or in terms of rays. A numerical example for typical tropospheric conditions indicates that a few of the lowest-loss modes are adequate to describe the field when source and observation points are on the boundary, rendering the ray contribution negligible. However, ray fields are expected to assume greater importance when the observation point, source point, or both, are located inside the duct.

1. INTRODUCTION AND CONCLUSIONS

In an elevated tropospheric duct, high frequency signals are guided with little attenuation by the outer boundary separating the duct (higher refractive index) medium from the exterior. The wave types responsible for this low-loss propagation mechanism are the whispering gallery (W.G.) modes, which have appreciable field strengths only in thin layers adjacent to the boundary. For an elevated source inside the duct, the field observed at long range may comprise many W.G. modes whose propagation and attenuation coefficients must be calculated very accurately for reliable prediction of the observed phase and amplitude of the transmitted signal. Because of interference between oscillatory modal fields, the total field is very sensitive to the contributions from the various relevant modal constituents.

An alternative procedure for calculating high-frequency fields is in terms of geometric-optical rays. Here, the presence of the concave boundary between the duct and exterior regions implies the existence of ray contributions undergoing many reflections before reaching a distant observation point inside the duct, and especially near the boundary. These multiply reflected ray fields of high order are not only tedious numerically but the notion of a ray-optical field becomes invalid when the ray travels too close to the boundary; in that event, one can no longer identify a local plane wave field (the boundary disturbs the phase front) which forms the foundation for a ray description.

The above observations suggest that a highly efficient and physically appealing method for analyzing high frequency tropospheric propagation would involve a mixture of rays and modes such that the fields extending relatively far from the duct boundary would be expressed in terms of rays while the fields clinging close to the boundary would be expressed in terms of W.G. modes. We have shown that such a hybrid formulation can indeed be achieved when the boundary is perfectly conducting (Ishihara, T. and L.B. Felsen, 1978) or characterized by a surface impedance (Ishihara, T and L.B. Felsen, 1979). The present paper extends this analysis to the tropospheric case where the boundary separates two media with different refractive indeces descriptive of tropospheric conditions. To establish the theory on the simplest model, the duct and exterior media are taken to be homogeneous, the duct boundary circular, and the propagation problem two-dimensional. Generalization to radial inhomogeneities and spherical geometry should pose no difficulties after the two dimensional model has been solved.

The analysis begins with the formulation of the exact electric line source Green's function in the two-medium cylindrical geometry. To include only the guiding effects of the interface between the source point Q and observation point P , without azimuthal periodicity of the field, the problem is posed in an infinitely extended angular space, equivalent to placing a "perfect absorber" at two radial planes. Such an absorber introduces spurious scattering effects from the radial coordinate origin, which are subsequently subtracted during the asymptotic (high-frequency) calculation of the field (Felsen, L.B. and N. Marcuvitz, 1973). The asymptotic Green's function is expressed in alternative forms comprising a) a discrete spectrum of proper eigenmodes plus a proper continuous spectrum, and b) a mixture of geometric optical fields, modal fields and a remainder integral, the contribution from the latter being negligible under suitable conditons. The detailed analysis, and corresponding numerical calculations for a model troposphere, have been performed for the special case when P and Q are widely separated and both are located on the interface. The mode-plus-continuous-spectrum representation has been used as a reference solution, with which the hybrid ray-mode formulation is compared. It is found that relatively few of the most tightly bound whispering gallery modes fully describe the field, rendering the ray contribution negligible. Therefore, this case is not best suited to demonstrating the utility of a ray-mode mixture. It is anticipated that this situation will change when the source point, the observation point, or both, are removed from the interface since the most tightly bound modal fields then have exponentially small amplitudes (Pappert, R.A. and C.L. Goodhart, 1977). Nevertheless, we have truncated the mode series by omitting some of the most tightly bound W.G. modes in order to force the ray contribution to be non-trivial. Under these circumstances, however, the validity of ordinary ray optics is being strained. The results show qualitative agreement that improves as the number of modes is increased and the number of rays reduced.

In summary, although the ray-optical field is unimportant for the present special example when all of the lowest-loss (most tightly bound) W.G. modes are included, the hybrid ray-mode formulation does demonstrate a systematic means for calculating the ducted fields in the most efficient numerical, and in a physically significant, manner. The number of rays and modes included in the representation is determined from well-defined criteria. Thus, the representation can be regarded as using rays to account quantitatively for the remainder field in a truncated mode series, or as using modes to account quantitatively for the remainder field in a truncated ray series, with the number of either being far less than when only one or the other is used exclusively. The method should have strong potential for dealing with lateral and longitudinal duct inhomogeneities, and with scattering by obstacles or localized scattering centers within the duct. These aspects are presently under study.

2. ALTERNATIVE FIELD REPRESENTATION

A. Green's function formulation

The physical model consists of two dielectric media separated by a cylindrical interface with radius $\rho = a$. The refractive indeces in medium $1(\rho < a)$ and in medium $2(\rho > a)$ are n_1 and n_2, respectively, with $n_1 > n_2$ but $n_1 \simeq n_2$. An electric line source is placed on the interface at an angular position $\phi = \phi'$. Since we are interested only in the guiding effects of the interface from the source point Q to an observation point P, it is necessary to remove the angular (2π) periodicity of the field in the cylindrical geometry. To this end, we place along two radial planes (which may be along a single diameter as in Fig. 1) a "perfect absorber for angularly propagating waves" which has the effect of extending the ϕ domain from its original $0 \leq \phi \leq 2\pi$ to $-\infty < \phi < \infty$, (Felsen, L. B. and N. Marcuvitz, 1973). Thus, waves originating at the source travel toward $|\phi| \to \infty$ without reflections, and the ϕ-dependence of the fields is not restrained to be periodic. The ficticious "perfect absorber" is, however, known to have the property of introducing a scattering center at the origin $\rho = 0$, which gives rise to spurious diffraction effects (Felsen, L. B. and N. Marcuvitz, 1973). These spurious contributions must be removed from the desired field representation. Since we are interested only in the lowest order asymptotic solution with respect to the large parameter $k_1 a$, where k_1 is the wavenumber in medium 1, this deletion can be accomplished when performing the asymptotic field calculation.

In the configuration of Fig. 1, the line-source-excited field (Green's function) $G(\rho, \phi; a, \phi')$ can be constructed in the following integral form (Felsen, L. B. and N. Marcuvitz, 1973):

$$G(\rho, \phi; a, \phi') = \frac{1}{2\pi} \int_C \tilde{g}(\rho, a; \nu) \exp(i\nu |\phi - \phi'|)\, d\nu \tag{1}$$

where

$$\tilde{g}(\rho, a; \nu) = \begin{cases} g(\rho, a; \nu) & \text{for } \operatorname{Re}(\nu) \geq 0 \\ g(\rho, a; -\nu) & \text{for } \operatorname{Re}(\nu) \leq 0 \end{cases} \tag{2}$$

and C is the integration path shown in Fig. 2. The radial Green's function $g(\rho, a; \nu)$ is obtained by imposing continuity on the tangential electric and magnetic fields at $\rho = a$, with the source placed at $\rho = a$:

$$g(\rho, a; \nu) = \frac{i\pi}{4} \begin{cases} \dfrac{\psi_1(k_1\rho)\psi_2(k_1 a)}{1 - R(\nu)} & , \rho \leq a \\[2ex] \dfrac{\psi_1(k_1 a)\psi_2(k_2 a)}{1 - R(\nu)} \quad \dfrac{H_\nu^{(1)}(k_2\rho)}{H_\nu^{(1)}(k_2 a)} & , \rho \geq \end{cases} \tag{3}$$

Here,

$$\psi_1(k_1\rho) = 2 J_\nu(k_1\rho) = H_\nu^{(1)}(k_1\rho) + H_\nu^{(2)}(k_1\rho) , \tag{4}$$

$$\psi_2(k_1\rho) = H_\nu^{(1)}(k_1\rho) + R(\nu) H_\nu^{(2)}(k_1\rho) , \tag{5}$$

$$R(\nu) = \frac{H_\nu^{(1)}(k_1 a)}{H_\nu^{(2)}(k_1 a)} \cdot \bar{R}(\nu) , \tag{6}$$

$$\bar{R}(\nu) = \frac{\left[H_\nu^{(1)'}(k_1 a)/H_\nu^{(1)}(k_1 a)\right] - \left[k_2/k_1 H_\nu^{(1)'}(k_2 a)/H_\nu^{(1)}(k_2 a)\right]}{\left[-H_\nu^{(2)'}(k_1 a)/H_\nu^{(2)}(k_1 a)\right] + \left[k_2/k_1 H_\nu^{(1)'}(k_2 a)/H_\nu^{(1)}(k_2 a)\right]} \tag{7}$$

and the wave numbers for the media on the concave and convex sides of the interface, respectively, are $k_i = k_o n_i$ $(i = 1, 2)$, with k_o representing the wave number in free space, and n_i representing the refractive index.

From (3), the resonance equation for angularly propagating modes is

$$1 - R(\nu_q) = 0 \tag{8}$$

or equivalently

$$\frac{J'_{\nu_q}(k_1a)}{J_{\nu_q}(k_1a)} - \frac{k_2}{k_1}\,\frac{H^{(1)'}_{\nu_q}(k_2a)}{H^{(1)}_{\nu_q}(k_2a)} = 0 \tag{9}$$

The resonance equation in (9) has two sets of roots, one set near $J'_\nu(k_1a) = 0$ or $J_\nu(k_1a) = 0$ and the other set near $H^{(1)'}_\nu(k_2a) = 0$ or $H^{(1)}_\nu(k_2a) = 0$ as shown in Fig. 2. The first set of roots represents whispering gallery modes trapped on the concave side of the boundary, while the second set describes fields similar to the creeping waves on a perfectly conducting convex surface. Since k_1a is much larger than unity and k_2 is very close to k_1, (9) can be approximated near $\nu = k_1a$ in terms of Airy functions (Abramowitz and Stegun, 1964):

$$\frac{V'(t_q)}{V(t_q)} - (1-\Delta n)\,\frac{W'_1(t_q + x_D)}{W_1(t_q + x_D)} = 0 \tag{10}$$

where $\Delta n = n_1 - n_2 << 1$ and,

$$\nu_q = k_1a + (k_1a/2)^{1/3}\,t_q \, , \tag{11a}$$

$$x_D = 2\Delta_n\,(k_1a/2)^{2/3} \tag{11b}$$

$$V(t) \equiv Ai(t) \tag{12}$$

$$W_1(t) \equiv Ai(t) - i\,Bi(t) \tag{13}$$

When ν is not near k_1a, one may employ the Debye approximations (see A(4) in Appendix 1) in (9) to obtain:

$$\exp(-i\pi/2)\exp(i2\zeta(w_m))\cdot\frac{\cos w_m - \sqrt{\cos^2 w_m - 2\Delta n}}{\cos w_m + \sqrt{\cos^2 w_m - 2\Delta n}} = 1 \tag{14}$$

where

$$\zeta(w) = k_1a\,[\cos w - (\pi/2 - w)\sin w]\, , \quad \nu = k_1\,a \sin w \, , \tag{15}$$

Equation (14) is valid for those whispering gallery modes where the Debye approximation may be applied (see (A 14) appendix 1). A similar equation can be obtained for the creeping waves, but the contributions from the modes with ν_q far from k_2a are negligible. Thus one may use (10) for all relevant ν_q in the creeping wave contribution.

For the special case when the observation point also lies on the interface (i.e., $\rho = a$), (3) reduces on use of the Wronskian relation for the cylindrical functions to

$$g(a,a;\nu) = \frac{1}{k_1a\left[\dfrac{J'_\nu(k_1a)}{J_\nu(k_1a)} - \dfrac{k_2}{k_1}\,\dfrac{H^{(1)'}_\nu(k_2a)}{H^{(1)}_\nu(k_2a)}\right]} \tag{16}$$

This case shall be considered first.

B. Guided mode and continuous spectrum representation

The integration contour in (1) can be deformed around the singularities of the integrand in the upper half of the ν-plane since the integrand decays at infinity there. Applying Cauchy's theorem, one may write the Green's function in (1) as (Fig. 2):

$$G(a,\phi;a,\phi') = \frac{1}{2\pi}\int_0^{i\infty}[g(a,a;\nu) - g(a,a;-\nu)]\exp(i\nu|\phi-\phi'|)d\nu + \sum_{m=1}^{M}\bar{G}_m + \sum_{\ell=1}^{\infty}\bar{G}_\ell \tag{17}$$

where

$$\bar{G}_q = \frac{i}{k_1a}\,\frac{\exp(i\nu_q|\phi-\phi'|)}{\dfrac{\partial}{\partial\nu}\left[\dfrac{J'_\nu(k_1a)}{J_\nu(k_1a)} - (1-\Delta n)\,\dfrac{H^{(1)'}_\nu(k_2a)}{H^{(1)}_\nu(k_2a)}\right]_{\nu=\nu_q}} \tag{18}$$

With $q = m$ or ℓ in (18), $\bar{G}_m$ and $\bar{G}_\ell$ represent the whispering gallery modes and creeping waves, respectively. When we employ the Airy function approximation near $\nu = k_1 a$, then

$$\bar{G}_q \sim \frac{-i}{2}\left(\frac{2}{k_1 a}\right)^{1/3} \frac{\exp(ik_1 d)\exp\left[i(k_1 a/2)^{1/3} t_q\right]}{\left\{t_q-(1-\Delta n)(t_q+x_D)+(1-\Delta n)\Delta n\left[\frac{W_1'(t_q+x_D)}{W_1(t_q+x_D)}\right]^2\right\}} \tag{19}$$

where d is $a|\phi-\phi'|$.

The branch cut integral in (17) can be written as

$$I = I_{c_o} + I_{s_1} + I_{s_2} \tag{20}$$

where
$$I_{c_o} = \frac{1}{2\pi}\int_{c_o} g_1(a,a;\nu)\exp(i\nu|\phi-\phi'|)\,d\nu \tag{21}$$

$$I_{s_1} = \frac{1}{2\pi}\int_0^{i\infty}[g_2(a,a;\nu)-g(a,a;-\nu)]\exp(i\nu|\phi-\phi'|)\,d\nu \tag{22}$$

$$I_{s_2} = \frac{-1}{2\pi}\int_{-i\infty}^{0} g_1(a,a;\nu)\exp(i\nu|\phi-\phi'|)\,d\nu \tag{23}$$

$$g(a,a;\nu) = g_1(a,a;\nu)+g_2(a,a;\nu) \tag{24}$$

$$g_1(a,a;\nu) = \frac{i\pi}{4}\psi_2^2(k_1a)/[1-R(\nu)] \tag{25a}$$

$$g_2(a,a;\nu) = \frac{i\pi}{4}H_\nu^{(2)}(k_1a)\,\psi_2(k_1a) \tag{25b}$$

and the integration path C_o lies along the imaginary axis in the ν-plane (Fig. 3). It can be shown (Appendix I) that I_{s_1} and I_{s_2} are $O(\frac{1}{k_1a})$ and represent spurious diffraction effects associated with the origin ($\rho = 0$). These integrals may therefore be neglected. Thus, the continuous spectrum is expressed as $I \sim I_{c_o}$ in (20) and I_{c_o} can be simplified as follows (see Appendix I):

$$I_{c_o} \sim \frac{1}{\pi}\int_0^\infty \frac{\exp[-k_1a|\phi-\phi'|\sinh v]\,dv}{2\Delta n/\cosh^2 v - i4\exp[-2ik_1a(\cosh v - v\sinh v)]\exp[-\pi k_1 a\sinh v]} \tag{26}$$

This integral is in a convenient form for numerical calculation. Thus, the Green's function in (17) becomes

$$G(a,\phi;a,\phi') = \sum_{m=1}^{M}\bar{G}_m + \sum_{\ell=1}^{\infty}\bar{G}_\ell + I_{c_o} \tag{27}$$

with $\bar{G}_m$ or G_ℓ given in (18) and I_{c_o} given in (26). For $\bar{G}_\ell$, and for $\bar{G}_m$ with $\nu_m \simeq k_1 a$ or $k_2 a$, one may employ the simplified expression in (19).

C. Hybrid ray-mode representation

To obtain a hybrid representation containing a mixture of ray and whispering gallery mode fields, we return to I_{c_o} in (21), deform the integration contour to the right across some of the whispering gallery poles, perform a partial ray expansion of the integrand over the deformed path, and then provide an estimate for the remainder integral. A typical deformed contour $\bar{C}$ is shown in Fig. 3. It can be shown that the integrand of I_{c_o} behaves properly as $|\nu| \to \infty$ to legitimize the deformation. We now express $g_1(a,a;\nu)$ in the integrand in (21) as

$$g_1(a,a;\nu) = \frac{\psi_2^2(k_1a)}{1-R(\nu)} = \psi_2^2(k_1a)\left[1+R(\nu)+\dots+R^n(\nu)+\frac{R^{n+1}(\nu)}{1-R(\nu)}\right] \tag{28}$$

where $\psi_2(k_1a)$ and $R(\nu)$ are given in (5) and (6), respectively. Employing Debye approximations and choosing $\bar{C} = C_N$ as defined below, the integral becomes (see Fig. 4)

$$\bar{I} \sim \frac{1}{4\pi}\sum_{n=0}^{N}\int_{C_N}[1+\bar{R}(w)]^2\,\bar{R}^n(w)\exp[iq_n(w)](-i)^n\,dw + R_{N\bar{M}} \tag{29}$$

where

$$q_n(w) = k_1a\left\{|\phi-\phi'|\sin w + 2(n+1)\left[\cos w - \left(\frac{\pi}{2}-w\right)\sin w\right]\right\} \tag{30}$$

$$R_{N\bar{M}} = \frac{1}{4\pi} \int_{\bar{C}_N} \frac{[1+\bar{R}(w)]^2 \bar{R}^{N+1}(w) \exp[iq_N(w)] \exp[i2\zeta(w)](-i)^{N+1}}{1-\bar{R}(w)\exp[i2\zeta(w)](-i)} dw \tag{31a}$$

$$\zeta(w) = k_1 a\,[\cos w - (\pi/2-w)\sin w] \tag{31b}$$

and

$$\nu = k_1 a \sin w \tag{31c}$$

The integral in (29) can be evaluated by the saddle point method (see Fig. 4). The saddle points w_n are determined by $dq_n(w)/=0^1$, and the integration path $\bar{C}_N$ is deformed into $\bar{C}_n$ (local steepest descent path). It may be shown that this deformation is possible, and the saddle point evaluation then yields:

$$\bar{I} = \sum_{n=0}^{N} G_n + R_{N\bar{M}} \tag{32a}$$

where

$$G_n = \frac{e^{i\pi/4}}{4} \sqrt{\frac{2}{\pi k_1 D_n}}\,(-i)^n [1+\bar{R}(w_n)]^2 \bar{R}^n(w_n) \exp(i k_1 D_n) \tag{32b}$$

with

$$\bar{R}(w) = \frac{\cos w_n - \sqrt{\cos^2 w_n - 2\Delta n}}{\cos w_n + \sqrt{\cos^2 w_n - 2\Delta n}} \tag{33a}$$

$$w_n = \pi/2 - |\phi - \phi'|/2(n+1) \tag{33b}$$

and

$$D_n = 2(n+1)\,a \sin[\,|\phi - \phi'|/2(n+1)] \tag{33c}$$

The expression for G_n in (32) represents a geometric-optical ray field reflected n times from the concave side of the boundary on its travel from the source point to the observation point (see Fig. 5). At each reflection, the field amplitude changes by the reflection coefficient $\bar{R}(w_n)$. The remainder integral $R_{N\bar{M}}$ in (30) is taken over the local steepest descent path $\bar{C}_N$ corresponding to the last ray ($n = N$) included in the expansion. The subscript $N\bar{M}$ signifies that $\bar{M}$ whispering gallery modes lie to the right of w_N. The residues of those modes with $0 < \mathrm{Re}\, w_m < \mathrm{Re}\, w_N$ must be extracted from the sum in (27) since they were eliminated by deformation of C_o into $\bar{C}_N$. Thus,

$$G(a,\phi; a,\phi') = \sum_{n=0}^{N} G_n + \sum_{m=1}^{\bar{M}} \bar{G}_m + \sum_{\ell=1}^{L_N} \bar{G}_\ell + R_{N\bar{M}} \tag{34}$$

The upper limit of L_N instead of infinity on the creeping wave sum indicates that the path $\bar{C}_N$ may intercept the pole sequence w_ℓ as shown in Fig. 4. Actually, this is of little practical consequence since the contributions from the higher order creeping waves are very small. It is shown in Appendix 2 that

$$R_{N\bar{M}} \sim \frac{-|\bar{R}(w_N)|}{1+|\bar{R}(w_N)|} G_N \tag{35}$$

provided that the saddle point w_N does not lie near poles w_m with $m = \bar{M}$ or $\bar{M}+1$.

3. NUMERICAL RESULTS

Numerical calculations were performed for a model troposphere with the following parameters: f = 900 MHz a = 6369 km, Δn = 30 units, for variable range $d \equiv a|\phi - \phi'|$. The guided-mode-plus-continuous-spectrum representation in (27) was used as a reference solution. It was found that 30 whispering gallery modes (i.e., M = 30) is adequate to yield accurate field values when the earth's boundary is assumed to lie at ρ = 6363 km. In fact, the maximum number of whispering gallery modes without the earth's surface is about $k_1 a/\pi$ ($\sim 3\times10^7$), and the decay factor does not increase very much as the number of modes increases. With the presence of the earth's surface, it is found that the decay factor becomes large for the higher order modes lying beyond the 40th mode (Cho, S. H. and J. R. Wait, 1978). Therefore, the number of modes providing the major contribution to the total field ranges from 30 to 40. For large distances d, as considered here, the contributions from the creeping waves and the continuous spectrum are negligible.

The hybrid ray-mode calculation from (34) was performed for various combinations of rays and whispering gallery modes, but the number of modes $\bar{M}$ always included all of the modes having very small decay factors (these have poles lying between $\pi/2-\sqrt{2\Delta n}$ and $\pi/2$). With the physical parameters chosen above, the number of these modes is six. The number of rays N was chosen such that $w_N < \pi/2 - \sqrt{2\Delta n}$. The physical interpretation of this condition is similar to that for the perfectly conducting boundary, and is schematized in Fig. 6. Also, the creeping waves are neglected in the hybrid representation, and the rays blocked by the earth's surface are eliminated in the calculation of the total field in order to compare the result with the field based on the whispering gallery modes only, in the presence of the earth. The rays eliminated (i.e., blocked by the earth's surface) are indicated in Fig. 6(b). In Fig. 7(a), the amplitude of the total field is normalized to the field in the free space, with the distance between source and observation points measured along the interface.

The numerical results show that the modes having very small decay factors account completely for the total field, thereby rendering the ray-optical contribution and the remainder $R_{\bar{N}\bar{M}}$ negligible. Thus, the hybrid ray-optical field formulation does not really show up to its best advantage for the present example where the source and observation points are both on the interface since a few tightly bound low-loss modes account efficiently for the observed field at long ranges. Nevertheless, the validity of the method is confirmed because the mode sum may legitimately be terminated after inclusion of only the lowest-loss modes, with the truncation error quantified by the (negligible) ray-optical field.

To force a situation where the geometric-optical field is not negligible, it is necessary to choose $\bar{M}$ so that not all low-loss modes are included. Accordingly, we have calculated the field also for $\bar{M} = 3, 4$. The results are shown in Fig. 7(b). The agreement with the reference solution is now poorer than before but the correct general trend is preserved. Since now $w_{\bar{N}} > \pi/2 - \sqrt{2\Delta n}$, the ray reflection coefficient cannot be approximated by Debye formulas but requires use of the Airy function:

$$\bar{R}(t) \cong \frac{\dfrac{W_1'(t)}{W_1(t)} - \dfrac{W_1'(t+x_D)}{W_1(t+x_D)}}{\dfrac{-W_2'(t)}{W_2(t)} + \dfrac{W_1'(t+x_D)}{W_1(t+x_D)}} \tag{36}$$

with

$$t = (k_1 a/2)^{2/3} \bar{w}_n^2 \quad , \bar{w} = \pi/2 - w \tag{36a}$$

It should be emphasized that the validity of the ray-optical formulation is strained in this parameter range since the rays strike the interface almost at grazing incidence.

When the source and (or) observation points are moved from the interface into the region $\rho < a$, the influence of the tightly bound whispering gallery modes with $\pi/2 - \sqrt{2\Delta n} < w_m < \pi/2$ is de-emphasized since their field decays rapidly away from the boundary (Pappert, R.A. and C. L. Goodhart, 1977). It is then expected that modes with $w_m < \pi/2 - \sqrt{2\Delta n}$ become important, thereby providing a better framework for the hybrid representation. This aspect is now being studied.

APPENDIX 1 - Approximation of continuous spectrum integral.

From (3) and the relation for the Hankel functions,

$$H_{-\nu}^{\overset{(1)}{(2)}}(k_1 a) = \exp(\pm i \nu \pi) H_{\nu}^{\overset{(1)}{(2)}}(k_1 a) \tag{A1}$$

one obtains

$$g(a, a, -\nu) = \frac{i\pi}{4} \frac{[\exp(i2\nu\pi) H_\nu^{(1)}(k_1 a) + H_\nu^{(2)}(k_1 a)]\psi_2(k_1 a)}{1 - \exp(i2\nu\pi) R(\nu)} \tag{A2}$$

Substituting (A2) into (22),

$$I_{s_1} = \frac{-i}{8} \int_0^{i\infty} \frac{\exp(i2\nu\pi) H_\nu^{(1)^2}(k_1 a)[1+\bar{R}(\nu)]^2}{1 - \exp(i2\nu\pi) \dfrac{H_\nu^{(1)}(k_1 a)}{H_\nu^{(2)}(k_1 a)} \bar{R}(\nu)} \exp(i\nu |\phi - \phi'|)\, d\nu \tag{A3}$$

One may employ Debye approximations for $|\nu - k_1 a| > (\frac{k_1 a}{2})^{1/3}$

$$H_\nu^{\overset{(1)}{(2)}}(ka) \sim \sqrt{\frac{2}{\pi k a \cos w}} \exp[\pm i \zeta(w) \mp i \pi/4] \tag{A4}$$

$$H_\nu^{\overset{(1)\prime}{(2)}}(ka) \sim \sqrt{\frac{2\cos w}{\pi k a}} \exp[\pm i \zeta(w) \pm i\pi/4]$$

where

$$\zeta(w) = ka[\cos w - (\pi/2 - w) \sin w] \tag{A4a}$$

and (A4) is valid for $|\arg(\zeta(w)^{2/3})| < 2\pi/3$.

Let

$$\nu = k_1 a \sin w \tag{A5a}$$

and

$$d\nu = k_1 a \cos w\, dw = i k_1 a \cosh v\, dv \tag{A5b}$$

where the last equality is valid on the imaginary axis in the ν-plane (Abramowitz and I. Stegun, 1964). Using the Debye approximation in (A4), with (A5a, b), in (A3), one may obtain:

$$I_{s_1} \sim -\frac{i}{4\pi}\int_0^{\infty} \frac{[1+\overline{R}(v)]^2 \exp[-k_1a(\pi+|\phi-\phi'|)\sinh v]\exp[i2k_1a(\cosh v - v\sinh v)]}{1-(-i)\exp[i2k_1a(\cosh v - v\sinh v)]\exp(-\pi k_1a\sinh v)}\,dv$$

$$\sim -\frac{i}{4\pi}[1+\overline{R}(0)]^2\left[\frac{\exp(i2k_1a)}{(\pi+|\phi-\phi'|)k_1a} + (-i)\overline{R}(o)\frac{\exp(i4k_1a)}{(2\pi+|\phi-\phi'|)k_1a} + \text{-----}\right] + O\left(\frac{1}{k_1^2a^2}\right) \quad \text{(A6)}$$

where

$$\overline{R}(0) = \frac{1-\sqrt{1-2\Delta n}}{1+\sqrt{1-2\Delta n}}$$

The asymptotic result in (A6) follows on geometric series expansion of the integrand and evaluation of each integral in the series by integration by parts. Similarly,

$$I_{s_2} = -\frac{i}{8}\int_0^{i\infty} \frac{\exp(i2\nu\pi)[H_\nu^{(1)}(k_1a)]^2[1+\overline{R}(\nu)]^2\exp(-i\nu|\phi-\phi'|)}{1-\overline{R}(\nu)\exp(i2\nu\pi)H_\nu^{(1)}(k_1a)/H_\nu^{(2)}(k_1a)}\,d\nu$$

$$\sim \frac{i}{4}[1+\overline{R}(o)]^2\left[\frac{\exp(i2k_1a)}{(\pi-|\phi-\phi'|)k_1a} + (-i)\overline{R}(o)\frac{\exp(+i4k_1a)}{(2\pi-|\phi-\phi'|)k_1a} \;---\right] + O\left(\frac{1}{k_1^2a^2}\right) \quad \text{(A7)}$$

To deal with the integral I_{c_o} in (21), we write:

$$I_{c_0} = \frac{-i}{(\pi k_1a)^2}\int_{-i\infty}^{i\infty} \frac{\exp(i\nu|\phi-\phi'|)\,d\nu}{[J_\nu'(k_1a) - J_\nu(k_1a)Z(\nu)][H_\nu^{(2)'}(k_1a) - Z(\nu)H_\nu^{(2)}(k_1a)]} \quad \text{(A8)}$$

where

$$Z(\nu) = \frac{k_2}{k_1}\,\frac{H_\nu^{(1)'}(k_2a)}{H_\nu^{(1)}(k_2a)} \quad \text{(A9)}$$

The denominator in (A8) may be rewritten as:

$$D = \frac{1}{2}\Big\{ H_\nu^{(1)'}(k_1a)H_\nu^{(2)'}(k_1a) - Z(\nu)[H_\nu^{(1)}(k_1a)H_\nu^{(2)'}(k_1a) + H_\nu^{(1)'}(k_1a)H_\nu^{(2)}(k_1a)]$$

$$+Z^2(\nu)H_\nu^{(2)}(k_1a)H_\nu^{(1)}(k_1a) + [H_\nu^{(2)'}(k_1a) - Z(\nu)H_\nu^{(2)}(k_1a)]^2\Big\} \quad \text{(A10)}$$

Using Debye approximations from (A4),

$$D \sim \frac{1}{2}\,\frac{2\cos w}{\pi k_1a}\left[\left(1+\frac{Z^2(w)}{\cos w}\right) + \exp(-i\pi/2)\left(1-\frac{Z(w)}{\cos w}\exp(i\pi/2)\right)^2\exp(-i2\zeta(w))\right] \quad \text{(A11)}$$

where

$$Z(w) = \frac{k_2}{k_1}\exp(i\pi/2)\cos w_2 \quad \text{(A12)}$$

with

$$\cos w_2 \sim \sqrt{1-\left(\frac{k_1}{k_2}\sin w\right)^2} \quad \text{(A12a)}$$

Thus, on the imaginary axis, for $v \geq 0$,

$$D \sim \frac{1}{2}\,\frac{2\cosh v}{\pi k_1a}\left[\frac{2\Delta n}{\cosh^2 v} + \exp(-i\pi/2)\left(1+\sqrt{1-\frac{2\Delta n}{\cosh^2 v}}\right)^2 \exp(-2i\zeta(v))\right] \quad \text{(A13)}$$

where $\zeta(v)$ is given by (31b) with $w = iv$. It follows that

$$I_{c_0} \sim \frac{1}{\pi}\int_0^{\infty} \frac{\exp(-k_1a|\phi-\phi'|\sinh v)\,dv}{\left[\frac{2\Delta n}{\cosh^2 v} + 4(-i)\exp(-\pi k_1a\sinh v)\exp[-2ik_1a(\cosh v - v\sinh v)]\right]} \quad \text{(A14)}$$

Here, the contribution from $-i\infty$ to 0 in (A8) is neglected since it is $O\left(\frac{1}{k_1a}\right)$ (see A(6)).

APPENDIX 2 - Approximation of $R_{N\bar{M}}$

From (30),

$$R_{N\bar{M}} = \frac{1}{4\pi} \int_{C_N} \frac{[1+\bar{R}(w)]^2 [\bar{R}(w)]^{N+1} (-i)^{N+1} \exp(i q_N(w)) \exp[i2\zeta(w)]}{1 - \bar{R}(w) \exp[i2\zeta(w)](-i)} \tag{A15}$$

We now rewrite

$$\frac{(-i)\exp[i2\zeta(w)]\,\bar{R}(w)}{1-(-i)\exp[i2\zeta(w)]\,\bar{R}(w)} = -\frac{1}{2}\left[1- \frac{1+(-i)\exp[2i\zeta(w)]\bar{R}(w)}{1+(-i)\exp[2i\zeta(w)]\bar{R}(w)}\right] \tag{A16}$$

$$= -\frac{1}{2}\left[1+i\tan[(\zeta(w)+\frac{\theta(w)}{2}+\pi/4]-i\frac{1}{2}\ \ell n\ |\bar{R}(w)|)\right] \tag{A17}$$

where

$$\bar{R}(w) = \exp[i\,\theta(w)]\exp[\ \ell n\ |\bar{F}(w)|] \tag{A18}$$

If the saddle point w_N for the last ray satisfies

$$\zeta(w_N) + \frac{\theta(w_N)}{2} + \pi/4 = j\pi \tag{A19}$$

where j is an integer, then the remainder term becomes approximately (see (32))

$$R_{N\bar{M}} \sim -\frac{1}{2}G_N\left[1-\frac{1-\bar{R}(w_N)}{1+\bar{R}(w_N)}\right] = -G_N\frac{|\bar{R}(w_N)|}{1+|\bar{R}(w_N)|} \tag{A20}$$

This result is used in (35).

---

[1] Note that the phase function $q_n(w)$ does not include the phase of the reflection coefficient terms that appear as amplitudes. Therefore, the ray paths do not have lateral shifts on the boundary. For the case of interest here, the shift effects are negligible when the Debye approximation can be applied.

REFERENCES

Abramowitz, M. and I. Stegun, 1964, "Handbook of Mathematical Functions," U S. Government Printing Office.

Cho, S. H. and J. R. Wait, 1978, "Analysis of Microwave Ducting in an Inhomogeneous Troposphere," Pure and Applied Geophysics, Vol. 116, Birkhäuser Verlag, Basel, Switzerland.

Felsen, L. B. and N. Marcuvitz, 1973, "Radiation and Scattering of Waves," Chap. 6, Prentice Hall, Englewood Cliffs, New Jersey.

Ishihara, T. and L. B. Felsen, 1978, "High Frequency Fields Excited by a Line Source on a Perfectly Conducting Concave Cylindrical Boundary," IEEE Trans. on Antennas and Propagation, Vol. AP-26 pp. 757-767.

Ishihara, T. and L. B. Felsen, 1979, "High Frequency Fields Excited by a Line Source Located on a Concave Cylindrical Impedance Surface," to be published in IEEE Trans. on Antennas and Propagation.

Pappert, R. A. and C. L. Goodhart, 1977, "Case Studies of Beyond the Horizon Propagation in Tropospheric Duct Environments," Radio Science, Vol. 12, No. 1, pp. 75-88

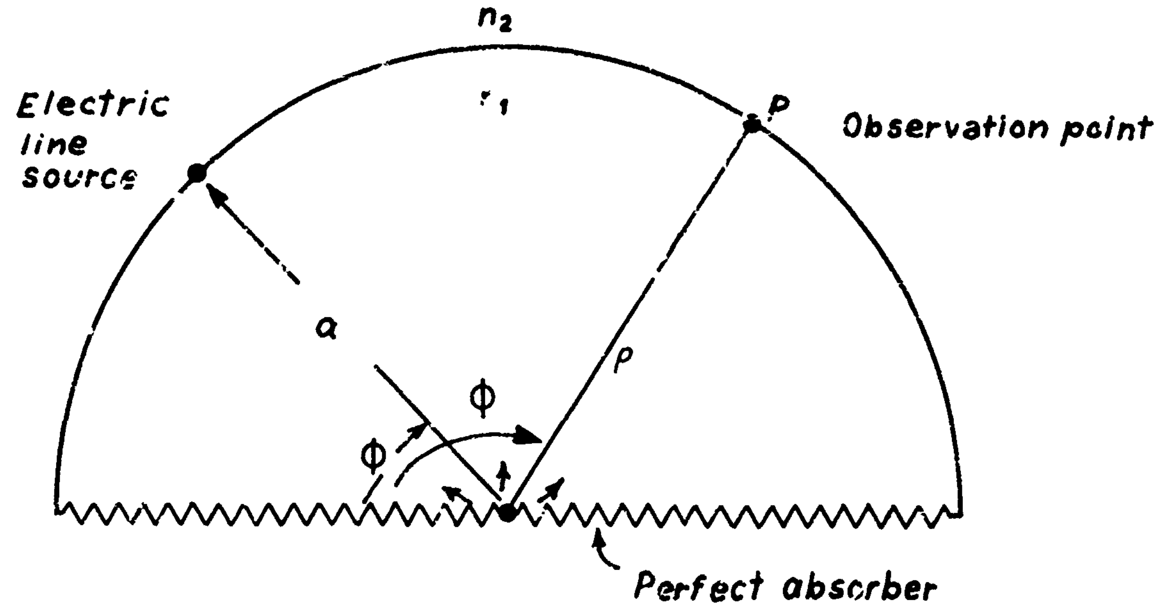

n_2
n_1
Electric line source
P
Observation point
a
ρ
Φ
Φ
Perfect absorber

Fig. 1 Physical configuration. The "perfect angular absorber" introduces a scattering center at $\rho = 0$

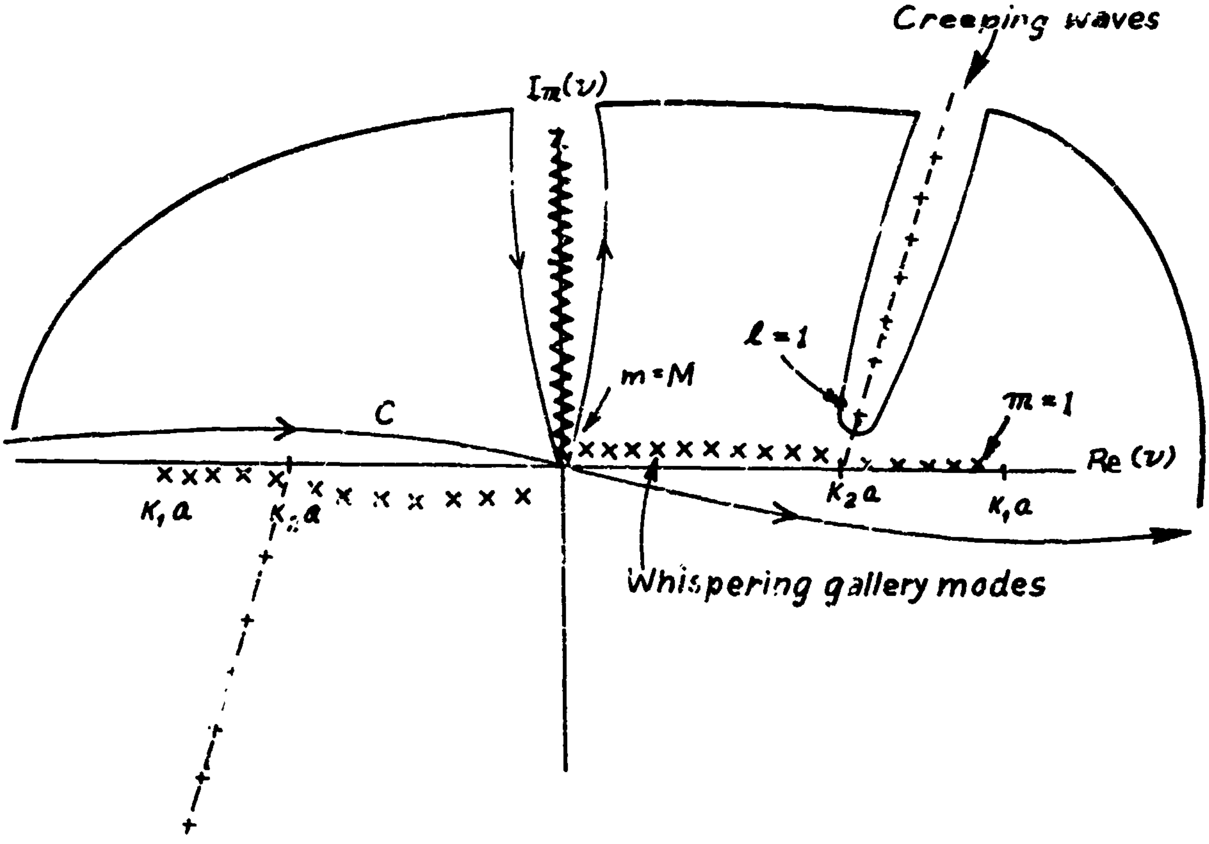

Creeping waves
Im(ν)
ℓ = 1
m = M
m = 1
C
Re (ν)
K_1a
K_2a
K_2a
K_1a
Whispering gallery modes

Fig. 2 Integration paths and singularities in complex ν-plane.

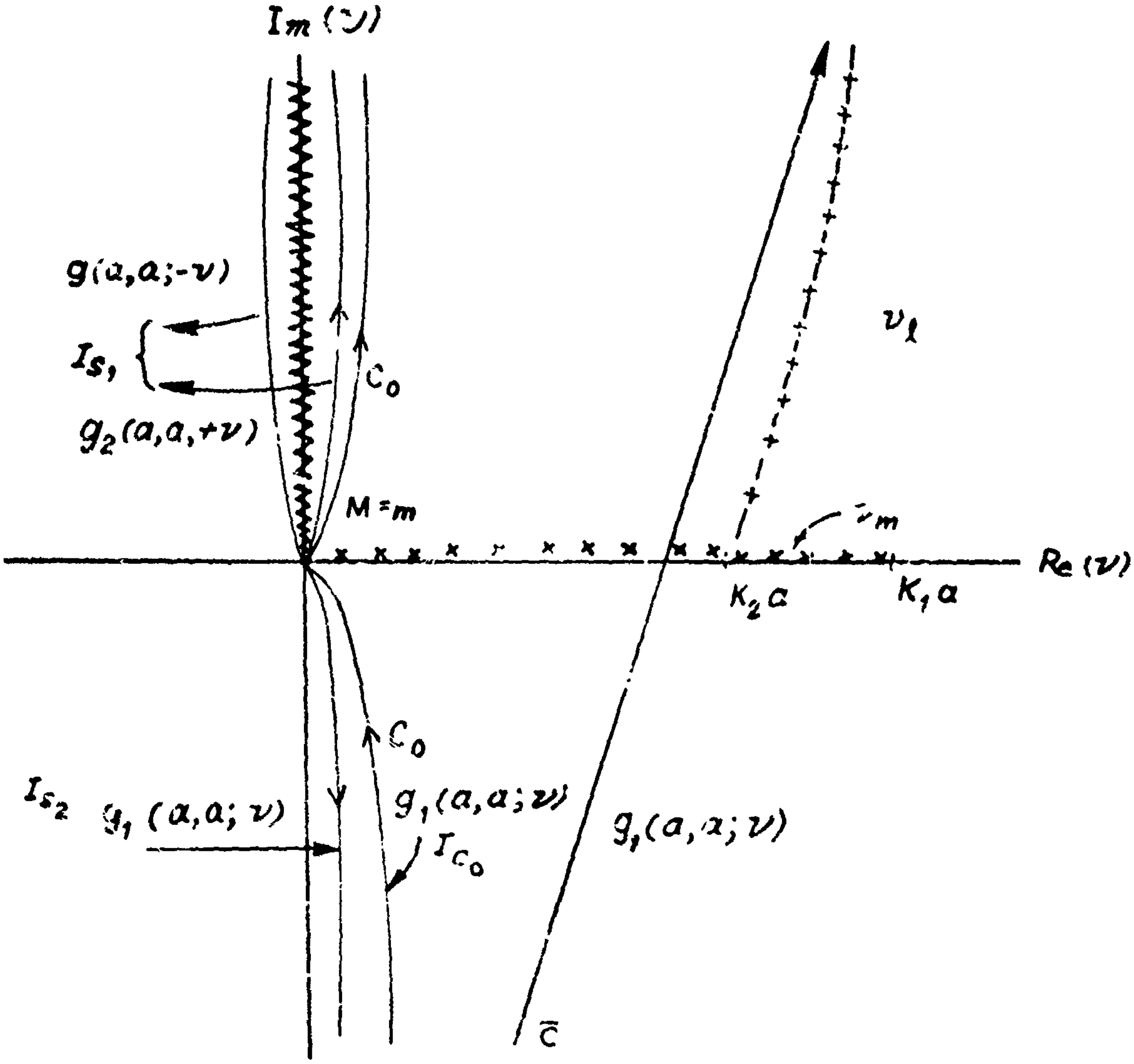

Im (ν)
Re (ν)
g(a,a;-ν)
I_{s_1}
g_2(a,a,+ν)
C_0
M = m
ν_m
ν_l
K_2a
K_1a
C_0
I_{s_2}
g_1 (a,a; ν)
g_1(a,a;ν)
I_{c_0}
g_1(a,a; ν)
$\bar{C}$

Fig. 3 Integration path decomposition in complex ν plane. Also shown are the integrals I_{c_o}, I_{s_1} and I_{s_2}. When the integration contour C_o is changed to $\bar{C}_N$ with the same integrand, one obtains in addition to the integral along $\bar{C}_N$ the residues representing the whispering gallery modes from m = M to m = M + 1.

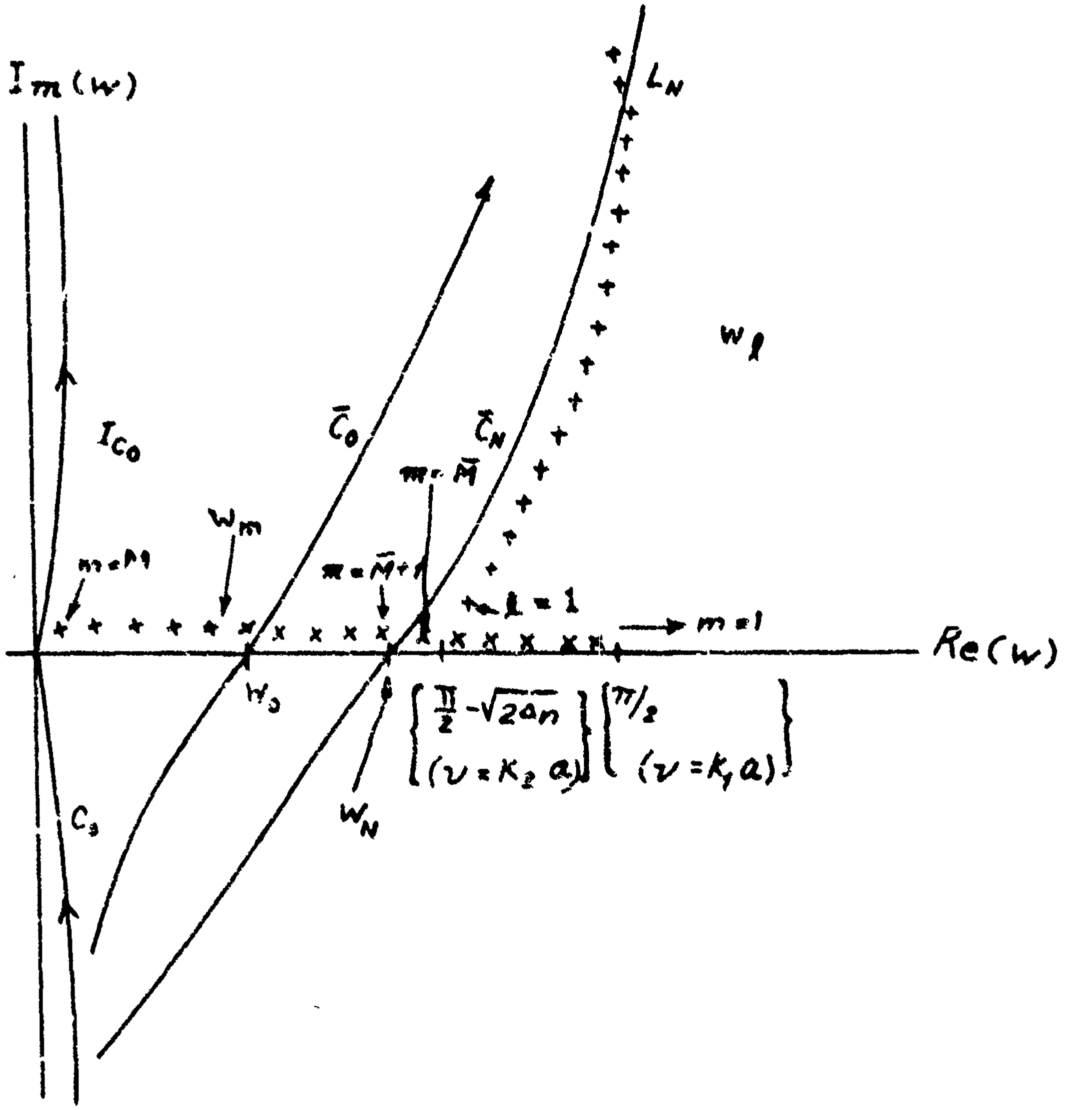
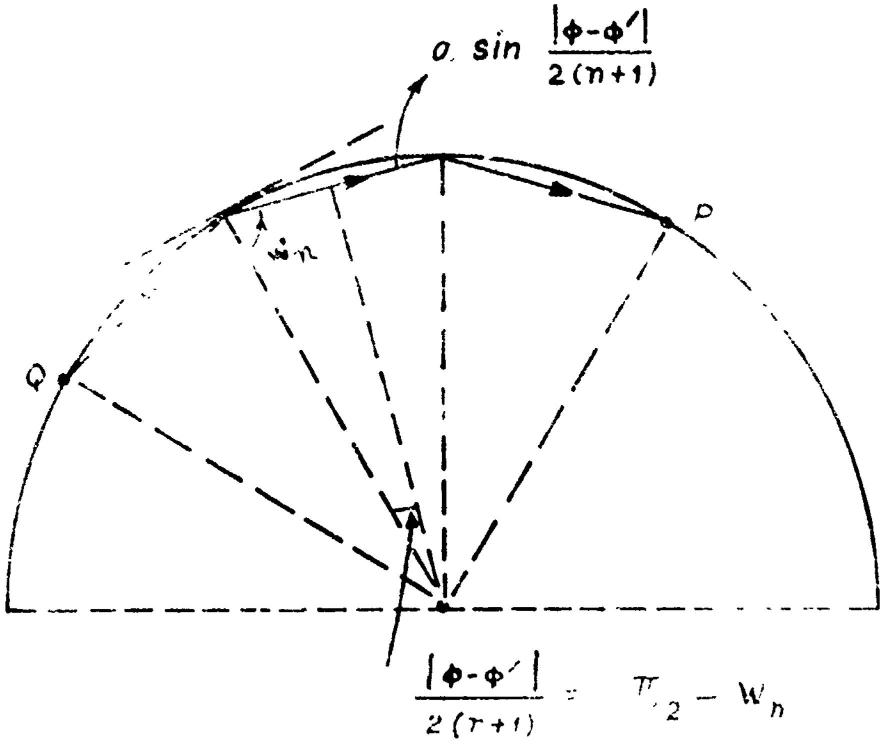
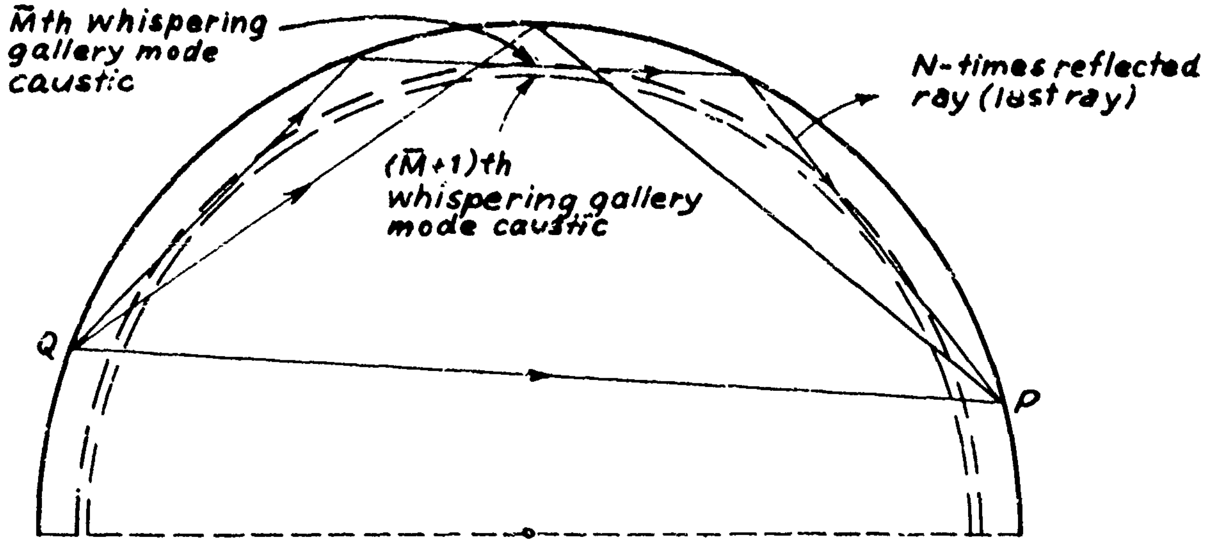

Im(w)
L_N
w_ℓ
I_{C_0}
$\bar{C}_0$
$\bar{C}_N$
$m = \bar{M}$
W_m
$m = M$
$m = \bar{M}+1$
$\ell = 1$
$m = 1$
Re(w)
W_0
$\left\{\begin{matrix}\frac{\pi}{2} - \sqrt{2a_n} \\ (\nu = k_2 a)\end{matrix}\right\}$
$\left\{\begin{matrix}\pi/2 \\ (\nu = k_1 a)\end{matrix}\right\}$
W_N
C_0

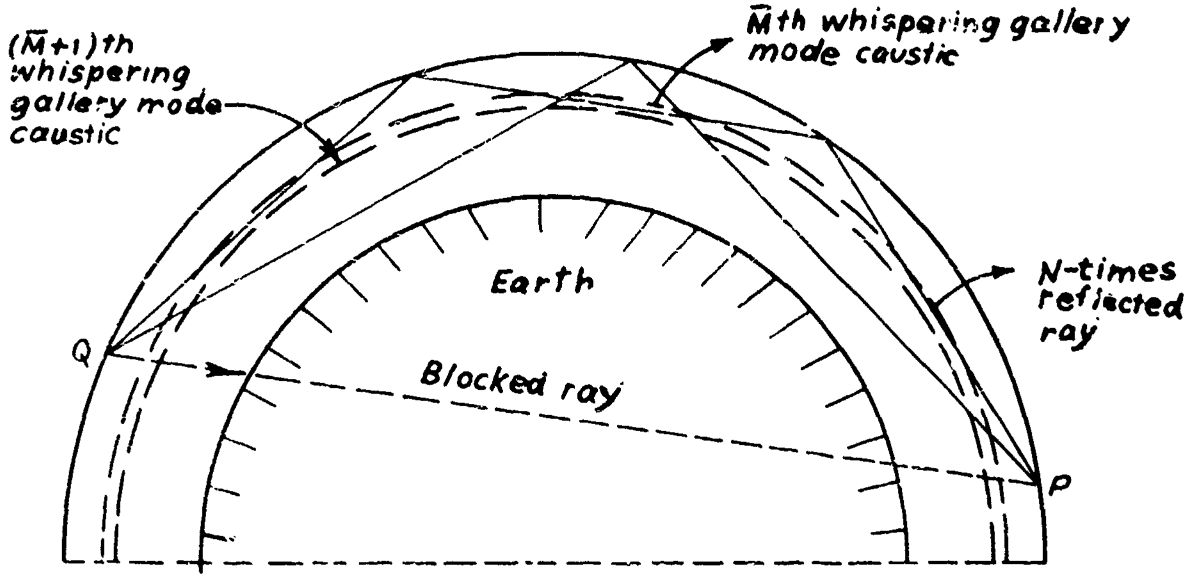
Fig. 1 Steepest descent paths and singularities in w-plane ($\nu = k_1 a \sin \alpha$). The saddle points for the direct ray and the N times reflected ray (last ray) are w_0 and w_N, respectively. When the remainder integral (see equation (31a)) is along $\tilde{C}_N$, one has in addition to the integral the contributions from the whispering gallery mode poles $m = \bar{M} + 1$ to M. Also the deformed path may capture creeping waves with $\ell \geq N$.


$a \sin \frac{|\phi-\phi'|}{2(n+1)}$
P
Q
W_n
$\frac{|\phi-\phi'|}{2(r+1)} = \pi/2 - W_n$

[illegible] Geometric optical r [illegible]


$\bar{M}$th whispering gallery mode caustic
N-times reflected ray (last ray)
($\bar{M}$+1)th whispering gallery mode caustic
Q
P

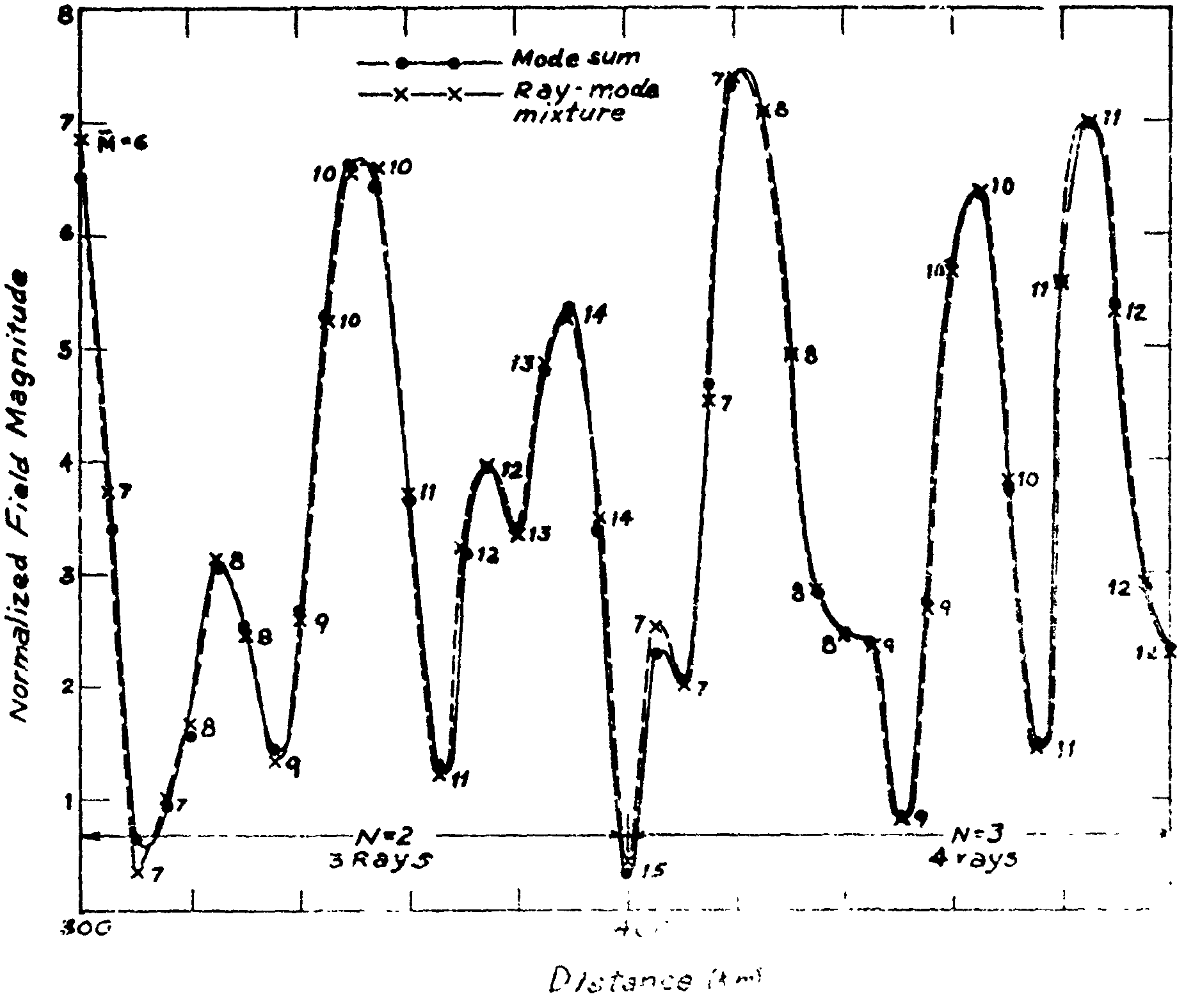
(a) In this sketch, without the earth's boundary, the effects of modes with caustics smaller than the $\bar{M}$th mode caustic are accounted for by ray optical fields with N 2.

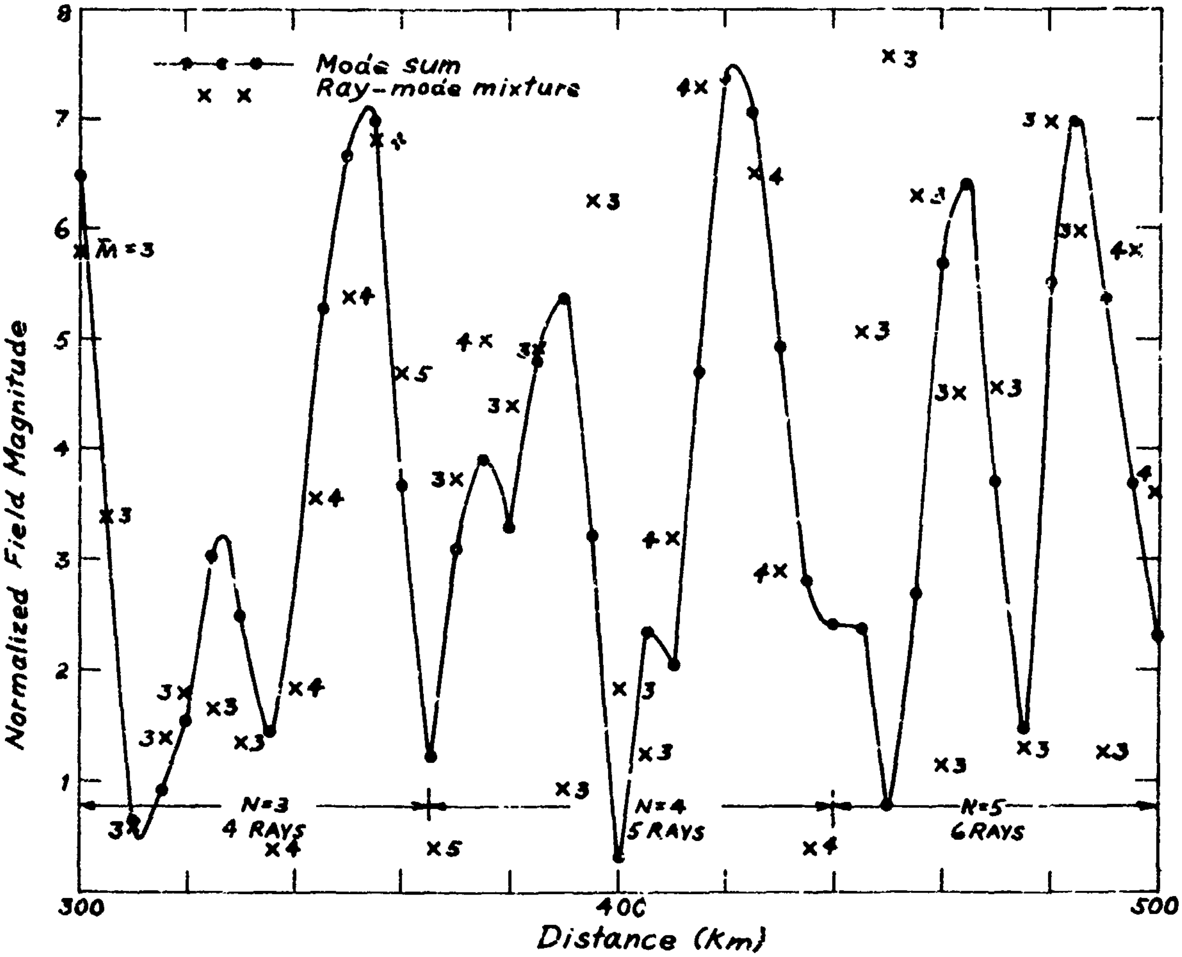

($\bar{M}$+1)th whispering gallery mode caustic
$\bar{M}$th whispering gallery mode caustic
Earth
N-times reflected ray
Q
Blocked ray
P

(b) In this sketch, in presence of the earth's boundary, the direct ray is blocked and omitted from the calculation.

Fig. 6 Pictorial representation of ray-mode combination. The $\bar{M}$ whispering gallery modes included are those whose caustics fit between the interface and the N-times reflected ray trajectory.

Fig. 7 Comparison of mixture of rays and whispering gallery modes with whispering gallery mode sum. The reference solution involves summation over 30 whispering gallery modes and is shown as the solid curve. The hybrid representation of modes and rays is shown as the dashed curve, and $\bar{M}$ and N indicate the number of modes and rays, respectively. Actually computed points are indicated by • for the mode sum and X for the ray-mode mixture, respectively, with the number next to X denoting $\bar{M}$. The number N of included rays is shown near the bottom scale. The direct ray (n = 0) is omitted from the calculation (see Fig. 6(b)).





7(b) 3 ≤ M < 6 so that some of the whispering gallery modes with very small loss are omitted. The ray-optical remainder may now be appreciable.

INFLUENCE DES MODELES D'IONOSPHERE SUR LES CALCULS DE PROPAGATION DES ONDES DECAMETRIQUES

R. FLEURY et P. GOURVEZ
Centre National d'Etudes des Télécommunications
Departement ETN/MIR - 22301 LANNION - FRANCE -

RESUME -

La fréquence critique foF2 de la couche F2 ionosphérique est l'un des parametres géophysiques les plus importants parmi ceux qui interviennent dans la propagation des ondes décamétriques. L'incertitude sur la connaissance de foF2 introduit notamment l'une des principales limitations à l'efficacité des outils très puissants actuellement disponibles pour les calculs de propagation (e.g. : la trajectographie numérique). Dans la présente étude, une comparaison systématique entre les distributions planétaires de la médiane mensuelle foF2 calculée à l'aide de modèles couramment utilisés et les distributions observées permet de déterminer des barres d'erreur pour foF2. Puis, des valeurs typiques de l'erreur commise sur foF2 sont introduites comme perturbations d'un modele de couche F2 utilisé conjointement à une méthode de tracé de rayon afin d'estimer leur influence sur les résultats de calculs de propagation des ondes décamétriques.

ABSTRACT -

The ionospheric F2-layer critical plasma frequency foF2 is one key geophysical parameter among those involved in H.F. waves propagation. This means that the efficiency of numerical methods available for H.F. waves propagation simulation (e.g. : ray tracing) is greatly limited by the uncertainties inherent to the foF2 input models. In the present study, a systematic comparison between observed planetary distributions of monthly median foF2 and those calculated by current models gives values of error bars on foF2. Further, an accurate ray tracing method used with F2-region models in which foF2 is perturbed by amounts equal to typical values of its error bars helps to assess the influence of these error bars on H.F. waves propagation results.

I - INTRODUCTION -

La simulation de la propagation des ondes décamétriques dans l'ionosphère requiert la modélisation de la propagation des ondes électromagnétiques d'une part, et celle du milieu de propagation d'autre part. Dans le cadre de l'optique géométrique, auquel nous nous limitons ici, la modélisation de la propagation est clairement établie. Dans ce cadre un modèle d'ionosphère doit essentiellement fournir une représentation des variations spatio-temporelles lentes (temps caractéristique > 20 mn, longueur caractéristique > quelques kms) de la fréquence plasma en fonction de divers paramètres géophysiques (activité solaire, saison, temps local, position géographique, altitude). Les modèles d'ionosphère existants sont suffisamment conformes à la réalité pour qu'ils aient pu permettre la mise au point de techniques de calcul de la propagation de très bonne précision (par exemple, la trajectographie numérique). Cependant, le degré de précision des modèles d'ionosphère n'est actuellement pas aussi satisfaisant que celui de la modélisation de la propagation elle-même. En conséquence dans les études de prévision de propagation ou d'évaluation des systèmes de télécommunication en ondes decamétriques, l'incertitude sur les résultats obtenus par simulation provient principalement de celle qui est introduite par la représentation de l'ionosphère. Le but de la présente étude est de contribuer à une tâche qui apparaît donc indispensable : rechercher une évaluation chiffrée de ces incertitudes.

Le but assigné a donne lieu à deux sous-études. La première consiste en une comparaison systématique entre des modèles d'ionosphère et des résultats de l'observation provenant du réseau mondial de stations ionosphériques. Elle est limitée au cas des modèles fournissant des valeurs médianes mensuelles de la fréquence critique ordinaire de la couche F2 (foF2). Des barres d'erreur pour foF2 calculée sont ainsi déterminées. La seconde sous-étude utilise des valeurs typiques de ces barres d'erreur sur foF2 afin d'estimer l'erreur qu'elles impliquent sur les valeurs calculées de paramètres caractérisant la propagation des ondes décamétriques.

II - CHOIX du PARAMETRE IONOSPHERIQUE ETUDIE ET DES MODELES TESTES -

II.1 - Choix du paramètre foF2

Du point de vue de l'optique géométrique des ondes décamétriques, les paramètres qui caractérisent l'état de chaque couche ionosphérique sont sa fréquence plasma critique, et l'altitude où celle-ci est localisée. Dans la basse ionosphère (couches E et F1), ces paramètres présentent des variations spatio-temporelles régulières, car en très grande partie liées à la variation de l'angle solaire zénithal. La modélisation donne une bonne représentation de leurs valeurs médianes mensuelles, l'écart relatif avec les médianes déduites de l'observation étant généralement inférieur à quelques pour cent (Rosich et col. 1973, Muggleton 1975). Par contre il n'en est pas de même pour les paramètres de la couche F2. Les variations spatio-temporelles de la fréquence critique foF2 et de l'altitude hmF2 où elle est localisée résultent en effet de mécanismes physiques plus nombreux, les mouvements d'ionisation jouant un rôle prépondérant.

En conséquence, foF2 et, dans une moindre mesure hmF2, présentent des variations beaucoup moins régulières que celles des paramètres de la basse ionosphère ; aussi la modélisation en rend-elle moins bien compte. D'autre part pour les radiocommunications en ondes décamétriques, les trajectoires d'énergie qui se réfléchissent dans la couche F2 sont les plus importantes. Or pour ces modes de propagation, la valeur de foF2 est déterminante par rapport à celle de hmF2. Ainsi à fréquence critique foF2 fixée, la simulation de ces modes donne des résultats très voisins quel que soit le profil vertical de fréquence plasma utilisé, à condition bien entendu que son allure générale ne soit pas trop irréaliste (Rush et col. 1975). De plus, un écart sur la valeur de foF2 influe beaucoup plus sur les résultats de la simulation des modes empruntant la couche F2 qu'un écart sur hmF2. Ceci est illustré, par exemple, par l'étude de l'influence respective de la variabilité jour à jour de foF2 et de hmF2 (Rush et col. 1974).

Nous limitons donc notre étude à celle des écarts entre les médianes mensuelles de foF2 déduites de l'observation et celles qui sont calculées à l'aide de modèles.

II.2 - Choix des modèles testés de foF2 -

La modélisation de la couche F2 peut être effectuée par résolution du système d'équations régissant la partie ionisée de l'atmosphère neutre terrestre, couplé à un modèle pour l'atmosphère neutre elle-même. Cette méthode conduit à des programmes de calcul beaucoup trop lourds pour qu'ils puissent être utilisés couramment dans la simulation de la propagation des ondes décamétriques. Aussi les modèles de foF2 médiane mensuelle testés ici sont ceux dont la construction repose sur l'analyse des médianes observées fournies par le réseau mondial de stations ionosphériques. Cette analyse consiste à approcher (au sens des moindres carrés) l'ensemble des valeurs médianes mensuelles observées de foF2 par des fonctions des paramètres géophysiques dont l'état de la couche F2 dépend. Les modèles de foF2 obtenus dépendent bien sûr tous des mêmes paramètres géophysiques (activité solaire, saison, temps local, position géographique) mais diffèrent par le type des fonctions choisies. Nous avons testé les plus couramment utilisés, à savoir :

- le modèle dit "d'OSLO" : les variations géographiques de la médiane mensuelle foF2 pour un mois et une heure TU donnés sont représentées par des séries de fonctions harmoniques sphériques de la latitude, de la longitude et de l'inclinaison magnétique modifiée. La variation diurne des coefficients de ces séries est représentée au moyen d'une série de Fourier. La variation de foF2 avec l'activité solaire est exprimée en fonction de la moyenne glissante sur 12 mois (R12) du nombre de Wolf mensuel. Pour un mois donné, foF2 est supposée croître linéairement pour R12 variant de 0 à 150 et être constante égale à sa valeur pour R12 égal à 150, lorsque R12 est supérieur à 150 (CCIR, Rapport 340, 1967). En fait, le modèle d'Oslo est issu des travaux de Jones et col. (1962, 1966) qui n'envisagent pas de régression sur foF2 en fonction de l'activité solaire. Pour notre part, nous disposons de plusieurs jeux de coefficients correspondant chacun à un mois et à une activité solaire donnés, établis depuis janvier 1967 par le NOAA. Ces jeux de coefficients sont déduits par régression parabolique fonction de R12 de ceux calculés par Jones et col. (1962, 1966) pour les années 1954 à 1958. Le modèle dit d'Oslo testé ici utilise la même régression parabolique en fonction de l'indice d'activité solaire R12.

- Le modèle dit "de NEW-DELHI" : le modèle précédent est construit en considérant chaque mois séparément et ne comporte aucune régression de foF2 en fonction de la saison. Le modèle de New-Delhi issu des travaux de Jones et col. (1970) prend en compte les variations saisonnières au moyen de séries de Fourier dont les coefficients sont calculés à l'aide de 60 mois de données. De plus, le modèle de New-Delhi comporte explicitement une représentation des variations de foF2 avec l'activité solaire. Il suppose une relation parabolique entre les coefficients des séries de fonctions de la position géographique et un indice d'activité solaire qui est soit R12, soit Ø12. Ø12 (en 10^{-22} Wm^{-2}) est la moyenne glissante sur 12 mois du flux solaire radioélectrique mesuré sur la longueur d'onde égale à 10.7cm. Le modèle de New-Delhi testé ici est celui dans lequel l'activité solaire est exprimée en fonction de Ø12.

- Le modèle de Ching et Chiu (1973, 1975) : il est construit à l'aide de fonctions analytiques choisies au vu des résultats d'études morphologiques des variations de foF2. L'avantage de cette représentation mathématique est qu'elle permet des calculs informatiques de foF2 moins coûteux en taille mémoire et temps de calcul. De plus, étant fondée sur des études morphologiques préalables, elle devrait à priori rendre mieux compte de la médiane foF2 vraie.

III - ETUDE de la QUALITE DES MODELES de foF2 -

III.1 - Mesure de la qualité des modèles -

Les écarts entre les valeurs observées de la médiane mensuelle foF2 et ses valeurs calculées pour les mêmes conditions géophysiques à l'aide d'un modèle permettent de juger la qualité de celui-ci. Pour une station, un mois et une heure donnés, la médiane observée n'est retenue comme donnée que si le nombre des mesures ayant servi à la déterminer est strictement supérieur à trois. De plus, pour la même station et le même mois, l'ensemble des médianes observées n'est utilisé que si le premier critère a tout de même permis de retenir la totalité des 24 valeurs horaires. D'autre part, les médianes mensuelles calculées à l'aide d'un modèle le sont en utilisant la valeur observée de l'indice d'activité solaire pour le mois considéré. Les écarts sur foF2 dus à l'erreur de prévision sur cet indice sont donc éliminés.

Pour une station, un mois et l'heure locale i, l'écart entre les médianes foF2 observée (f_{mi}) et calculée (f_{ci}) sera considéré :

- en valeur absolue : $\Delta f_i = f_{ci} - f_{mi}$ (MHz)
- et en valeur relative à la médiane observée : $(\Delta f/f)i = \Delta f_i/f_{mi}$ (%).

En fait, dans le but de réduire le nombre des grandeurs à manipuler, nous etudierons la moyenne des écarts sur les 24 heures :

$$x = \frac{1}{24} \sum_{i=1}^{24} x_i \quad , \quad \text{où} \quad x_i = \Delta f_i \quad \text{ou} \quad (\Delta f/f)_i$$

. Nous étudierons conjointement l'écart type :

$$Sx = \frac{1}{24} \left[\sum_{i=1}^{24} (x_i - x)^2 \right]^{1/2} \quad \text{où} \quad x = \Delta f \quad \text{ou} \quad \Delta f/f,$$

qui, pour le mois et la station donnes, caracterise la dispersion diurne des ecarts autour de leur valeur moyenne x.

L'objectif est d'obtenir des valeurs typiques des ecarts et dispersions ainsi definis. Mais dans le même temps, nous recherchons si des relations caracteristiques existent entre ces paramètres et les paramètres geophysiques en fonction desquels sont faites les principales regressions sur lesquelles reposent les modèles (la position géographique et l'activite solaire). En effet la mise en evidence de telles relations pourrait permettre d'eventuelles corrections des modèles testes.

III 2 - les ecarts en fonction de la position géographique -

Pour un mois donné, et donc une activité solaire fixee, les ecarts x et leur dispersions diurnes Sx sont représentés en fonction de la latitude geographique d'une part, et de la longitude d'autre part. La seule variation caracteristique qui en ressorte est liée à un effet saisonnier et concerne les distributions en latitude pour les années d'activité solaire faible ou moyenne (R12 ⩽ 105). Nous présentons à titre d'exemple les distributions géographiques des ecarts relatifs $\Delta f/f$ et de leurs dispersions diurnes $S_{\Delta f/f}$ afférents au modèle de New-Delhi pour Janvier et Juillet 1965. Ce sont deux mois pour lesquels les activite solaires sont très voisines ($\Phi 12$ égale respectivement $73.8 \cdot 10^{-22}$ Wm^{-2} et $76.4 \cdot 10^{-22}$ Wm^{-2}). Les répartitions en longitude sont quelconques (Figures 1 et 2). Par contre, un effet saisonnier peut être mis en évidence sur les distributions en latitude. $\Delta f/f$ pour le mois de Janvier (Figure 3a) est généralement négatif dans l'némisphère d'été. Dans l'hemisphère d'hiver (Figure 3a), il peut être négatif ou positif, la moyenne en latitude étant positive. $\Delta f/f$ pour le mois de Juillet est generalement négatif dans l'hémisphère d'éte (Figure 4a). Dans l'hémisphère d'hiver il peut être négatif ou positif, sa moyenne en latitude etant ici positive, voisine de zéro. En resumé, les valeurs calculees de foF2 sont (en moyenne diurne) inférieures aux valeurs observees l'ete, et supérieures (ou egales) aux valeurs observées l'hiver. De plus la dispersion diurne $S_{\Delta f/f}$ des ecarts relatifs est plus faible dans l'hémisphère d'été que dans celui d'hiver (figures 3b et 4b). La variation diurne est donc mieux modélisee dans le cas de l'été. Les mêmes conclusions peuvent être tirées de l'etude des ecarts absolus Δf et de leurs dispersions $S_{\Delta f}$. Elles ressortent aussi de l'analyse en fonction de l'inclinaison magnétique modifiée qui est en realité la variable latitudinale intervenant dans les modèles du type de celui de New-Delhi. Cet effet saisonnier n'a pu être mis en evidence que pour les annees d'activite solaire faible ou moyenne. Il sera montré plus précisément dans le paragraphe traitant des ecarts en fonction de la saison.

D'autre part, les valeurs des écarts et de leurs dispersions diurnes sont dans l'ensemble comprises à l'intérieur de domaines limités (sur les exemples precedents : $|\Delta f/f| \leqslant 15$ %, $S_{\Delta f/f} \leqslant 10$ % ; soit pour les écarts absolus : $|\Delta f| \leqslant 0.75$ MHz et $S_{\Delta f} \leqslant 0.5$ MHz). Mais quelques points sont très en dehors de ces limites et de plus entre deux stations très voisines les variations des ecarts et dispersions peuvent être très importantes. La disparité dans la qualité des mesures d'une station à l'autre est probablement à l'origine de ces points aberrants et discontinuités.

Ces resultats de l'analyse en fonction de la position géographique conduisent à etudier les écarts sur foF2 et leurs dispersions diurnes moyennes sur trois bandes de latitude : la zone Nord, la zone Sud et la zone Equatoriale (± 20° de latitude), zone de transition entre les deux premieres. Cette procédure permet de réduire l'influence du facteur d'erreur propre à chaque station et d'obtenir des valeurs typiques des écarts et de leurs dispersions diurnes. Pour qu'elles soient significatives et comparables d'une bande de latitude à l'autre, les moyennes sont calculees en respectant deux regles :

- les stations dont les mesures sont utilisées sont reparties le plus uniformement possible en longitude (aucune longitude n'est privilégiée) ;
- tout en respectant cette première règle, le plus grand nombre possible de stations est retenu (l'erreur aléatoire d'une station à l'autre en est ainsi d'autant plus reduite).

Les moyennes en latitude sont ainsi calculees à l'aide de 10, 8 et 12 points pour respectivement la zone Nord, Equatoriale et Sud.

III.3 - Les ecarts en fonction de l'acitivite solaire -

Les écarts moyens sur 24 heures et leurs dispersions diurnes sont etudies en fonction de l'activité solaire en valeurs moyennes par bande de latitude. Conjointement ils sont etudies en valeurs non moyennees en latitude. Une évaluation globale de la qualité de la régression en fonction de l'activite solaire est en effet obtenue à l'aide d'histogrammes du nombre de points (i.e. de stations) par intervalle de valeurs non moyennees en latitude des écarts d'une part et de leurs dispersions diurnes d'autre part. Ces histogrammes sont determinés à l'aide du maximum disponible de points de mesure.

Dans les deux cas, pour un mois fixe, le mois de Janvier, quatre niveaux d'activite solaire sont retenus.

JANVIER	1965	1972	1970	1958
R_{12}	11.7	70.8	105.6	199.0
Φ_2 (10^{-22}W.m^{-2})	73.80	120.50	154.60	241.60
Nombre de points utilisés pour les histogrammes	98	94	[illegible]	121

Seule l'étude des écarts relatifs sur foF2 est ici présentée de manière détaillée. L'allure générale des variations des écarts absolus est en effet identique à celle des écarts relatifs. De plus, dans l'étude de l'influence des écarts affectant foF2 sur la propagation des ondes décamétriques, ce sont les écarts relatifs qui interviennent.

III.3. A) Cas du modèle d'Oslo :

Dans les zones Equatoriale et Sud, les écarts relatifs moyens en latitude augmentent avec l'activité solaire, tandis que dans la zone Nord ils varient peu (figure 5a). Pour les activités solaires peu élevées ($R_{12} \lesssim 70$), les écarts relatifs concernant foF2 calculée pour la zone Equatoriale sont les plus faibles. Mais ils deviennent les plus forts (avec ceux concernant la zone Sud) pour les activité solaires élevées ($R_{12} \gtrsim 70$) Les écarts absolus et relatifs maximum (i.e. : quelle que soit la bande de latitude) sont respectivement voisins de 0.4 MHz et 7.6 %, si l'année 1958 ($R_{12} \simeq 200$) n'est pas prise en compte. Si elle l'est, ils valent respectivement 1 MHz et 9 %.

Dans les zones Equatoriales et Nord, les dispersions diurnes des écarts, moyennes en latitude, augmentent avec l'activite solaire lorsque celle-ci est faible ou moyenne ($R_{12} \lesssim 100$) et diminuent avec l'activité solaire lorsque celle-ci est élevée ($R_{12} > 100$) (figure 5b). Les dispersions diurnes les plus faibles concernent la zone Sud où, quelle que soit l'activité solaire, $S_{\Delta f} \simeq 0.45$ MHz et $S_{\Delta f/f} \simeq 7$ %. Les dispersions diurnes maximales des écarts relatifs et absolus concernent les zones Equatoriale et/ou Nord, pour l'année 1970 ($R_{12} \simeq 100$) et sont respectivement voisines de 0.9 MHz et 16 %.

En résumé, la régression en fonction de l'activité solaire devient mauvaise, en moyenne sur 24 heures, lorsque cette activité est très élevée. Mais dans ce cas la variation diurne est mieux cernée par le modèle. Au total, les bornes supérieures des domaines de variation de Δf et de $\Delta f/f$ sont fournies par les valeurs trouvées pour les années de soleil calme ou moyennement actif. Notons toutefois que ces bornes varient peu avec l'activité solaire. (Figure 5c, trait épais). Le domaine typique de variation de l'écart relatif (quelles que soit l'activité solaire, la position géographique et l'heure) est donc :

$$-14\ \% \leqslant \frac{\Delta foF2}{foF2} \leqslant 23\ \%.$$

Ces résultats sont confirmés par l'examen des histogrammes du nombre de points par intervalle de valeurs des écarts relatifs et de leurs dispersions diurnes respectivement, non moyennées en latitude. Pour les années d'activité solaire faible ou moyenne (figures 7, 8, 9 ; courbes en trait fin), les écarts relatifs sont en majorité contenus dans un domaine centré au voisinage de 0 % : par exemple, pour l'année 1965, 81 % des valeurs sont comprises dans le domaine ± 9 %. Pour ces mêmes annees, les dispersions diurnes des écarts relatifs (Figures 12, 13, 14, courbes en trait fin) sont en majorite contenues dans un domaine centré au voisinage de 9 % ; pour 1965 par exemple, 87 % des valeurs sont dans le domaine 9 ± 6 %. Par contre, pour l'année 1958, le domaine contenant la majorité des écarts relatifs (figure 10, courbe en trait fin) est centré au voisinage de 6 %, 87 % des points étant dans le domaine 6 ± 9 %. Mais le domaine des dispersions diurnes est maintenant centre au voisinage de 7 %, 87 % des points étant dans le domaine 7 ± 7 %. Pour l'année de forte activité solaire et à l'échelle planétaire, nous retrouvons le décalage des écarts vers des valeurs plus élevées ainsi que celui des dispersions diurnes vers des valeurs plus faibles.

III.3. : B) Cas du modèle de New-Delhi :

Les écarts absolus et relatifs moyens en latitude pour les fortes activités solaires sont réduits par rapport au cas du modèle d'Oslo (figure 5a). Leurs valeurs maximales sont respectivement voisines de 0.7 MHz et 6 %, contre 1 MHz et 9 % précedemment. Les dispersions diurnes des écarts sont pratiquement inchangées (figure 5b). La régression en fonction de l'activité solaire donne donc, pour les fortes activités solaires, des valeurs de foF2 plus proches des valeurs vraies (quelle que soit l'heure puisque les dispersions diurnes sont inchangées sinon plus faibles). Néanmoins, les écarts relatifs pour les années d'activité solaire moyenne (1972, 1970) sont sensiblement accrus, notamment pour les zones Equatoriale et Nord (Figure 5a). En définitive, le domaine typique de variation des écarts relatifs (quelle que soit l'heure, la position géographique et l'activité solaire) est lui-même légèrement accru : $-16\ \% \leqslant \frac{\Delta foF2}{foF2} \leqslant 26\ \%$

(figure 5c, trait pointillé)

Les histogrammes du nombre de stations par intervalle de valeurs non moyennées des écarts relatifs confirment ces conclusions. Pour l'année 1958, le modèle de New-Delhi est globablement meilleur que celui d'Oslo. L'histogramme des écarts (figure 10, trait épais) est maintenant centré au voisinage de 0 %, 86 % des points étant situés dans le domaine ± 9 %. Celui des dispersions diurnes (figure 14, trait épais) est pratiquement inchangé. Pour les années d'activité solaire moyenne, le modèle de New-Delhi est légèrement moins bon. Pour l'année 1972, l'histogramme des écarts (figure 8, trait épais) est centré au voisinage de + 4 % (contre 1 % pour le modèle d' Oslo, figure 8, trait fin). De même pour l'annee 1970, 87 % des points sont dans le domaine ± 15 % (figure 9, trait épais) ; au lieu de 87 % de points dans le domaine ± 12 % pour les écarts déduits du modèle d'Oslo (figure 9, trait fin).

D'autre part, les histogrammes par intervalle de valeurs des dispersions diurnes (figures 11 à 14) sont pratiquement inchangés passant du modèle d'Oslo (trait fin) à celui de New-Delhi (trait épais); quelle que soit l'activité solaire.

III.3. : C) Cas du modèle de Ching et Chiu :

La variation en fonction de l'activité solaire des écarts moyens en latitude est semblable à celle des écarts afférents au modèle de New-Delhi (figure 5a). Mais (sauf pour les zones Equatoriales et Sud dans le cas d'activités solaires faibles ou moyennes ($R_{12} \leqslant 100$))les écarts moyens sur foF2 sont ici plus importants (figure 5a). En particulier, pour l'activité solaire très élevée ($R_{12} \simeq 200$), les écarts absolus et relatifs maximums sont respectivement voisins de - 2.5 MHz et - 24 %. En moyenne sur les 24 heures foF2 calculée a donc tendance à être inférieure à foF2 observée. D'autre part, les dispersions diurnes des écarts varient avec l'activité solaire de manière similaire à celles afférentes au modèle de New-Delhi, mais sont généralement supérieures (valeurs entre 10 et 26 %, contre 6 et 17 %, figure 5b). En fait, l'examen des écarts en fonction de l'heure, en diverses stations, révèle que la variation diurne de foF2 est correctement modélisée mais souvent en retard par rapport à celle de foF2 observée (retard $\gg$ 1h surtout aux heures de lever du soleil ; cf. les comparaisons effectuées par Ching et Chiu 1973). Ceci explique au moins en partie les écarts et dispersions diurnes importants. En tout état de cause, les domaines de variation de ΔfoF2 et ΔfoF2/foF2 suivent, en fonction de l'activité solaire, la même évolution que précédemment, mais sont plus larges (figure 5c, trait tireté). Le domaine extrême de variation des écarts relatifs (i.e. : quelles que soient l'activité solaire, la position géographique et l'heure) est :

$$- 47\ \% \leqslant \frac{\Delta \mathrm{foF2}}{\mathrm{foF2}} \leqslant 40\ \%.$$

Les histogrammes par intervalles de valeurs non moyennées en latitude des écarts relatifs, pour les activités solaires faibles et moyennes (figures 7, 8, 9, traits en pointillés), présentent un maximum plat et large. Pour l'année 1965 ($R_{12} \simeq 10$) par exemple, 84 % des points sont situés dans le domaine ± 21 % (dans le cas des modèles précédents, un pourcentage similaire de points était trouvé dans le domaine ± 9 %). En outre, les écarts forts sont ici plus nombreux ; pour 1965 (respectivement 1972), 16 % (resp. 19 %) des points sont au-delà des limites ± 21 %. Pour l'année d'activité solaire très élevée (figure 10, traits en pointillés), l'histogramme présente un maximum prononcé au voisinage de - 17 %, 81 % des points étant dans le domaine - 17 ± 20 %. Ceci confirme que pour l'activité solaire très élevée ($R_{12} \simeq 200$), les valeurs de foF2 calculées à l'aide du modèle de Ching et Chiu sont, en moyenne sur 24 heures, généralement inférieures aux valeurs observées. Par ailleurs, quelle que soit l'activité solaire, les histogrammes par intervalles de valeurs des dispersions diurnes présentent un maximum prononcé au voisinage de 20 %, plus de 80 % des points étant compris dans le domaine 20 ± 10 % (figures 11 à 14, traits en pointillés). Ceci confirme que la variation diurne de foF2 n'est pas correctement cernée.

III.4 - Les écarts en fonction de la saison :

L'effet saisonnier mis en évidence sur les distributions en latitude des écarts moyens sur 24 heures et sur celles de leurs dispersions diurnes pour les activité solaires faibles et moyennes (paragraphe III.2) est ici précisé dans le cas du modèle de New-Delhi pour l'année 1972. L'année 1972 correspond à une activité solaire moyenne et variant au plus de 13 % au cours des 12 mois (valeurs extrêmes de Φ_{12} : 123 et 109.10^{-22} Wm^{-2}, pour mai et décembre respectivement). Les écarts relatifs moyens en latitude (figure 6a) sont positifs pour l'hiver, de valeur maximum voisine de 8 % pour l'hiver-Nord, et de 12 % pour l'hiver-Sud. Ils sont négatifs pour l'été-Nord de valeur absolue maximum voisine de 10 % et négatifs voisins de zéro ($\approx \pm 2$ %) pour l'été-Sud. Autrement dit, en moyenne sur 24 heures, foF2 calculée pour les activité solaires moyennes présente une anomalie d'hiver renforcée par rapport à l'anomalie d'hiver réelle. Pour les activités solaires faibles, foF2 calculée présentera une anomalie d'hiver qui n'existe pas en réalité. D'autre part les dispersions diurnes moyennes en latitude des écarts relatifs (figure 6a) présentent un maximum voisin de 14 - 16 % l'hiver et un minimum voisin de 4 - 6 % l'été. La variation diurne de foF2 calculée est donc plus proche de celle de foF2 observée dans le cas de l'été que dans le cas de l'hiver. En définitive, le domaine de variation de $\Delta f/f$ quelle que soit l'heure (figure 6b) comprend principalement des valeurs positives dans le cas de l'hiver ; il ne comprend que des valeurs négatives dans le cas de l'été en hémisphère Nord, et est situé autour de zéro dans celui de l'été en hémisphère Sud. Ceci confirme que l'anomalie d'hiver affectant foF2 calculée est plus marquée qu'elle ne devrait l'être, ou bien existe contrairement à la réalité. Le domaine maximum de variation de $\Delta f/f$ (quelles que soient la position géographique, l'heure, la saison) obtenu pour l'année 1972 est : $- 12\ \% \leqslant \Delta \mathrm{foF2}/\mathrm{foF2} \leqslant 28\ \%$ (figure 6b). Le domaine obtenu pour Janvier seul (déjà déductible de la figure 5c) est : $- 6\ \% \leqslant \Delta \mathrm{foF2}/\mathrm{foF2} \leqslant 23\ \%$. Les valeurs des écarts affectant foF2 dans le cas de l'été élargissent donc sensiblement le domaine de variation de ΔfoF2/foF2.

IV - ERREUR SUR LES RESULTATS DE CALCULS DE LA PROPAGATION INDUITE PAR L'ERREUR SUR foF2

L'erreur commise sur la médiane mensuelle foF2 entraîne une erreur sur les caractéristiques calculées pour les trajectoires des ondes décamétriques qui empruntent la couche F2. Une estimation quantitative de cette seconde erreur est obtenue par comparaison entre les caractéristiques de la propagation calculées par tracé de rayon à travers un modèle de couche F2 médiane et celles qui le sont lorsque la fréquence critique foF2 est modifiée selon des pourcentages typiques des valeurs de l'erreur sur foF2 déduites de l'analyse statistique précédente.

Le tracé de rayon consiste en la résolution numérique des équations de l'optique hamiltonienne (Haselgrove 1954) ; le programme de calcul informatique développé est similaire à celui établi par R.M. Jones et col. (1975). L'indice de réfraction incorporé dans les équations différentielles régissant les trajectoires d'énergie ne tient compte ni du champ géomagnétique, ni des collisions entre les électrons et les ions et/ou les particules neutres. Les résultats sont valides pour le mode ordinaire en cas de prise en compte du champ géomagnétique.

Le modèle tridimentionnel de fréquence plasma choisi pour la couche F2 est celui de Ching et Chiu (1973-1975). Le profil vertical qu'il utilise est donné pour une formule de Chapman, qui pour les calculs de propagation donne des résultats réalistes (Rush et col. 1975). En outre, l'application de la méthode montre que, à pourcentages de perturbation de foF2 identiques, les variations relatives des caractéristiques de la propagation obtenues sont du même ordre, que le modèle de la couche F2 initiale soit celui d'Oslo (ou New-Delhi) ou celui de Ching et Chiu. Un avantage pratique important de ce dernier est qu'il permet de diviser par deux (sinon plus) le temps de calcul d'une trajectoire.

Les caractéristiques de la propagation calculées pour les modèles non perturbé et perturbé de la couche F2 concernent le premier bond des trajectoires et sont : la distance au sol en fonction de l'élévation d'émission, pour différentes fréquences d'émission (e.g. : figure 15) ; la distance au sol en fonction de la fréquence maximum utilisable (MUF) qui lui correspond (e.g. : figure 16) ; l'angle d'élévation à l'émission de la trajectoire correspondant à la MUF en fonction de cette MUF (e.g. : figure 16b).

Un exemple des résultats obtenus à l'aide de la méthode qui vient d'être décrite concerne la propagation aux latitudes moyennes européennes, l'émetteur étant situé à Lannion (48.45°N, 356.73°E) et l'azimut de la direction d'émission étant voisin de 60°. Le modèle de couche F2 utilisé est celui qui est obtenu pour le mois de Janvier 1958 à 1200 TU. Pour ce cas, le plus petit domaine de valeurs pour l'erreur sur foF2 calculée à l'aide des modèles actuels est tel que : $|\Delta foF2/foF2| \leqslant 15\ \%$ (figure 5c). Les bornes supérieures (± 15 %) de ce domaine sont donc utilisées comme valeurs du pourcentage de perturbation foF2.

La distance en fonction de l'élévation, calculée à l'aide du modèle non perturbé est représentée sur la figure 15 (en trait épais) pour des fréquences d'émission égales à 20, 30 et 40 MHz. A fréquence fixée, partant des élévations faibles et à élévation croissante, la distance au sol décroit, passe par un minimum (Dmin), puis croît rapidement jusqu'à ce que les élévations atteignent le seuil au-delà duquel les trajectoires traversent la couche ionosphérique. La fréquence d'émission fixée est la MUF pour la distance Dmin. Notons que les fréquences réfléchies par la couche F2 appartiennent ici au haut de la gamme H.F. étant donné la situation de très forte activité solaire ($R_{12} \simeq 200$). Pour les mêmes fréquences, la distance en fonction de l'élévation, calculée lorsque le modèle est perturbé par un accroissement de 15 % de foF2 est représenté en trait pointillé sur la figure 15. L'effet de cette erreur simulée de + 15 % sur foF2 est une décroissance de la distance atteinte, quelles que soient l'élévation et la fréquence d'émission. L'amplitude de cette décroissance augmente avec l'élévation et la fréquence. Elle est inférieure à ou voisine de 10 % pour les distances atteintes par les trajectoires réfléchies dans le bas de la couche F2 ; par exemple celles de la fréquence 20 MHz pour les élévations allant de 0 à 40°, ou celles des fréquences 30 et 40 MHz pour les élévations inférieures à 12° et 22° respectivement. Pour les trajectoires réfléchies plus près du maximum de la couche F2, la diminution de la distance au sol induite par l'erreur de 15 % sur foF2 est plus importante : par exemple, à 30 MHz, elle atteint - 36 % pour l'élévation égale à 27°. Notons que les distances en fonction de l'élévation, obtenues lorsque l'erreur simulée sur foF2 est égale à - 15 %, présentent la même allure mais correspondent à un accroissement d'ensemble.

La distance au sol atteinte par une trajectoire en fonction de la MUF qui lui correspond est calculée dans le cas du modèle initial (figure 16a, courbe B), et dans les cas où foF2 est diminuée de 15 % (courbe A) ou accrue de 15 % (courbe C). L'erreur relative induite sur cette distance (qui n'est autre que Dmin) par l'erreur simulée de ± 15 % sur foF2 croît avec la fréquence. De plus, à MUF donnée, l'amplitude de l'erreur sur Dmin est plus forte dans le cas de l'erreur négative sur foF2. Ainsi, pour une MUF égale à 25 MHz, elle vaut + 27 % si foF2 est réduite de 15 % et - 22 % si foF2 est accrue de 15 % ; pour une MUF égale à 40 MHz, les valeurs correspondantes sont + 39 % et - 20 %. Néanmoins, à distance donnée, l'erreur relative sur la MUF correspondante est constante quelle que soit cette distance et voisine de ± 15 % selon que l'erreur sur foF2 vaut elle-même ± 15 %.

L'élévation de la trajectoire correspondant à la MUF en fonction de celle-ci fournit des résultats analogues (figure 16b, courbes A, B, C). Pour une élévation d'émission donnée, l'erreur constante de ± 15 % sur la MUF est évidemment retrouvée ici. De plus, l'erreur relative sur l'élévation due à celle sur foF2 croît avec la fréquence, passant de ± 25 % environ pour une fréquence égale à 24 MHz à - 46 % et + 30 % pour une fréquence égale à 40 MHz. Là encore, l'erreur négative (- 15 %) sur foF2 induit une erreur plus forte, ici sur l'élévation, que celle induite par l'erreur positive. Cet effet est marqué surtout aux fréquences élevées.

Des résultats identiques aux précédents ont été obtenus pour un cas de faible activité solaire (Janvier 1965), l'erreur simulée sur foF2 étant prise égale à ± 15 % (figure 5c). Les fréquences alors réfléchies par la couche F2 sont plus faibles ($\leqslant$ 20 MHz). Par ailleurs les calculs effectués pour les mêmes cas de propagation, mais l'erreur simulée sur foF2 étant égale à ± 30 %, montrent que les erreurs résultantes sur les caractéristiques du premier bond sont de morphologies semblables à celles décrites ci-dessus mais d'amplitudes presque doubles.

CONCLUSION

L'étude des écarts entre les valeurs de la médiane mensuelle foF2 calculées à l'aide de modèles et les valeurs déduites des mesures du réseau mondial d'ionosondes a permis de tester la qualité des régressions utilisées pour la construction des modèles. Elle a en outre permis de déterminer des valeurs numériques pour les barres d'erreur à affecter à foF2 calculée. L'erreur relative sur foF2 calculée à l'aide des modèles utilisant des séries de fonctions est comprise dans le domaine (- 15 %, + 25 %) quelles que soient l'activité solaire, l'heure et la position géographique. Le domaine de variation de l'erreur sur foF2 calculée à l'aide du modèle phénoménologique de Ching et Chiu est sensiblement plus large, ce qui semble dû à un déphasage entre les variations diurnes calculée et réelle de foF2. De plus, la construction de ce modèle est basée sur l'utilisation des données d'une cinquantaine de stations seulement. Il pourrait être amélioré plus facilement qu'un autre étant donnée sa structure reposant sur des fonctions déduites d'études morphologiques de la couche F2.

Les calculs rigoureux de la propagation des ondes décamétriques nécessitent des ressources informatiques importantes. Le modèle d'ionosphère entrant dans ces calculs doit donc être aussi performant que possible, ce qui est le cas du modèle de Ching et Chiu. Utilisant ce modèle pour la couche F2 conjointement à une technique de trajectographie numérique, des barres d'erreur sur les caractéristiques calculées de la propagation résultantes de celles qui affectent foF2 ont été déterminées. Des barres d'erreur typiques de ± 15 % sur foF2, voisines des erreurs minimales possibles à l'aide des modèles actuels de couche F2, induisent un domaine d'erreur (- 20 %, + 30 %) sur les distances, (- 25 %, + 25 %) sur les élévations et (- 15 %, + 15 %) sur les MUF calculées pour le premier bond des trajectoires des ondes de la gamme décamétrique.

REFERENCES -

- C.C.I.R., "Atlas des caracteristiques ionosphériques", Rapport 340-Oslo 1966, U.I.T. Genève, 1967.
- CHING B.K., Y.T. CHIU, "A phenomenological model of global ionospheric electron density in the E-, F1- and F2- region", J. Atmosph. Terr. Phys. Vol. 35, pp. 1615-30, 1973.
- CHIU Y.T., "An improved phenomenological model of ionospheric density" J. Atmosph. Terr. Phys., Vol. 37, pp. 1563-70, 1975.
- HASELGROVE J., "Ray theory and a new method for ray tracing", in the Physics of the Ionosphere.p. 355, The Phys. Soc., London, 1954.
- JONES W.B., R.M. GALLET, "Representation of diurnal and geographic variations of ionospheric data by numerical methods", J. of Research of the N.B.S., Vol. 66D, pp. 419-38, 1962.
- JONES W.B., R.P. GRAHAM, M. LEFTIN, "Advances on ionospheric mapping by numerical methods", N.B.S. Technical Note 337, 1966.
- JONES W.B., D.L. OBITTS; "Global representation of annual and solar cycle variations of foF2 monthly median 1954-1958", O.T./I.T.S. Research Rep. 3, ITS-Boulder-Co., 1970.
- JONES R.M., J.J. STEPHENSON, "A versatile three dimensional ray tracing computer program for radio waves in the ionosphere", O.T. Rep. 75-76, 1975.
- MUGGLETON L.M., "Une méthode pour la prévision de foE à un instant et en un lieu quelconques", J. des Télecom., Vol. 42, pp. 413-18, 1975.
- ROSICH R.K., W.B. JONES, "The numerical representation of the critical frequency of the F1-region of the ionosphere", O.T. Rep. 73-22, ITS-Boulder-Co., 1973.
- RUSH C.M., D. MILLER, J. GIBBS, "The relative daily variability of foF2 and hmF2 and their implications for H.F. radio propagation", Radio Science, Vol. 9, pp. 749-56, 1974.
- RUSH C.M., W.R. EDWARDS, "Sensitivity of H.F. circuit simulations to electron density models", Radio Science, Vol. 10, pp. 867-74, 1975.

LEGENDES DES FIGURES

Figure 1a :	Ecart relatif moyen sur 24 heures en fonction de la longitude géographique, dans le cas du modèle de New-Delhi, pour le mois de Janvier 1965.
Figure 1b :	Dispersion diurne de l'écart relatif en fonction de la longitude géographique, dans le cas du modèle de New-Delhi, pour le mois de Janvier 1965.
Figure 2a :	Idem que 1a pour le mois de Juillet 1965.
Figure 2b :	Idem que 1b pour le mois de Juillet 1965.
Figure 3a :	Ecart relatif moyen sur 24 heures en fonction de la latitude geographique, dans le cas du modèle de New-Delhi, pour le mois de Janvier 1965.
Figure 3b :	Dispersion diurne de l'ecart relatif en fonction de la latitude géographique, dans le cas du modèle de New-Delhi, pour le mois de Janvier 1965.
Figure 4a :	Idem que 3a pour le mois de Juillet 1965.
Figure 4b :	Idem que 3b pour le mois de Juillet 1965.
Figure 5a :	Ecart relatif moyen sur 24 heures, moyen en latitude, en fonction de l'activité solaire, pour les zones Nord, Equatoriale, Sud. De gauche à droite : cas du modèle d'Oslo, de New-Delhi, de Ching et Chiu.
Figure 5b :	Idem que 5a pour la dispersion diurne de l'écart relatif.
Figure 5c :	Domaine de variation de l'écart relatif (à gauche) et absolu (à droite) en fonction de l'activité solaire quelles que soient l'heure et la position géographique. Trait épais : cas du modèle d'Oslo. Trait pointillé : cas du modèle de New-Delhi. Trait tireté : cas du modèle de Ching et Chiu.
Figure 6a-b :	Ecart relatif moyen sur 24 heures, moyen en latitude (6a à gauche) et dispersion diurne de l'écart relatif moyen en latitude (6a à droite) en fonction de la saison pour les zones Nord, Equatoriale et Sud, dans le cas du modèle de New-Delhi. Domaine de variation de l'écart relatif (6b) pour les zones Nord, Equatoriale, Sud, quelle que soit l'heure, en fonction de la saison, dans le cas du modèle de New-Delhi.
Figure 7 :	Nombre de points (i.e. de stations) par intervalle de valeurs de l'écart relatif moyen diurne, dans les cas du modèle d'Oslo (trait fin), de New-Delhi (trait épais) de Ching et Chiu (trait pointillé) ; pour le mois de Janvier 1965.
Figures 8-9-10 :	Idem pour le mois de Janvier 1972, 1970, 1958 respectivement.
Figure 11 :	Nombre de points (i.e. de stations) par intervalle de valeurs de la dispersion diurne de l'écart relatif dans le cas du modèle d'Oslo (trait fin), de New-Delhi (trait épais), de Ching et Chiu (trait pointillé) ; pour le mois de Janvier 1965.
Fig. 12-13-14 :	Idem pour le mois de Janvier 1972, 1970, 1958, respectivement.
Figure 15 :	Distance au sol en fonction de l'angle d'élévation à l'émission, pour le premier bond des trajectoires d'ondes de fréquences fixées (en MHz) réfractées dans la couche F2 ; courbes en trait plein : cas du modèle initial de couche F2 ; courbes en trait pointillé : cas du modèle perturbé par une erreur simulée de + 15 % sur foF2.
Figure 16a :	Distance au sol minimale (Dmin) en fonction de la MUF correspondante, pour le premier bond des trajectoires d'ondes décamétriques réfractées dans la couche F2 ; courbe B : cas du modèle initial de couche F2 ; courbes A et C : cas du modèle perturbé par une erreur simulée sur foF2 de - 15 % et + 15 % respectivement.
Figure 16b :	Idem pour les angles d'élévations à l'émission des trajectoires correspondants à la MUF.

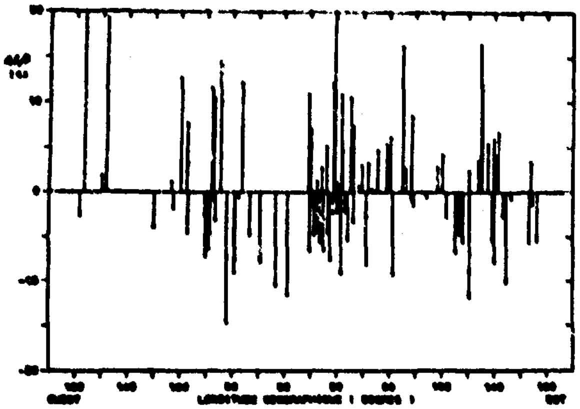


Figure 1a

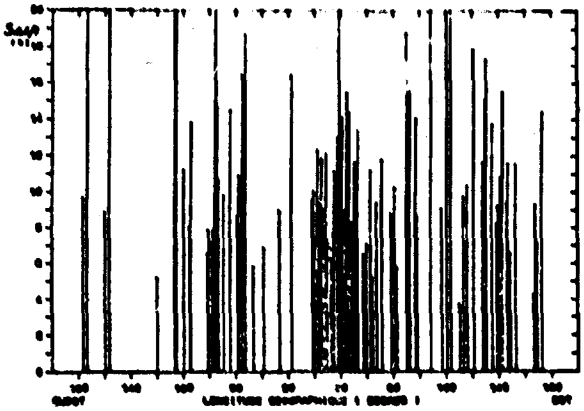


Figure 1b

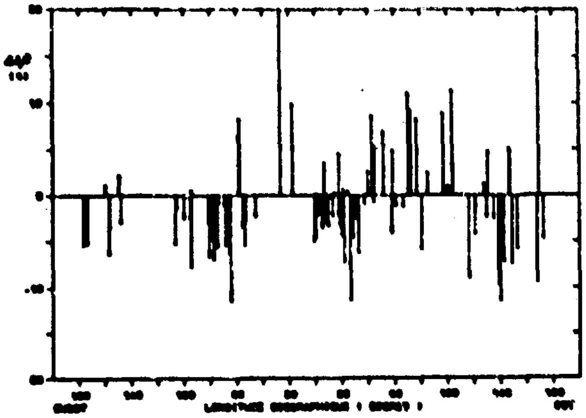


Figure 2a

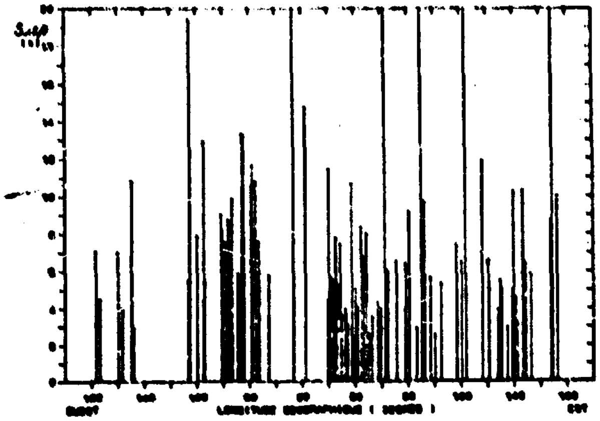


Figure 2b

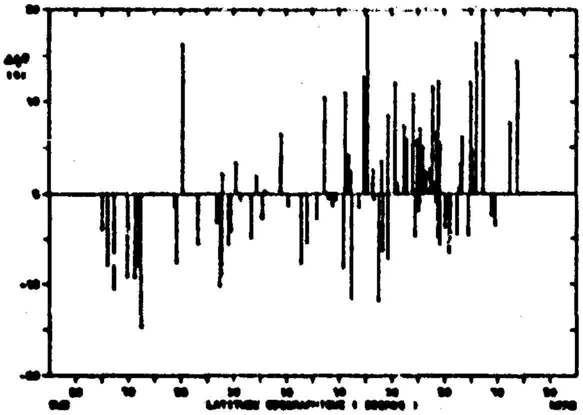


Figure 3a

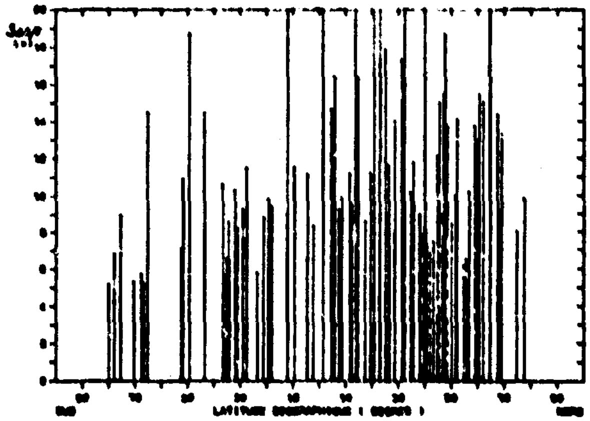


Figure 3b

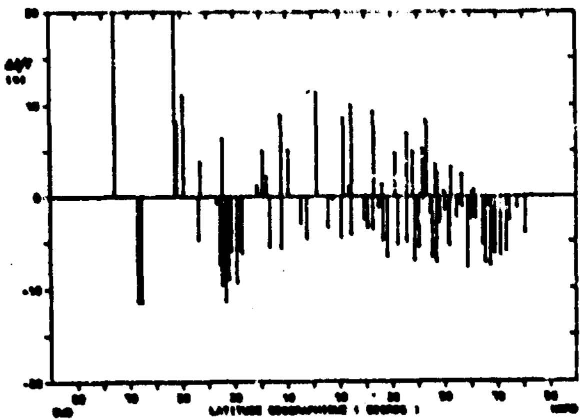


Figure 4a

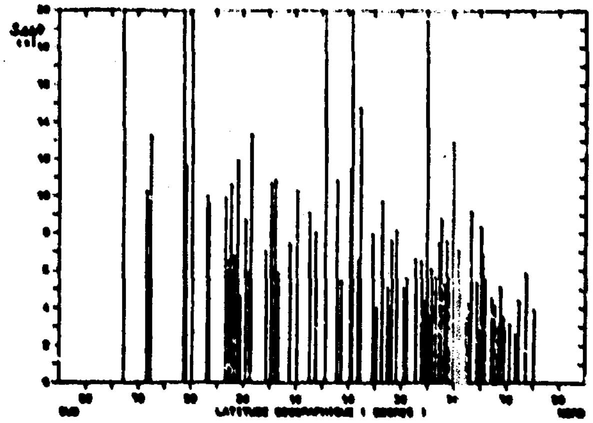
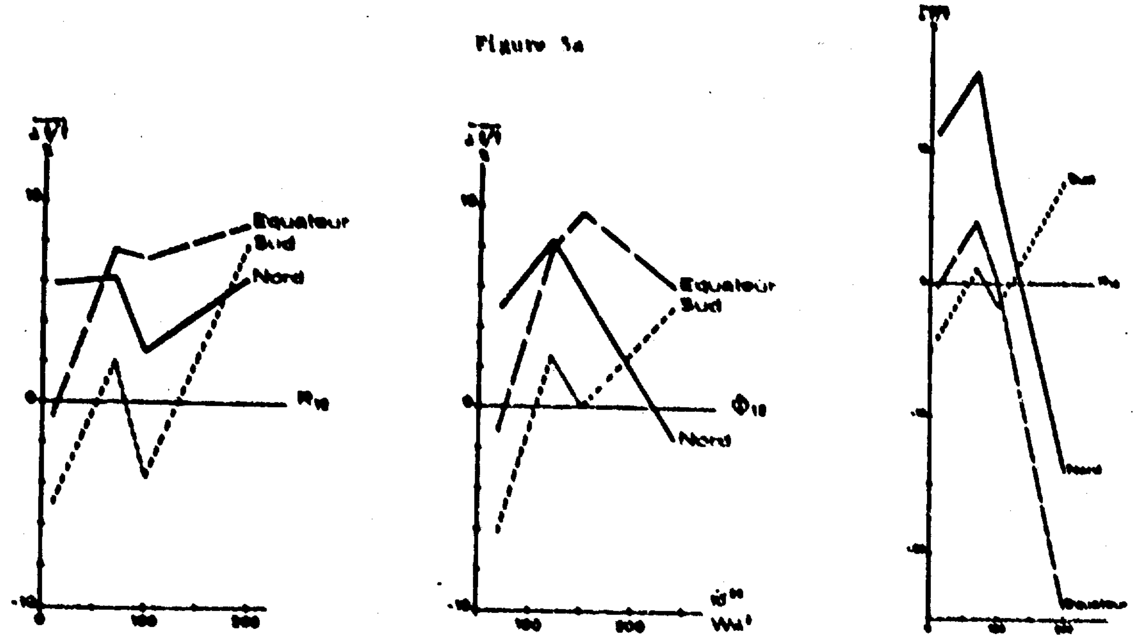


Figure 4b

Figure 5a



Cas du modèle d'Oslo

Cas du modèle de New-Delhi

Cas du modèle de Ching et Chiu

Figure 5b

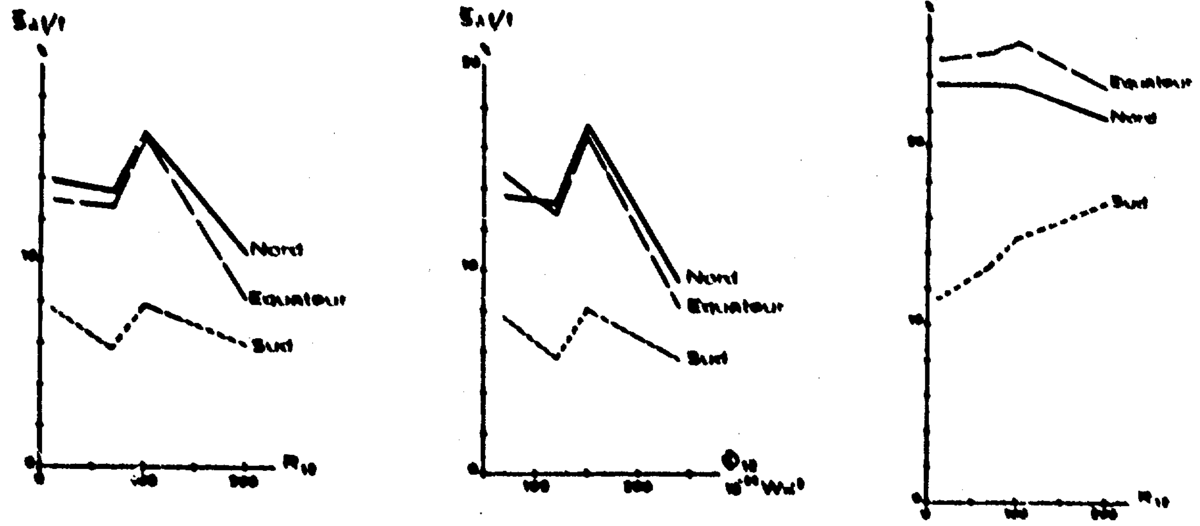


Figure 5c

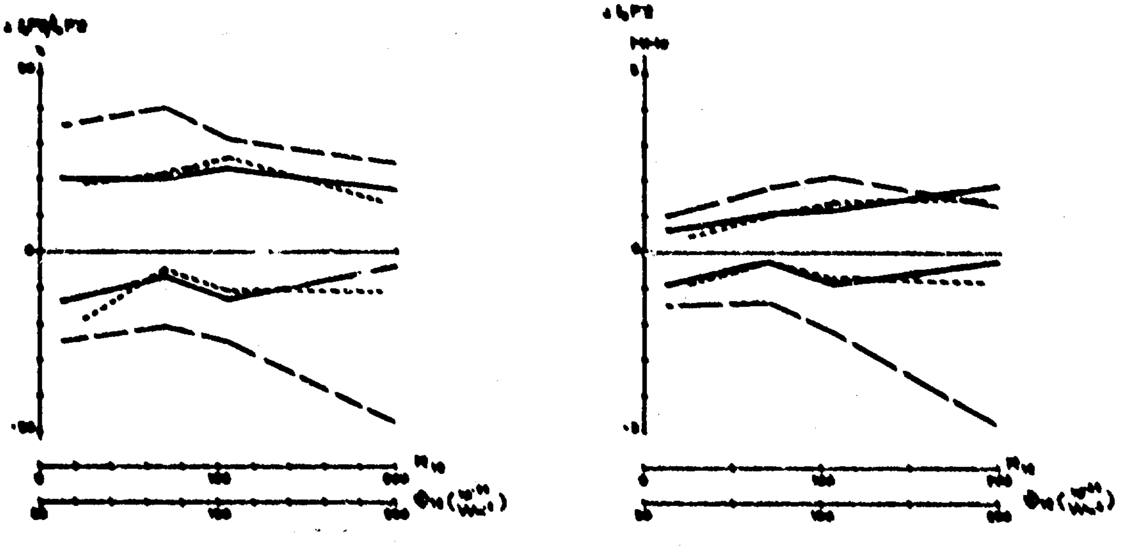


Figure 6a

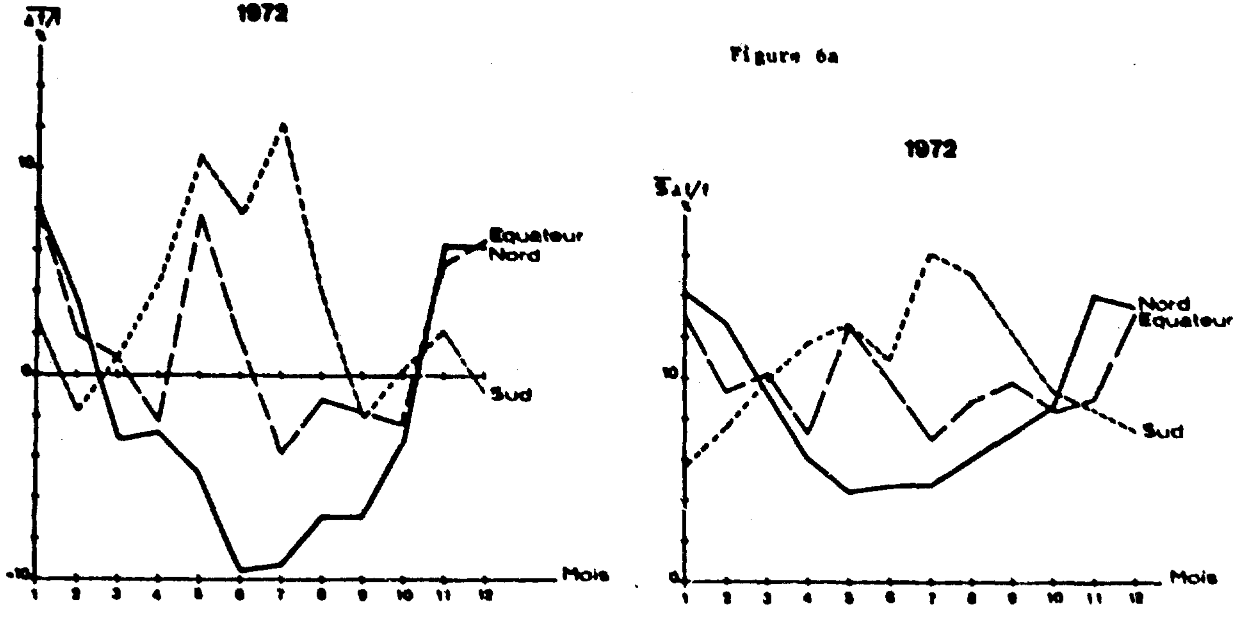
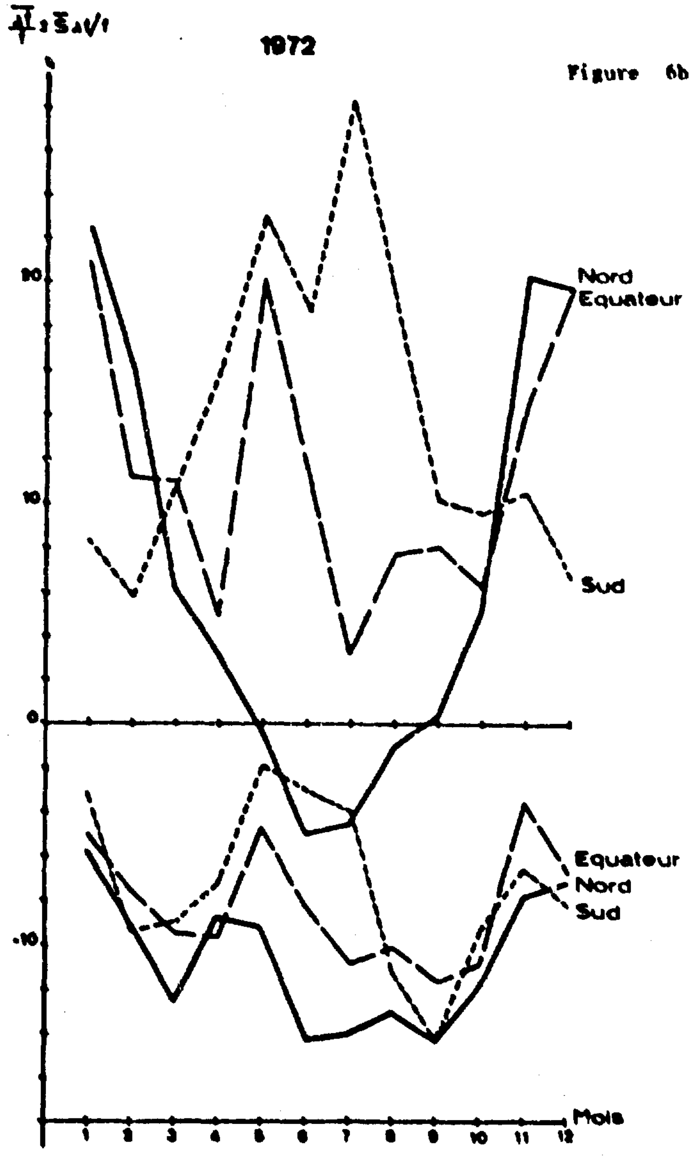


Figure 6b



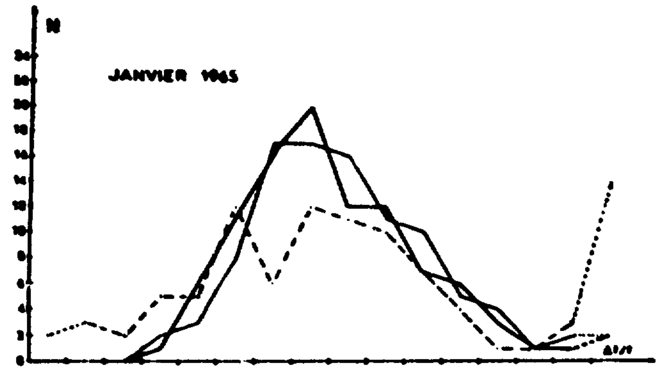


Figure 7

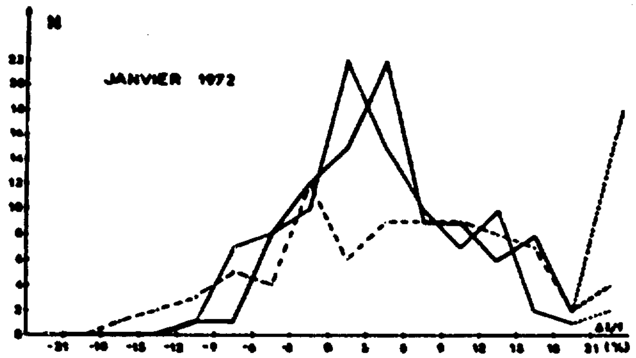


Figure 8

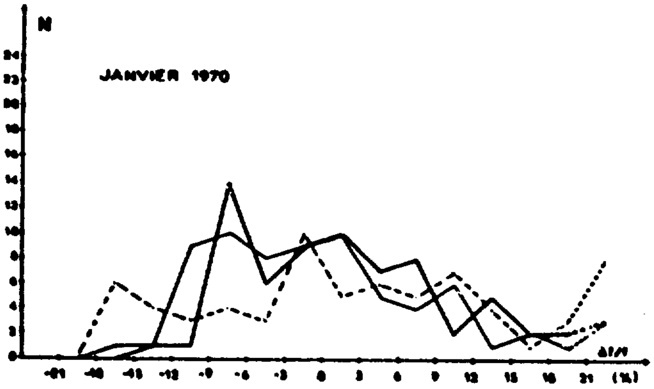


Figure 9

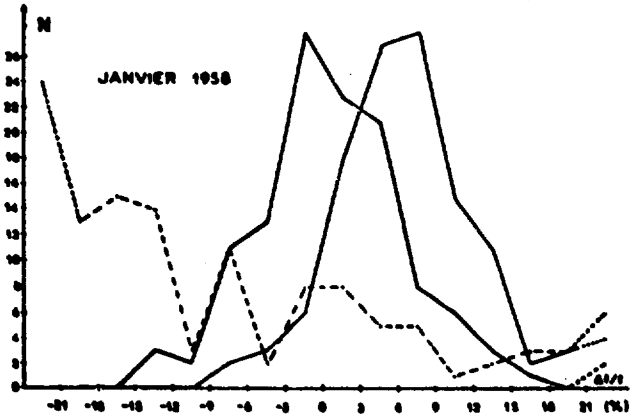


Figure 10

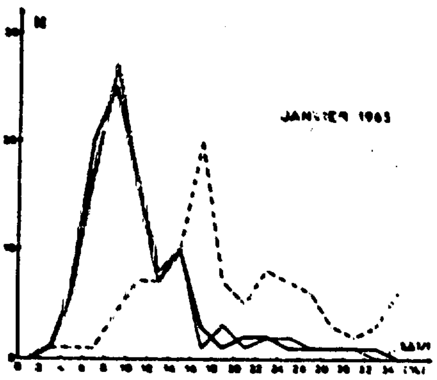


Figure 11

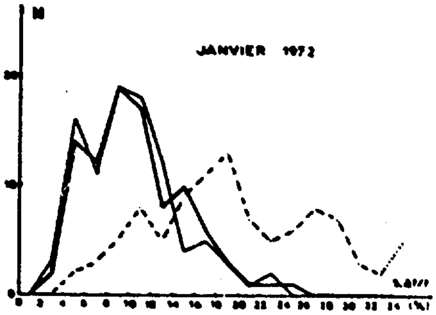


Figure 12

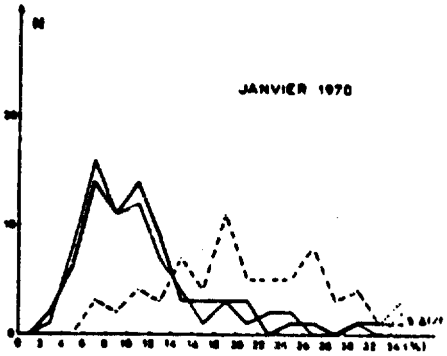


Figure 13

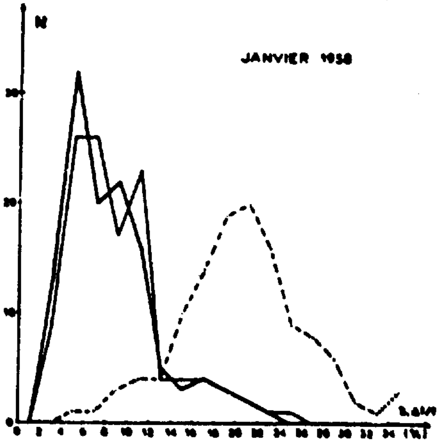


Figure 14

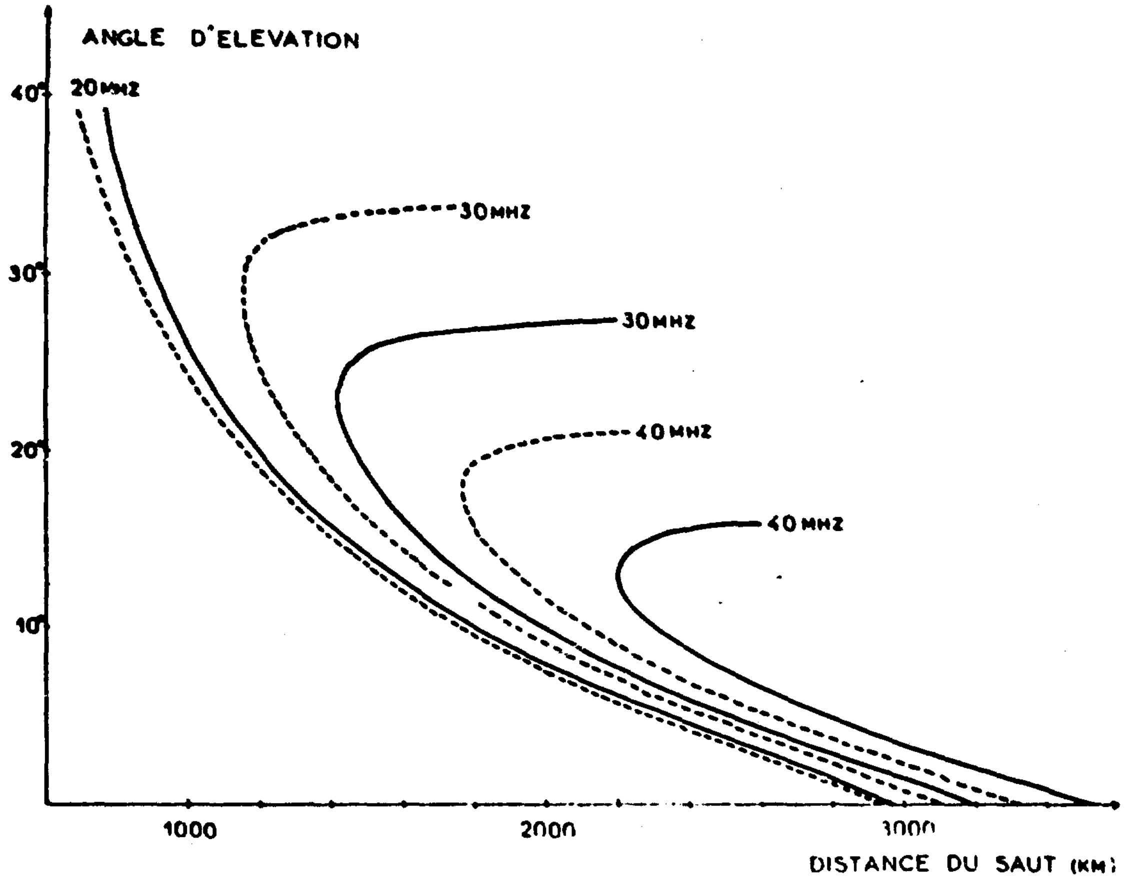


Figure 15

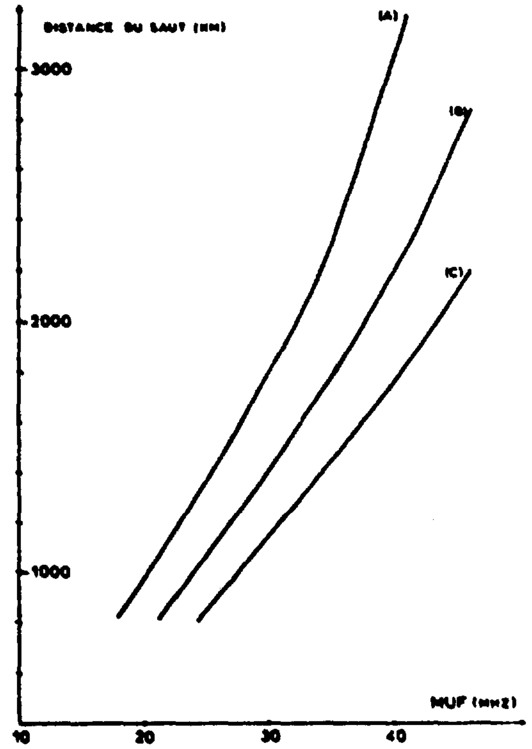


Figure 16a

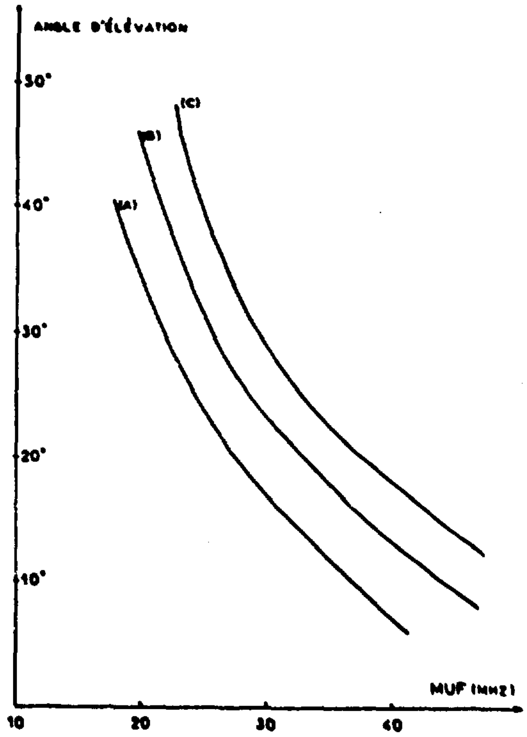


Figure 16b

DISCUSSION

L.W. Barclay, UK

Section 11.2. of the paper indicates that a linear variation of f_0F2 is assumed in the Oslo model for values of R_{12} between 0 and 100. However, Figure 5c, for example, includes an analysis of this model for values of R_{12} up to 200. What variation of f_0F2 with R_{12} is assumed for values between 100 and 200? CCIR Report 340 adopts a linear variation for R_{12} values up to 150.

Author's Reply

In the CCIR Report 340, Oslo 1966, the given planetary f_0F2 model, the so-called "Oslo model", is based on the work of Jones et col., 1962. In this work any regression with the solar activity is not considered. However, in the Oslo model a linear relationship of f_0F2 with the solar index R_{12} is assumed. For the high values of R_{12} the actual relationship noticeably departs from a linear one. In the Oslo model for any R_{12} value greater than 150 f_0F2 is calculated assuming R_{12} equal to 150. This permits to attenuate the unrealistic features due to the linear extrapolation but is in fact a very crude simplification. In the present study we have preferred to use the Oslo model in a more realistic way and have fitted the available uSK coefficients to a parabolic regression with the solar activity, similarly to what is done in the New-Delhi model. But perhaps because of the differences in the methods of construction of respectively the Jones et col. 1962 and the New-Delhi models, the parabolic regression gives in the first case a quasi-linear part and in the second a more curved part of a parabola. In fact, all through the presentation of this paper, rigorously I had to say a "modified" Oslo model in place of Oslo model.

METHODS OF DETERMINING IONOSPHERIC STRUCTURE FROM OBLIQUE SOUNDING DATA

R. F. DuBroff, N. Narayana Rao and K. C. Yeh
Department of Electrical Engineering
University of Illinois at Urbana-Champaign
Urbana, IL. 61801

SUMMARY

Methods of inverting the leading edge of a backscatter ionogram for ionospheric structure are presented. Two models of the electron density distribution are utilized for this purpose: (a) the quasi-parabolic (QP) layer with locally uniform parameters, and (b) the locally quasi-parabolic (LQP) layer with horizontal gradients. In each case, the model ionosphere may be represented by a multi-dimensional vector of the form $\vec{\xi} = [\xi_1, \xi_2, \ldots]$, which we may call the "ionostate" vector. The inversion procedure consists of starting with an initial estimate of $\vec{\xi}$ and employing an iterative technique to obtain a final value of $\vec{\xi}$ such that the mean squared error between the measured values of leading edge data and those computed from $\vec{\xi}$ is minimized. The inversion procedures are illustrated by means of examples for both cases.

1. INTRODUCTION

1.1. The Inversion Problem

It is well known that the group path P'(f) of a radio wave at frequency f is a nonlinear functional of the ionization density profile N. While it is a simple matter to compute P' for a given N either analytically for some simple profiles or numerically for more complex profiles, the inverse problem of computing N for a given P' becomes very difficult or nearly impossible in general.

It happens that, for vertical incidence, the group delay integral (magnetoionic effects ignored) can be transformed into the form of a convolution integral whose solution can be readily found. This technique and its variation have been used to invert the vertical ionograms to obtain true height profiles. However, for oblique rays, no such transformation has been found and its solution remains very difficult. To assist the solution of this difficult problem it is noted that the ionospheric morphologies and modeling have been subjects of study for many years. The rough first-order ionospheric behavior is therefore known. It is also possible that the ionospheric profile immediately above an oblique sounder can be deduced by sounding vertically from a co-located vertical sounder. In any event, it is assumed that knowledge exists, based on which a first-order ionospheric profile can be constructed. If this first-order profile does not depart from the true profile appreciably, it is possible to reformulate the problem by a linearization procedure. This reformulated problem is linear and much simpler to solve.

1.2. Oblique Sounding Data

There are two kinds of oblique sounding data: (a) Point-to-point oblique ionograms, and (b) Backscatter leading edge. Point-to-point oblique ionograms are traces of group path versus frequency for oblique propagation between two fixed points separated by some distance. For the one-hop mode of propagation, the reflection points of the rays therefore lie in the neighborhood of the midpoint between the transmitter and the receiver, irrespective of the frequency. Similarly, for the multiple-hop mode of propagation, reflection for each hop occurs in the neighborhood of a single location, for all frequencies capable of propagating in that mode. It can be easily conjectured that a one-hop mode oblique ionogram is capable of providing an equivalent profile of electron density valid near the midpoint between the transmitter and the receiver, but is not by itself too useful for deriving the horizontal gradients of electron density. The use of a two-hop mode trace in addition to the one-hop mode trace would increase the utility of the ionogram for deriving the horizontal gradients but still does not permit their determination continuously along the azimuthal direction from the transmitter to the receiver.

In the case of the backscatter leading edge, the transmitter and the receiver are located at about the same location and the time delay of the transmitted signal backscattered from the ground and then received at the receiver is measured as a function of frequency. For a given frequency, many returns are possible corresponding to all elevation angles of transmission and reception within the beamwidths of the transmitting and receiving antennas. There is however a minimum value for the time delay which occurs near the transition from the low angle ray mode of propagation to the high angle ray mode of propagation, as shown in Fig. 1(a). Thus for each frequency, a continuum of backscattered returns beginning with the minimum time delay return will be received. Alternatively, the situation can be thought of as a continuum of point-to-point oblique ionograms corresponding to continuously increasing values of ground range of the backscatter location away from the transmitter. Such a continuum of ionograms is shown in Fig. 1(b). The tangent curve to these ionograms is the "backscatter leading edge," and is generally the only useful portion of the backscatter ionogram since all the other returns cannot be distinguished from one another, except in the case of high resolution ionograms.

It can now be seen from Fig. 1(b) that points along the backscatter leading edge correspond to rays reflecting at continuously increasing distant locations from the transmitter. Hence the backscatter leading edge contains information concerning horizontal gradients of electron density outward from the transmitter location along the azimuthal direction corresponding to the backscatter ionogram. It is the purpose of this paper to discuss methods of obtaining such information. To do this, we consider two ionospheric models: (a) the quasi-parabolic (QP) layer with locally uniform parameters, and (b) the locally quasi-parabolic (LQP) layer with horizontal gradients. In the following section, we shall discuss each model and the corresponding inversion technique and present examples of inversion.

2. QUASI-PARABOLIC (QP) LAYER

The quasi-parabolic layer is defined by the variation of electron density N_e with the radial distance r from the center of the earth as given by

$$N_e(r) = \begin{cases} N_m\left[1 - \left(\frac{r-r_m}{y_m}\right)^2\left(\frac{r_b}{r}\right)^2\right] & \text{for } r_b < r < r_m \frac{r_b}{r_b - y_m} \\ \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where

$N_m = f_c^2/80.6$ = maximum value of electron density

f_c = critical frequency

r_m = value of r at which N_e is equal to N_m

r_b = value of r at the base of the layer

$y_m = r_m - r_b$ = semithickness of the layer

The slight modification over the parabolic model, which is defined by ignoring the factor $(r_b/r)^2$ in (1), enables the derivation of exact closed form expressions for the ray path parameters for the quasi-parabolic layer (Croft and Hoogasian, 1968), by application of Bouguer's rule for ray tracing in a spherically symmetric layer with no magnetic field. For a signal of frequency f and elevation angle of transmission β, the expression for the group path P' is given as follows (Rao, 1974):

$$P' = 2\left(1 - \frac{F^2}{A}\right) r_b \sin\gamma - 2r_0 \sin\beta - \frac{BF}{2A^{3/2}} \ln \frac{U}{V} \tag{2}$$

where

$F = f/f_c$

$\gamma = \cos^{-1}\left(\frac{r_0}{r_b}\cos\beta\right)$ = elevation angle at the base of the layer

r_0 = radius of the earth

$U = B^2 - 4AC$

$V = 2Ar_b + B + 2r_b F\sqrt{A}\sin\gamma$

$W = 2\sqrt{C}\, Fr_b \sin\gamma + 2C + Br_b$

$A = F^2 - 1 + (r_b/y_m)^2$

$B = -2r_m r_b^2/y_m^2$

$C = \left(\frac{r_b r_m}{y_m}\right)^2 - F^2 r_0^2 \cos^2\beta$

To find the minimum group path, which is equal to the minimum time delay times the velocity of light in free space, we first note that for a given frequency and for a given set of layer parameters, the elevation angle of transmission corresponding to the minimum group path ray is given by the solution of equation

$$\frac{\partial P'}{\partial \beta} = 2\left(1 - \frac{F^2}{A} + \frac{BF^2}{AV}\right) r_o \sin\beta \cdot \cos\gamma$$

$$- 2\left(1 - \frac{2B}{\sqrt{A}}\frac{F^3}{U}\, r_o \sin\beta\right) r_o \cos\beta$$

$$= 0 \tag{3}$$

Recognizing that $\partial P'/\partial\beta$ varies continuously from a large negative value near zero elevation angle to a large positive value near maximum elevation angle corresponding to penetration of the ray through the layer. we can solve (3) for β in an iterative manner. The penetration condition occurs for the ray apogee radius equal to $- B/2A$ and hence for the elevation angle of transmission given by

$$\beta_p = \cos^{-1}\left\{\frac{1}{Fr_o}\left[\left(\frac{r_b r_m}{y_m}\right)^2 - \frac{B^2}{4A}\right]^{\frac{1}{2}}\right\} \tag{4}$$

Thus starting with values of β near zero and β_p, the value of β for which $\partial P'/\partial\beta$ is a small specified value can be found in an iterative manner. Substitution of this value of β in (2) then gives the value of the minimum group path, P'_{min}.

3. INVERSION FOR QP LAYER

3.1. Basic Technique

In earlier work of backscatter leading edge inversion, the QP layer model has been employed by Hatfield (1970) for obtaining the critical frequency, with the other two parameters having the assumed values, and by Rao (1974) for obtaining all three parameters from three data points on the leading edge. The method we employ here is a modification of that by Rao (1974).

The modification involves the minimization of the following sum-squared error function:

$$E\,(r_b,\, r_m,\, f_c) = \sum_{i=1}^{3} \left[P'_{c_i}\,(r_b,\, r_m,\, f_c,\, f_i) - P'_i\right]^2 \tag{5}$$

where P'_{ci} are computed minimum group paths corresponding to f_i and for a given set of values of r_b, r_m, and f_c, and P'_i are the actual minimum group hath values from the leading edge. Fig. 2 dipicts the strategy graphically. Assume that $(f_1,\, P'_1)$, $(f_2,\, P'_2)$, and $(f_3,\, P'_3)$ are the points we have chosen from the backscatter leading edge. We start the procedure with an initial set of layer parameters, which are denoted by $(r_b,\, r_m,\, f_c)_0$. Corresponding to these starting parameters is a unique backscatter ionogram leading edge trace $[P'_c\,(f)]_0$. We note that the error function E is simply the sum of the squared distances between the ordinates of the $[P'_c\,(f)]_0$ curve corresponding to the frequencies f_1, f_2, and f_3. and the chosen backscatter ionogram minimum group paths P'_1, P'_2, and P'_3. As r_b, r_m, and f_c are varied to minimize E, we see that we are effectively fitting the $P'_c(f)$ curve to the three data points in the least-squared error sense. If the data points are from a synthetic curve, computed by assuming the Q-P model, the error E will then be zero.

The error E is in general a nonlinear function of r_b, r_m, and f_c and an iterative method is required to seek its minimum. Numerical minimization techniques are available which seek a local minimum for a given starting point in parameter space. The nonlinear minimization alogorithm employed in this study is described by Fletcher and Powell (1963) and was adapted for use with our computer. The Fletcher-Powell method performs, for each iteration step, a linear minimization along a direction determined by the current gradient and an updated estimate of the Hessian. The size of the steps taken through the parameter space as the minimum is sought is proportional to the difference between the current value of the function to be minimized (E), and the user-supplied minimum value, typically taken to be 0.001 km^2, and inversely proportional to the current magnitude of the function gradient ($|\vec{\nabla}E|$). In this manner, rapid convergence is assured. The minimization procedure is terminated when the function value has not changed by more than a user-specified value (EPS), or if $|\vec{\nabla}E|$ has become less than EPS.

3.2. Inversion for Three-Dimensional Ionosphere

Thus far we have discussed the procedure for the inversion of three data points on a backscatter leading edge to Q-P layer parameters. Since each point on the leading edge corresponds to a different ground range, it is obvious that the three points must be chosen fairly close together if horizontal gradients are present. The three data points must corespond to three ionospheric propagation paths which are in such horizontal proximity as to validate the assumption of a locally horizontally uniform ionosphere. These considerations give rise to a method of deriving the horizontal ionization gradients from the leading edge. Three closely spaced points on the leading edge just above the critical frequency at the backscatter sounder site can be used to determine the layer parameters of the ionosphere at a range which is in close proximity to the site. Next, three closely spaced points further up the leading edge can be used to obtain the layer parameters at a greater range than that corresponding to the previous set of data points. In obtaining these layer parameters, the solution for the first set of three data points can be used as the starting solution. A repetition of this process continuously along the leading edge for successive sets of three data points yields the layer parameters as a function of distance away from the sounder site. When the leading edges are available for a number of azimuthal directions away from the sounder, it is then possible to obtain a three dimensional distribution of electron density made up of locally uniform QP layers.

3.3. Example

To test the inversion technique, backscatter leading edges are generated for two azimuths (330° and 358°) from a common location by three-dimensional ray tracing in a mid-latitude ionospheric model. The generated minimum group path values versus frequency are shown in Table 1. To invert these leading edge data, a QP layer is fitted to the actual vertical electron density profile existing above the common (transmitter) location. By using these QP layer parameters as the starting solution, overlapping sets of three data points along the leading edge are inverted for QP layer parameters. In inverting a given set other than the first one, the solution found for the previous set is used as the starting solution. For each set of data inverted, the corresponding ground range is computed and the QP layer parameters are assigned to a location half that distance from the transmitter and along the pertinent azimuth. The layer parameters and the corresponding ranges generated in this manner are listed in Tables 2 and 3

To compare the applicability of the deduced three-dimensional ionosphere by inversion of the leading edge data relative to that of the originally assumed three-dimensional model, rays are now traced at one frequency (16 MHz) in the original as well as the computed ionospheres for an azimuth of 340° (lying between 330° and 358° corresponding to the inverted backscatter data) and for several elevation angles of transmission. For each ray, the group path (P'), the ground range (R) and the quantity (P'-R) are computed. The quantity (P'-R) is shown plotted versus P' for the two cases in Fig. 3. A comparison of the two curves indicates that the computed ionosphere gives the correct range to within ±2 km for a wide range of group path values. The discrepancies greater than ±2 km toward the low end of the curves are attributable to the fact that the region is close to the minimum group path for 16 MHz, whereas the discrepancies greater than ±2 km toward the high end of the curves are attributable to the fact that the region is beyond the range corresponding to the inverted backscatter data.

4. THE LOCALLY QUASI-PARABOLIC (LQP) LAYER WITH HORIZONTAL GRADIENTS

To extend the backscatter leading edge inversion technique for a layer with horizontal gradients in electron density, we now consider a quasi-parabolic layer with its parameters varying in a specified manner with distance away from the sounder location. Thus the electron density distribution is given by

$$N_e(r,\theta) = N_m(\theta)\left\{1 - \left[\frac{r-r_m(\theta)}{y_m(\theta)}\right]^2 \left[\frac{r_b(\theta)}{r}\right]^2\right\} \tag{6}$$

where r is the radial distance from the center of the earth and θ is the angular displacement away from the sounder location. Since for a fixed value of θ, eq.(6) has the form of an ordinary quasi-parabolic layer given by (1), we call this model the locally quasi-parabolic (LQP) layer model.

In view of the horizontal gradients, the electron density distribution is no longer spherically symmetric and hence the ray path parameters cannot be obtained in closed form as in Sec.2. It is necessary to resort to numerical integration of the ray equations. These ray equations are usually written in a form referred to as the Haselgrove equations (Haselgrove, 1957; Yeh and Liu, 1972). In the present case, however, the problem is two dimensional. Furthermore, neglecting the earth's magnetic field and following the notation of Yeh and Liu (1972), the pertinent ray equations may be written as

$$\frac{dr}{d\tau} = \frac{\sigma_r}{n^2} \tag{7}$$

$$\frac{d\theta}{d\tau} = \frac{\sigma_\theta}{rn^2} \tag{8}$$

$$\frac{d\sigma_r}{d\tau} = \frac{1}{n}\frac{\partial n}{\partial r} + \sigma_\theta\frac{d\theta}{d\tau} = \frac{1}{n}\frac{\partial n}{\partial r} + \frac{\sigma_\theta^2}{rn^2} \tag{9}$$

$$\frac{d\sigma_\theta}{d\tau} = \frac{1}{rn}\frac{\partial n}{\partial\theta} - \frac{\sigma_\theta}{r}\frac{dr}{d\tau} = \frac{1}{rn}\frac{\partial n}{\partial\theta} - \frac{\sigma_\theta\sigma_r}{rn^2} \tag{10}$$

with

$$n = \text{phase refractive index}$$

$$\sigma_r = n\hat{k}\cdot\hat{r}$$

$$\sigma_\theta = n\hat{k}\cdot\hat{\theta}$$

where $\hat{k}$ is the unit vector in the wave normal direction and $d\tau$ represents the differential element of phase path. Two additional equations have also been used in order to allow us to keep track of the cumulative group and phase paths as the ray progresses. These additional equations are:

$$\frac{dP}{d\tau} = 1 \tag{11}$$

and

$$\frac{dP'}{d\tau} = \frac{1}{n^2} \tag{12}$$

with the first equation representing the phase path and the second equation representing the group path.

The set of initial conditions associated with equations (7) through (12) is based upon the assumption that there is a free space region (n=1) between the surface of the Earth (source location) and the base of the ionosphere. Thus, the initial conditions are:

$$r = 6370 \text{ km} = \text{radius of Earth} \tag{13}$$

$$\theta = 0° \text{ (selection of source location as angular reference)} \tag{14}$$

$$\sigma_r = \sin\phi_0 \; (\phi_0 = \text{initial ray elevation angle}) \tag{15}$$

$$\sigma_\theta = \cos\phi_0 \tag{16}$$

$$P = P' = 0 \tag{17}$$

The net result of using these initial conditions in connection with a given model for the phase refractive index, n, is that the only remaining degree of freedom for a ray of some specified frequency is provided by the initial elevation angle of the ray (ϕ_0) and consequently the value of group path, P', which is computed when the ray has left the source, reflected from the ionosphere, and returned back to the surface of the Earth is determined by ϕ_0. Of course, certain choices of ϕ_0 may result in the ray penetrating the ionosphere. Nevertheless, for those values of ϕ_0 which result in rays returning to the surface of the Earth there will be some value of ϕ_0 which results in a minimum group path.

5. INVERSION FOR THE LQP LAYER

5.1. The Ionostate Vector

The technique of inversion for the LQP layer is similar to that discussed in Sec.3.1. We however need to assume some form of dependence of the QP layer parameters on θ. The simplest form consists of a nominal value which is given by the value of the parameter directly above the source (θ=0) plus a constant gradient in the θ direction multiplied by the angular displacement away from the source [e.g. $r_b(\theta) = r_{bo} + \frac{dr_b}{d\theta} \cdot \theta$]. Since there are three parameters, r_{bo}, r_{mo}, and f_{co}, as well as three gradients (one in each parameter), there is then a total of six parameters needed to describe the LQP model. These six parameters may be considered to define a basis for a six dimensional space in which a simultaneous set of the six parameters may be thought of as constituting a vector which we may define as:

$$\vec{\xi} = [\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6] \tag{18}$$

with

$$\xi_1 = r_{bo}$$

$$\xi_2 = r_{mo}$$

$$\xi_3 = f_{co}$$

$$\xi_4 = dr_b/d\theta$$

$$\xi_5 = dr_m/d\theta$$

$$\xi_6 = df_c/d\theta$$

We shall call the vector (18) the "ionostate" vector or simply the ionostate.

Having defined the ionostate, the problem, stated in its simplest terms, becomes how to determine the six components of the ionostate in some optimum sense. The criterion which we have selected is essentially the same that given by (5). Thus the backscatter leading edge which would be observed if the ionosphere were LQP, should match as closely as possible the backscatter leading edge as experimentally measured by the ionospheric sounder operating at the same set of frequencies and observing the actual ionosphere. More precisely, if we denote the set of sounding frequencies by f_i (for i=1 to N) and denote the measured backscatter leading edge by $P'(f_i)$, while denoting the LQP model backscatter leading edge as $\hat{P}'(\vec{\xi},f_i)$, the problem may be stated as:

Find $\vec{\xi}$ such that

$$e^2(\vec{\xi}) \equiv \frac{1}{N} \sum_{i=1}^{N} [P'(f_i)-\hat{P}'(\vec{\xi},f_i)]^2 \tag{19}$$

is a minimum.

5.2. Examples

To test the inversion technique, we assumed that the ionosphere was locally quasi-parabolic and was completely specified by the ionostate, ξ_0 where

$$\begin{aligned}\vec{\xi}_0 &= [\xi_1,\ \xi_2,\ \xi_3,\ \xi_4,\ \xi_5,\ \xi_6] \\ &= [6570\text{ km},\ 6720\text{ km},\ 5\text{ MHz},\ 0\text{ km/rad},\ 0\text{ km/rad},\ 2\text{MHz/rad.}]\end{aligned}$$

Based upon this $\vec{\xi}_0$, we then computed the minimum group paths, corresponding to a backscatter ionogram at discrete sounding frequencies. We then "estimated" an initial ionostate and using the mean squared error e^2, defined in equation (19) as a criterion of optimality, we attempted to reconstruct the original ionostate, $\vec{\xi}_0$.

As an example, the cases illustrated in Figure 4 result from specifying the ξ_2, ξ_3, ξ_4, and ξ_5 components of $\vec{\xi}$ as "fixed", thereby reducing the problem to a two dimensional minimization. In all these cases the fixed components were set to their correct values (i.e. the values of the corresponding components of $\vec{\xi}_0$). Using three different sets of starting values (marked by x's in Figure 4) for the remaining components (ξ_1 and ξ_6), the program was allowed to continue in an iterative manner until the mean squared error became less than 1 km². The underlined numbers in this figure represent the mean squared error, and the dotted curve corresponds roughly to the contour curve along which the mean squared error is approximately 7. A somewhat more complete set of contour curves is shown in Figure 5. As can be seen in comparing Figures 4 and 5, the trajectory of the point (ξ_1, ξ_6) proceeds primarily along the direction of a narrow valley in the mean squared error surface. Figure 6 shows a similar convergence when ξ_2 and ξ_6 are used as variable ionostate components.

6. DISCUSSION AND CONCLUSION

In conclusion, we have in this paper discussed techniques of inverting the leading edge of a backscatter ionogram for ionospheric structure and illustrated them by means of examples involving synthetic data. The inversion for locally uniform QP layer parameters when used with successive sets of three data points along the leading edges for various azimuths from the backscatter sounder was found to be capable of providing a three dimensional distribution. The deduced distribution in this manner was found to yield ground range for a given group path at a specified frequency to within ±2 km of the value corresponding to the originally assumed ionosphere for synthesizing the leading edges. This observation holds despite the fact that the minimum group path rays all reflect from below a certain apogee height and hence the vertical profiles obtained from inversion of the leading edges cannot be used for tracing rays reflecting above that height.

A simple explanation for the success of the technique in finding the range for a given group path to a good accuracy, despite the above mentioned limitation of the inversion technique can be given by referring to Fig. 7 which illustrates the fact that the backscatter ionogram is a continuum of oblique ionograms corresponding to successively increasing values of ground range away from the transmitter. Let us suppose we have found the range corresponding to a particular point A (f_1, P_1') on the leading edge. Then a point B, corresponding to the same group path as that of A but to a frequency $f_2 < f_1$, corresponds to a range greater than that of A. Alternatively, it corresponds to a range corresponding to that of a point C on the leading edge, lying above the point A. Hence the ray of frequency f_2 and group path P_1' reflects below the apogee height corresponding to the point C. Since this argument is true for any point on the low-ray portions of the oblique ionograms, it follows that one can always find the range corresponding to such a point, although the backscatter leading edge is not capable of yielding the ionospheric profile above the apogee heights corresponding to the minimum group path rays.

The inversion for LQP layer parameters was found to yield results which are in good agreement with the originally assumed parameters. However, the convergence of the solution along a narrow valley in the mean squared error surface, as shown for example in Fig. 5, indicates that it is important to start the inversion process with a good estimate of the LQP layer parameters. Such an estimate can be obtained by first inverting the leading edge for locally uniform layer parameters along the azimuthal direction from the sounder and then using those parameters to obtain the necessary gradients.

ACKNOWLEDGEMENT

The work described in this paper was supported by the Rome Air Development Center Deputy for Electronic Technology.

REFERENCES

Croft, T. A. and H. Hoogasian, 1968, "Exact ray calculation in quasi-parabolic ionosphere with no magnetic field," Radio Sci., 3, 69-74.

Fletcher, R. and M. J. D. Powell, 1963, "A rapidly convergent descent method for minimization," Computer Journal, 6, 163-168.

Haselegrove, J., 1957, "Oblique ray paths in the ionosphere," Proc. Phys. Soc. London, Sec B70, 653-662.

Hatfield, V. E., 1970, "Derivation of ionospheric parameters from backscatter data," In Ionospheric Forecasting, edited by V. Agy, AGARD Conf. Proc., 49, paper 16.

Rao, N. N., 1974, "Inversion of sweep-frequency sky wave backscatter leading edge for quasiparabolic ionospheric layer parameters," Radio Sci., 9, 845-847.

Yeh, K. C. and C. H. Liu, 1972, "Theory of ionospheric waves," Academic Press, New York, p. 240.

Table 1. Computed minimum group path data for two azimuths from a common location.

f, MHz	P'_{min}, km, for azimuth of 330°	358°
10	798.35	805.04
11	884.79	894.16
12	972.60	984.55
13	1062.92	1077.68
14	1155.56	1173.13
15	1251.66	1273.45
16	1349.12	1376.13
17	1450.55	1483.70
18	1535.25	1596.70
19	1665.4[illegible]	1714.96
20	1780.63	

Table 2. Computed QP layer parameter values and the corresponding ground ranges for 330° azimuth

Range, km	f_c, MHz	r_b, km	r_m, km
327.56	9.2406	6556.4	6676.2
377.18	9.3493	6561.9	6679.5
431.21	9.3344	6561.2	6679.1
486.87	9.2511	6557.4	6676.4
531.07	9.7489	6565.8	6694.9
577.99	10.0895	6573.9	6707.4
632.21	10.1063	6574.6	6708.0
689.80	10.0427	6572.1	6705.8
749.4[illegible]	9.9592	6568.2	6702.9

Table 3. Same as Table 2 except for 358° azimuth

Range, km	f_c, MHz	r_b, km	r_m, km
373.28	8.3981	6509.2	6661.6
414.73	8.9054	6519.1	6680.1
461.30	9.0253	6526.2	6681.9
515.7[illegible]	8.9677	6522.8	6680.9
568.43	8.9757	6523.2	6681.1
619.54	9.0693	6529.0	6682.8
676.62	9.0482	6527.7	6682.4
732.96	9.2278	6528.4	6690.8

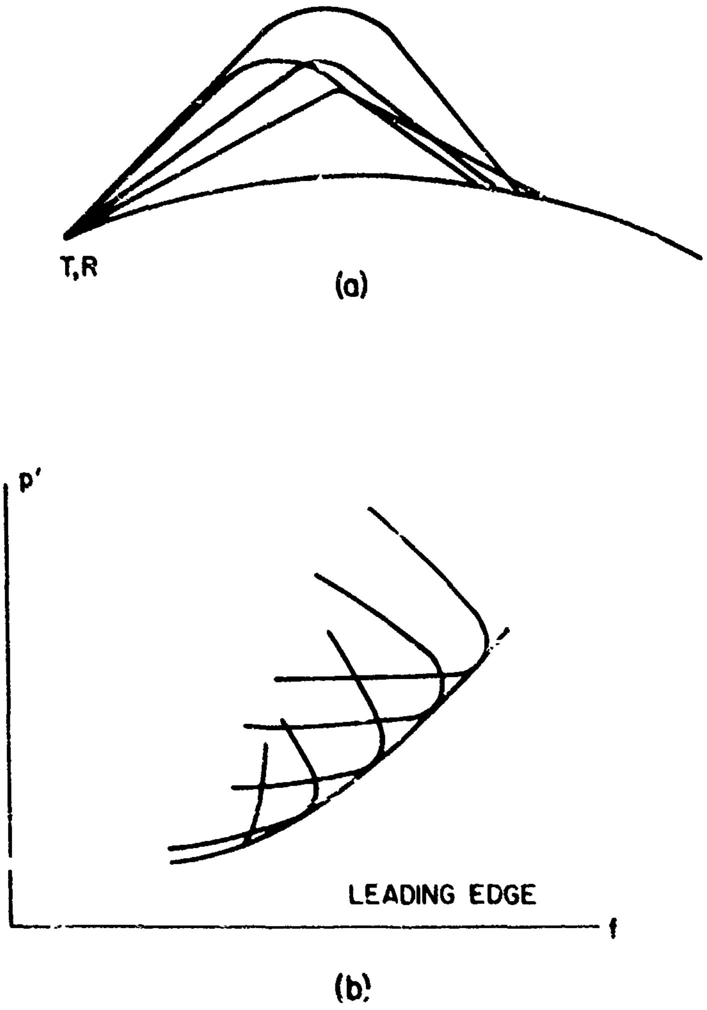


Figure 1. (a) Ray paths for backscattered rays at one frequency
(b) Showing that the leading edge of a backscatter ionogram is the tangent curve to a continuum of point-to-point oblique ionograms.

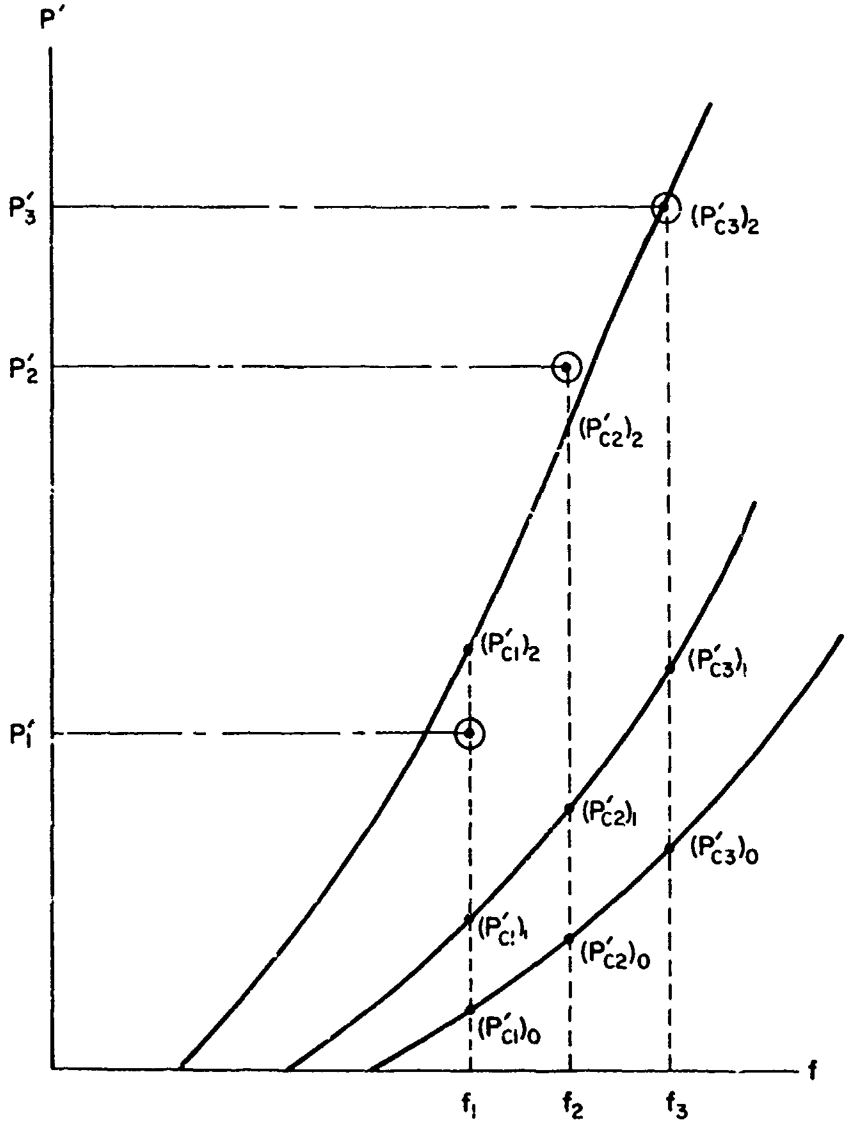

P′
P′3
P′2
P′1
(P′C3)2
(P′C2)2
(P′C1)2
(P′C3)1
(P′C2)1
(P′C1)1
(P′C3)0
(P′C2)0
(P′C1)0
f
f1
f2
f3

Figure 2. Graphical illustration of the least sum-squared error method of backscatter ionogram leading-edge inversion for a Q-P ionosphere.

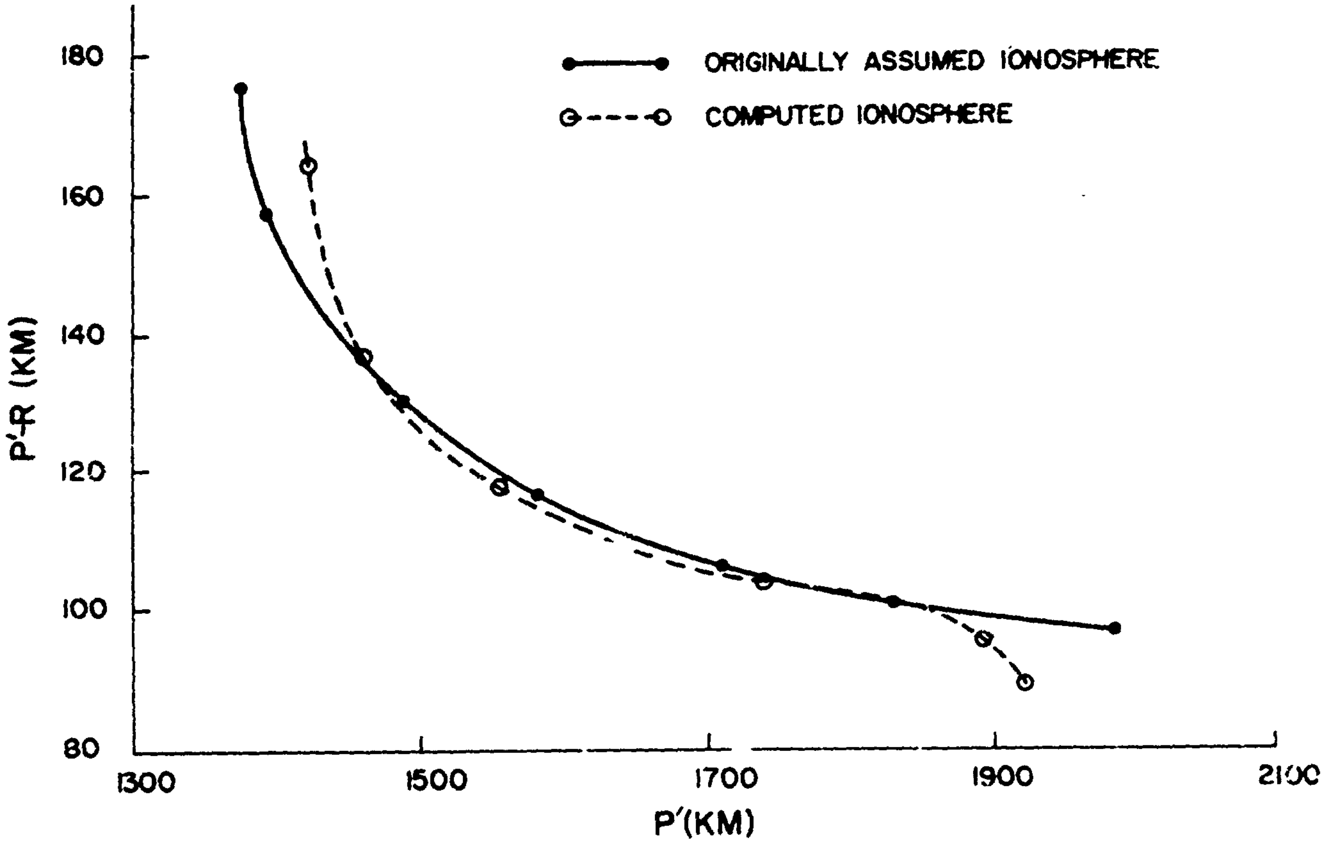


Figure 3. For comparing the reconstruction of ground range from group path for originally assumed and computed ionospheres.

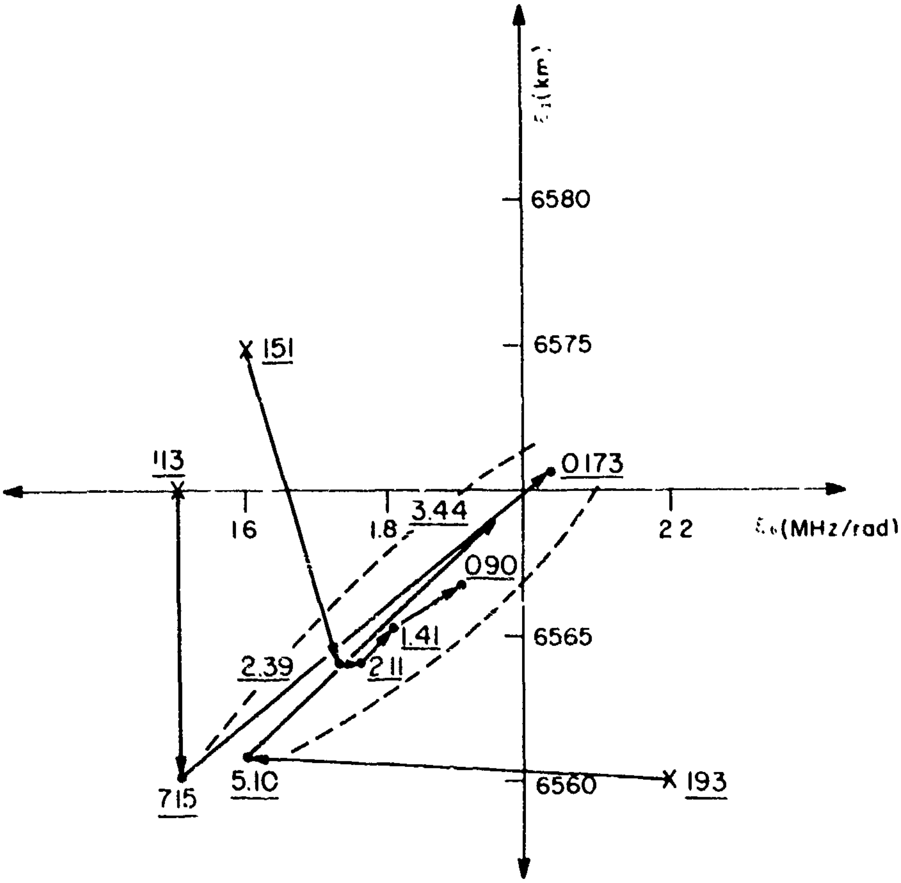


Figure 4. Two component (f_c and r_b) convergence of the mean square error.

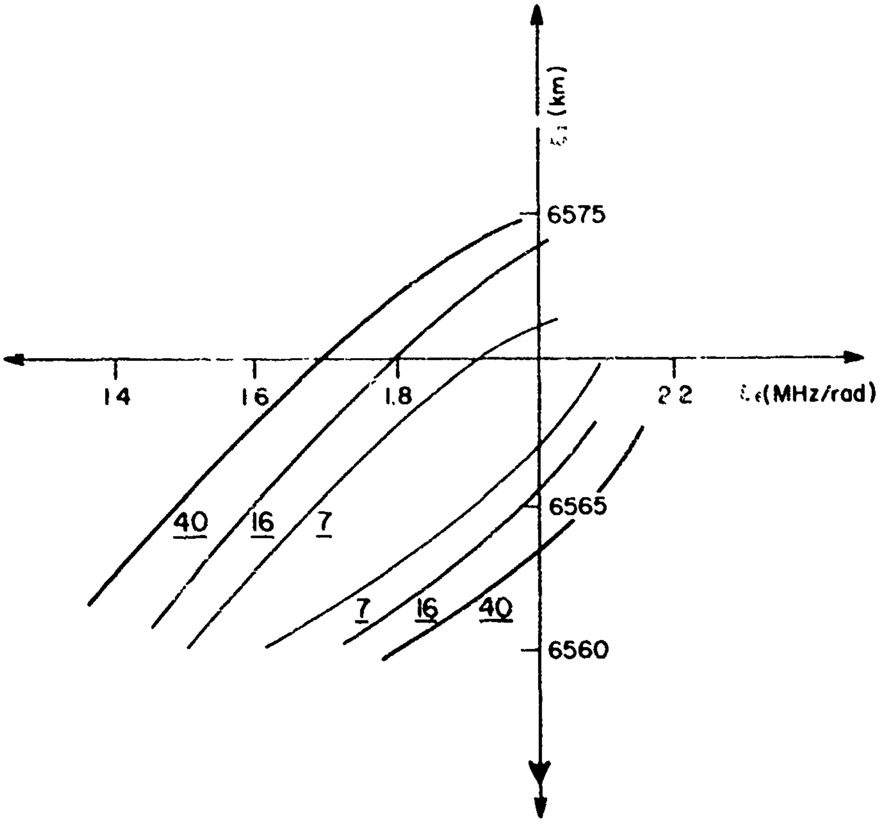


Figure 5. Level curves of mean square error surface.

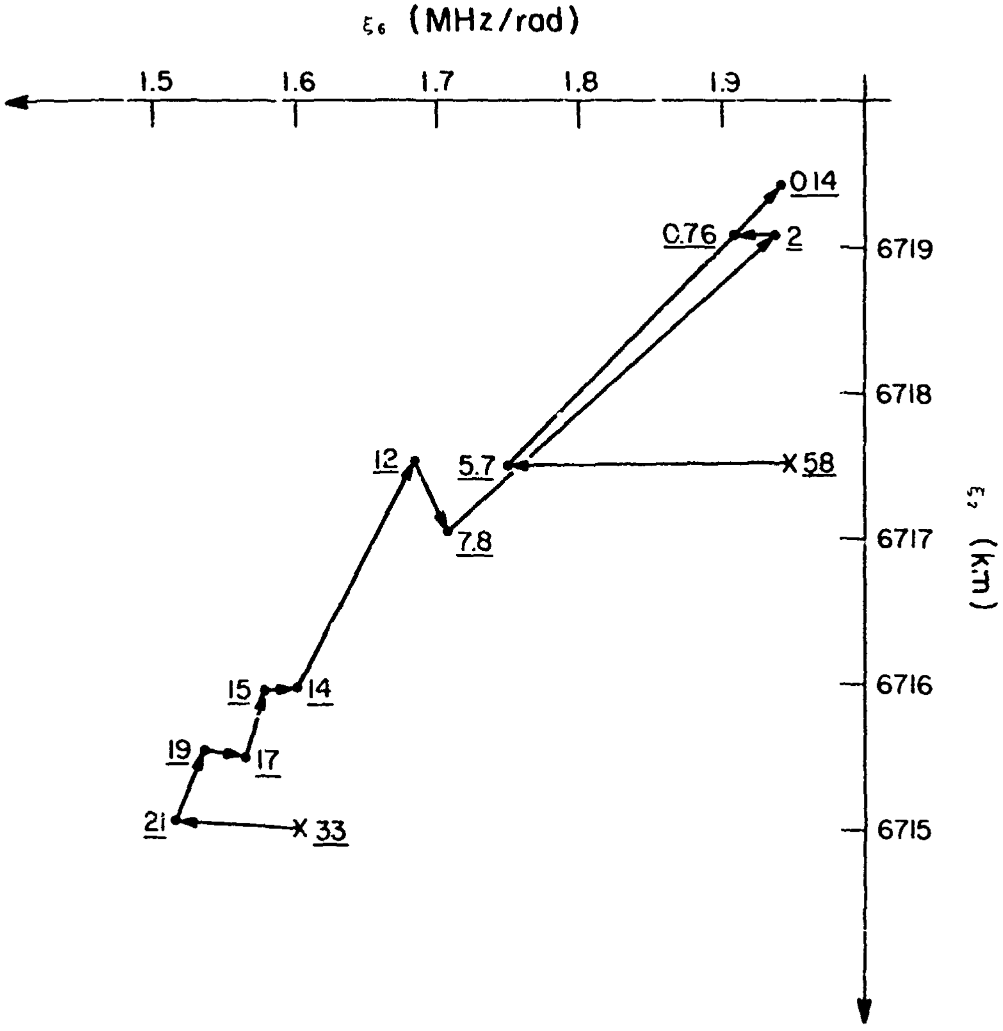


Figure 6. Two component (f_c and r_m) convergence of the mean square error.

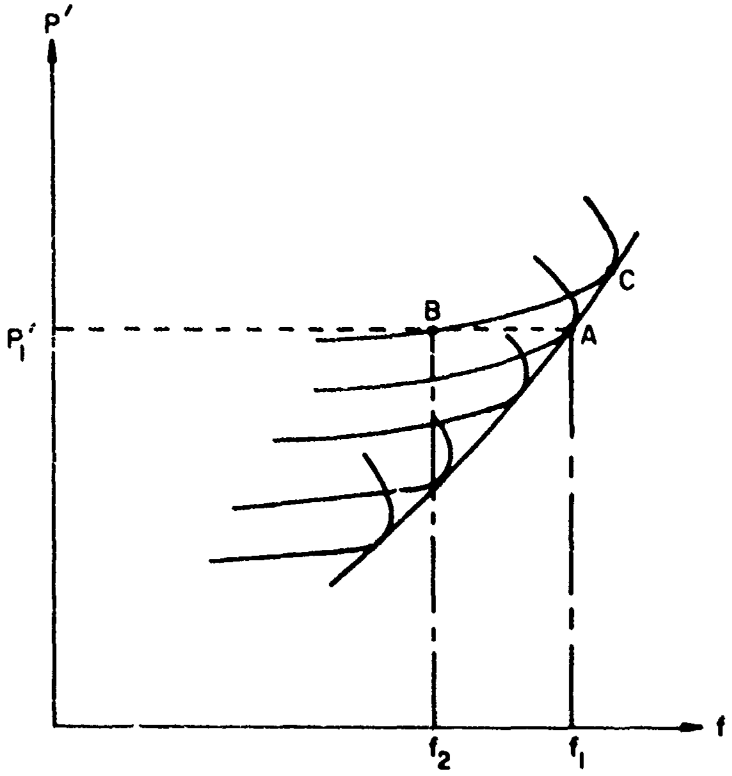


Figure 7. For explaining the success of the backscatter leading edge inversion technique for finding the ground range for a given group path, despite the limitations of the backscatter data.

DISCUSSION

C.S. Goutelard, Fr

Cet exposé est très intéressant et je voudrais faire deux commentaires. Tout d'abord en ce qui concerne le procede de calcul, il n'apparait pas evident que le calcul itératif converge dans tous les cas. Nous avons étudié ce problème d'inversion par des méthodes similaires à celle présentée ici et nous avons traite plusieurs centaines de cas sur lesquels on a pu déterminer des zones où le système de calcul ne converge pas. De plus, certains paramètres influent peu sur le temps de focalisation, et leur determination se fait avec une grande imprecision. L'orateur pourrait-il préciser s'il a étudié ces points et donner ces conclusions.

Le second point porte sur l'utilisation de cette méthode dans des systèmes. Il aurait éte intéressant d'etudier l'influence de la précision des mesures sur la précision du modele déduit. En effet, les resultats presentes partent de données theoriques ou les imprécisions des mesures n apparaissent pas. Nous avons constate que des erreurs même faibles sur la mesure du temps de focalisation entraine de grandes erreurs sur certains paramètres du modèle. L'orateur pourrait-il commenter l'influence de la précision des mesures sur la precision du modèle qu'il construit?

Author's Reply

In our study we have chosen the initial "ionostate" reasonably close to the true ionostate. Of ten cases we have studied fast convergence has been found. Of course, one can imagine cases for which the initial ionostate may be sufficiently far away from the true ionostate so the the convergence is not achieved. This can be especially troublesome when working with real data. Even when working with synthesized data we have found that the level curves of the error surface can form very long narrow valleys. This can come about because of the compensating effects between two or more components of the ionostate. For example, an increase in critical frequency will reduce the group delay and this reduction can be compensated by raising the ionospheric layer height. Even though it results in a different ionostate when used to predict the ground distance of a propagation path these different ionostates will predict ground distances that do not deviate very much as shown in the paper.

B.W. Reinisch, US

Are you implying that there is a unique solution to the problem and does your iteration procedure converge toward this solution when you have more than 2 unknown parameters?

Author's Reply

Since the group delay is a non-linear function of the ionization profile the question of uniqueness is still unanswered mathematically at present. However, for about ten cases we have studied the convergence is rather rapid for up to 4 parameters.

THE GEOMORPHOLOGY OF THE HF BREAKTHROUGH PHENOMENON

John M. Goodman
Communication Sciences Division
Naval Research Laboratory
Washington, D. C., 20375 USA

SUMMARY

The phenomenon of HF radio "breakthrough" to the topside ionosphere is briefly reviewed with emphasis upon that information about the ionospheric personality which may be deduced from satellite measurements of HF noise and/or signals. The overall effort calls for parameterization of the ionospheric escape cone (iris) in terms of specified ionospheric conditions as represented by a convenient model which depicts the salient geomorphological features. Included among these features are the midlatitiude trough, the Appleton anomaly, and the day-night terminator. A general three-dimensional raytrace routine is utilized to demonstrate selected results. A case is made for a general environmental assessment system using topside data, either active or passive. This concept is especially viable for the totally uncooperative link. Problems in the utilization of the HF breakthrough phenomenon over oceanic areas are discussed and the potential of acti topside sounding to solve the Navy problem is noted.

1. INTRODUCTION

The phenomenon of high frequency (HF:3-30 MHz) radiowave "breakthrough" has been reported by a number of experimenters (Atkins and Chapman, 1963; Nelms, 1964; Huguenin and Papagiannis, 1965) and various propagation factors pertaining to the effect have been discussed by several workers including Rawer, (1967) and Maliphant, (1967). Currently the U. S. Air Force is operating a passive ionospheric monitor experiment on board the DMSP-F2 satellite which utilizes terrestrial HF noise to extract estimates of foF2 beneath the satellite (Thompson, 1978), and preliminary results have been discussed by Rush et al, (1978), Rush and Buchau, (1977) and Rush and Ziemba, (1978).

A number of satellites placed in orbits above the nominal F2 maximum of ionization have included experimental packages for measurement or indication of the breakthrough of terrestrial high frequency radiation. Generally, the breakthrough phenomenon is observed at frequencies in excess of the sub-satellite foF2. There is some indication, however, that frequencies below foF2 may reach a satellite via various modes of anomalous propagation (Rush and Ziembe, 1978; Hagg et al, 1969). This anomalous effect, although of interest, is not detailed herein. Among the satellites observing HF breakthrough are TRANSIT 2A, RAE-1, ARIEL III, DMSP-F2, and the topside sounders Alouette I/II and ISIS 2. Other satellite HF experiments, including those flown on TOPSI and ORBIS, have also provided useful information; these experiments have used the inverse procedure of radiating a fixed HF frequency which was monitored by a ground station.

The first extensive attempt to determine the sub-satellite foF2 from HF noise (passive) measurements was made by Rush and Buchau, (1977). ISIS-2 topside ionograms were utilized in conjunction with an AGC voltage read-out to infer increased noise levels above foF2. The authors found that over regions of the earth where the density of HF noise sources was high and the ionosphere did not possess strong gradients, measurements of breakthrough could determine foF2 to within 1 MHz over 80% of the time. Over regions where there were few sources of terrestrial noise, the procedure was unreliable as expected. Strong gradients also lowered the confidence of the results. Recent work by Rush and Ziemba, (1978) indicated that passive determination of foF2 was much more accurate than a model determination at night; during the day the passive technique was no better than use of a model. Control was provided by comparison with ground-based ionosonde at Ottowa, Canada. The poor daytime accuracy was attributed, at least in part, to reduction in the number of sources caused by increased daytime absorption. The first satellite experimental package designed specifically to measure foF2 using HF breakthrough was the passive ionospheric monitor (SSIP) aboard DMSP-F2. Although the package is not working as well as designed due to possible coupling problems, there is some potential for extraction of useful information (Thompson and Patterson, 1979). The U. S. Air Force has also developed a plan for orbiting an active topside sounder (SSI) in its DMSP series F8 through F11 (Vette et al, 1978) as a follow-up to the SSIP experiments.

The Naval Research Laboratory is developing a plan for use of topside measurements such as those obtained from the SSJ and SSIP packages aboard DMSP. Specific interest is directed toward development of operational scenarios for climatological update of ionospheric paramenters over oceanic areas for use in HF communication assessment and possible improvement in the allocation of signal intercept resources. Other benefits which might accrue from an active topside sounder program include oceanic measurements of topside spread-F which is at least circumstantially related to satellite scintillation effects. Thus an operational topside sounder program would aid in the selection of the proper mix of HF and SATCOM resources for achieving connectivity between the major Communication Area Master Station (CAMS) and other Naval Communication Stations (NAVCOMSTAS) and the Fleet. In the process of analyzing the possible uses of topside data to assist in various Naval mission areas, it was necessary to examine the HF breakthrough phenomenon. The present paper describes some preliminary work on this topic.

A central purpose of this paper is to examine the potential for Naval force utilization of HF noise monitoring equipment in spacecraft. The issues in HF link assessment/prediction for the Navy context may be considered to be contained within three requirements. First, there is the requirement for connectivity between communication stations and various fleet assets over oceanic areas (doubly cooperative links). In this case both ends of the link are known, can be controlled, and both may have acess to enviromental data directly. These links are most amenable to adaptive system design. Secondly, there is the requirement for link vulnerability analysis from a fixed (known) terminal to arbitrary (unspecified) terminals covering a wide range of azimuths and path links (singly-cooperative). Finally, there is the possible requirement for assessment over completely arbitrary links (uncooperative). In this case the terminals are unspecified a priori Figure 1 illustrates the degree of competitiveness of forecasting systems and quasi-adaptive systems as a function of the degree of link cooperativeness. It is clear that doubly cooperaive links require the least quantity of enviromental data input and the uncooperative link requires the most.

2. PLAN OF ATTACK

The preliminary effort undertaken, and reported herein, relies primarily upon the utilization of a three-dimensional raytrace program in conjunction with a reasonable ionospheric model which includes all of the usual macroscopic features of the geomorphology. Such a capability has been provided to us courtesy of Thomason et al, (1979). The raytrace program was developed by Jones and Stephenson, (1975) and the global ionospheric model was developed jointly between NRL and the Institute for Telecommunication Sciences (ITS) at Boulder, Colorado. Additional work was carried out using a code developed at NRL for analysis of high altitude nuclear effects upon radio waves (Mead and Wagner, 1971).

The plan of attack was to characterize the ionospheric "iris" as a function of solar activity (i.e., sunspot number, SSN), magnetic activity (i.e., planetary K-index, Kp), season, latitude, longitude, time of day, transmission frequency and mode of propagation (i.e., O or X). The characterization parameters include contours of fixed launch elevation through the "iris" projected to the satellite height and the displacement of the centroid of the iris contours with respect to the zenithal ray as well as the zenith itself. A complete treatment would also include annotation of D and E region absorption for each iris characterization.

To accomplish the goals of this study, it is clearly required to restrict the range of values assumed for each variable. Even so, a considerable amount of computer time is needed if the iris is to be characterized fully. To remedy this situation it was decided to probe the matrix of variables using only a single ray which might be indicative of situations for which more detailed calculations would be warranted. The ray selected was the zenithal ray (i.e., $\chi = 0^0$) which "ran" the fastest on the NRL advanced scientific computer. The minimum output data required from each raytrace was down-range displacement from the position of ray launch, the bearing (azimuth) of the ray, and the absorption level. The down-range ray displacement and its direction are indicative of the magnitude and direction of any horizontal gradients which affect the ray. The importance of the absorption is obvious. As of this writing this prescription has not been fully realized.

3. THE IONOSPHERIC IRIS

Using Snell's law, it may be shown that terrestrial sources of HF noise or transmission will be "blocked" from reaching above the height of the F2 maximum for rays having zenith angles greater than some angle θ_0. This blocking effect is essentially limited to a zone of rays (or "iris") in which the half-angle θ_0 is

$$\theta_0 = \cos^{-1} \frac{f_p}{f} \qquad (1)$$

where f_p is the local overhead foF2 (to a good approximation) and f is the transmitted frequency. This expression assumes no anisotropy and a plane stratified ionosphere. Note that θ_0 defines an angle which increases as f approaches f_p from above. For $f \leq f_p$, θ_0 is undefined and total reflection occurs below foF2. Thus frequencies just slightly above the local plasma frequency will penetrate the F region but will be largely zenithal. As f increases substantially above the local overhead value of foF2, θ_0 increases to such as extent that localization of the terrestrial source of transmissions is severely limited. The best localization is obviously obtained where f is just slightly in excess of foF2. Table I indicates how θ_0 varies with f_p/f. Also tabulated is 300 tan θ_0 (i.e. R_{300}) which is the horizontal projection of the radius of the escape cone or iris at an altitude of 300 kilometers. Even in this first order treatment R_{300} is an underestimate due to earth curvature.

TABLE I

f_p/f	θ_0 (degrees)	R_{300} (kilometers)
.999	2.56	13
.995	5.73	30
.99	8.11	43
.95	18.19	99
.90	25.84	145
.85	31.79	186
.80	36.87	225
.75	41.41	264
.70	45.57	306
.65	49.46	351
.60	53.13	400
.55	56.63	455
.50	60.0	520

The reader will note that Maliphant (1967) has developed some useful nomograms for deducing the path of the rays which emerge from the ionosphere assuming spherical stratification. In this procedure, Maliphant uses of fixed chapman-like electron density profile shape and the ionospheric parameters which may be varied are hF2 and f_pF2. It is deemed to be a good tool for testing absurdities which may "crop-up" in raytracing using realistic models.

Figure 2 illustrates the geometrical extent of the ionospheric iris at 300 kilometers and at a satellite height of 834 kilometers assuming a sub-satellite plasma frequency of 6.2 MHz and a transmission (noise) frequency of 10 MHz. In any operational system using HF breakthrough as an indication of the plasma frequency f_p of the ionosphere, it is obvious that one must determine the greatest lower bound of noise breakthrough as accurately as possible. This restricts the leakage of unwanted noise by virtue of iris restriction as well as gives the more appropriate estimate of f_p. Situations such as those illustrated in Figure 2 are probably unacceptable. Taking 10 MHz as a reference frequency and following a typical DMSP orbit occurring during March at 1800 GMT, for a sunspot number of 100, we may construct the variation of the ionospheric iris along the track. The CCIR global atlas is utilized (1966). Results for every 5 minutes of satellite position are exhibited in Figure 3 assuming first order definition of the iris indicated by equation 1. Notice that the iris disappears at certain equatorial positions; at these points the plasma frequency is in excess of 10 MHz and noise penetration is disallowed.

4. IONOSPHERIC MODEL AND RAYTRACING CALCULATIONS

Having examined the first-order approaches to the determination of iris dimensions, it was deemed necessary to examine the distortions in the iris shape using more detailed ionospheric models and 3 dimensional raytracing.

The ionospheric model employed in this study is based upon the work of Thomason et al, (1979). Suffice it to say that it contains all of the usual macroscopic geomorphology of the ionosphere (D-region, E-region, E-F2 valley, F1 region, F2 region, topside). The topside description is a modified version of that developed by Bent, (1975). The F2 region critical frequencies provided by the ITS-78 "BLUE DECK" version (Jones and Obitts, 1970) of the basic program described by Barghausen et al (1969) are used but have been modified as appropriate through encorporation of certain polar corrections based upon magnetic activity (Miller and Gibbs, 1975). The raytracing routine is due to Jones and Stephenson (1975).

Figure 4 is a plot of the plasma frequency contours for March 1100 GMT between the equator along the 69° W Meridian, through the pole, and on to the opposite equator along the 111° E meridian. A sunspot number of 70 is assumed along with a Kp value of 3. The left-hand-side of the map refers to the dawn meridian and the right-hand-side refers to dusk. Figures 5A and 5B are raytracings, from the origin of Figure 4 assuming f = 10 MHz and initial launch azimuths (AZ) of 0° (North) and 270° (West) respectively. Notice that the initial breakthrough occurs for an elevation near 30° in each instance, provided we ignore the breakthrough due the "phantom" source down-range for the AZ = 270° case. Notice also that the 90° elevation rays in both cases are displaced down-range. We find that these displacements amount to 14.3 kilometers in range at a bearing measured counter-clockwise from North at ~ 59°. This implies an electron density gradient above the transmitter directed toward the ESE (specif: AZ ~ 121°). This is in general agreement with other model representations including CCIR for the same conditions. The absorption for the zenithal ray is found to be 1.16 dB.

Figure 6 is similar to figure 4 but for GMT = 0500. Thus it constitutes a representation of ray trajectories along the midnight-midday meridian. Notice that the gradient at the origin (i.e., the geographic equator at midnight) is toward the south. We have no E-W set of contours to determine if there is an additional zonal gradient but we shall see from the companion raytracings that it is negligible. Figures 7A and 7B show that the initial breakthrough is at a much higher elevation than in the former test, here being ~ 60° for AZ = 0° and ~ 65° for AZ = 270°. This is because of the higher overhead plasma frequency leading to a smaller value of $\theta_0 = \cos^{-1} f_0/f$ since f is fixed at 10 MHz. The zenithal ray is bent toward the north by 48 kilometers as reckoned at an altitude of 840 kilometers. Thus the 90° elevation ray exhibits a forward tilt for AZ = 0° but "appears" to be unrefracted for AZ = 270°. The absorption for the zenithal ray is 0.45 dB in this case.

Figures 8 and 9 are contours of HF breakthrough in a range-azimuth format with elevation angle as parameter for both the ordinary and extraordinary rays respectively. The calculations were for a sunspot number of 0 and a $K_p = 3$. The local time is noon, the month is March, and the rays are launched from a point located at 75° W, 20° N. Penetration is most clearly favored toward the North and refractive effects are slightly more evident for the extraordinary ray than for the ordinary ray.

5. FUTURE PLANS

As indicated in section 2.0 it is planned to modify the existing software so that zenithal ray characteristics may be computed as a function of other variables on a single submission. Presently the program only allows for iteration of elevation angle at a fixed bearing with all other parameters fixed. Thus many computer runs are necessary to characterize the ionospheric iris with a substantial increase in cost and turn-around time. As of this writing, such modifications have not yet been debugged. Once we have characterized the zenithal characteristics of the iris as a function of sunspot number, magnetic Kp, month, time of day, geography, and mode of propagation, specific cases will be run at a full set of azimuth and elevation angles.

Since there is a potential need to assess the environment over oceanic areas, especially for the uncooperative link, NRL has been examining various system approaches to the problem. Figure 10 gives the classical picture of a forecasting system based either upon physical or climatological models. NRL and NOSC have been working on a system of this type for HF frequency management based in part upon SOLRAD HI 1-8 Å solar x-ray data which is a determinant for SID modifications of the LUF. However it is a difficult task at this time to deduce the MUF (or the FOT) from first principles, and climatological models are relatively unsatisfactory for this purpose especially over oceanic areas. Figure 11 is a cartoon which depicts the generic system which would be envisioned to solve the problem at hand. This system would employ empirical models and extrapolation algorithms driven by real-time inputs from both ground-based and satellite data. However the type of inputs to the real-time assessment system are viewed to be more directly related to the plasma density which is the important parameter. It is emphasized that Navy needs imply that strong emphasis must be placed upon topside ionosonde as the primary sensor with (passive) HF breakthrough receivers serving as a backup. Also note that the prediction system exhibited in Figure 10 surfaces as a default system in Figure 11.

6. ACKNOWLEDGEMENTS

The author is indebted to J. Thomason for providing an advance copy of an NRL report (Thomason et al, 1979) and for allowing use of his NRL global model of the ionosphere in preparation of this paper. F. D. Clarke is acknowledged for programing assistance and D. R. Uffelman is thanked for his helpful comments.

REFERENCES

ATKINS, E. A., and J. H. CHAPMAN, 1963, "Satellite Measurements of Radio Breakthrough", Can. J. Phys., 41, 1388-1393.

BARGHAUSEN, A. F., J. FINNEY, L. PROCTOR, and L. SCHULTZ, 1969, "Predicting Long Term Operational Parameters of High-Frequency Sky-Wave Telecommunication Systems", ESSA Tech. Rpt. ERL 110-ITS 78.

BENT, R. B., S. G. LLEWELLYN, G. NESTERCZUK, and P. SCHMID, 1975, "The Development of a Highly Successful Worldwide Emperical Ionospheric Model and its Use in Certain Aspects of Space Communications and Worldwide Total Electron Content Investigations" in Effect of the Ionosphere on Space Systems and Communications edited by J. M. Goodman, Govt. Printing Office.

CCIR Report 340, 1966, ITU, Oslo.

EISELE, J. A., and P. R. V. SHANNON, 1977, "Ocean Surveillance Analysis by Computer (the OSAC System)", NRL Memo Rpt. 3503.

HAGG, E. L., E. J. HEWINS, and G. L. NELMS, 1969, "The Interpretation of Topside Sounder Ionograms", Proc. IEEE, 57 (6), 949-960.

HUGUENIN, C. R., and M. D. PAPAGIANIS, 1965, "Space-Borne Observations of Radio Noise from 0.7 to 7.0 MHz and Their Dependence on the Terrestrial Environment", Ann. Astrophysics, 28, 239.

JONES, R. M., and J. J. STEPHENSON, 1975, "A Versatile Three-Dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere", Office of Telecommunications Report 75-76.

JONES, W. B., and D. L. OBITTS, 1970, "Global Representation of Annual and Solar Cycle Variation of foF2 Monthly Median 1954-1958", OT ITS Research Rpt. 3.

MALIPHANT, R. G., 1967, "Ionospheric Refractions of High Frequency Radio Waves Propagating between the Earth and Orbiting Satellites" in Propagation Factors in Space Communications edited by W. T. Blackband, AGARD Proc. #3, W. & J. MacKay Ltd., London.

MEAD, J. B., and L. S. WAGNER, 1971, NRL Report 7284.

MILLER, D. C., and J. GIBBS, 1975, "Ionospheric Analysis and Ionospheric Modeling", AFCRL Rpt. 75-0549.

NELMS, G. L., 1964, "Ionospheric Results from the Topside Sounder Satellite Alouette", in Space Research IV edited by P. Muller, North-Holland, 437-451.

RAWER, K., 1967, "Noise Produced by Terrestrial Sources in Near Earth Space", in Propagation Factors in Space Communications edited by W. T. Blackband, NATO/AGARD, 383-402.

RUSH, C. M., D. NELSON, A. L. SNYDER, V. PATTERSON, T. TASCIONE, and E. ZIEMBA, 1978, "HF Noise in Space" in Effect of the Ionosphere on Space and Terrestrial Systems edited by J. M. Goodman, Govt. Printing Office.

RUSH, C. M. and E. ZIEMBA, 1978, "On the Usefulness of Topside HF Noise Measurements in Determining foF2", J. Atmospheric Terrest. Phys., 40, 1073-1079.

RUSH, C. M., and J. BUCHAU, 1977, "Determining the F-Region Critical Frequency from Satellite Borne Noise Measurements", J. Atmospheric Terrest. Phys. 39, 277-286.

THOMASON J., G. SKAGGS, and J. LLOYD, 1979, "A Global Ionospheric Model", NRL Report (to be published).

THOMPSON, R. L, 1978, "User Requirements of Aerospace Propagation-Environment Modelling and Forecasting", NATO/AGARD (OTTOWA) paper 1-2.

THOMPSON, R. L., and V. PATTERSON, 1979, Private Communication.

ETTE, J. I., R. W. VOSTREYS, and R. HOROWITZ, 1978, "Report on Active and Planned Spacecraft and Experiments", NSSDC/WDC-A-R&S.

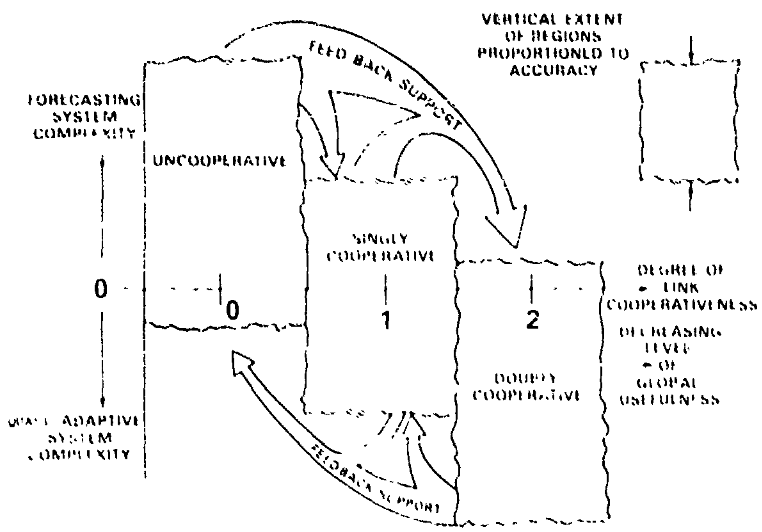

VERTICAL EXTENT OF REGIONS PROPORTIONED TO ACCURACY
FEED BACK SUPPORT
FORECASTING SYSTEM COMPLEXITY
UNCOOPERATIVE
SINGLY COOPERATIVE
0
0
1
2
DEGREE OF LINK COOPERATIVENESS
DECREASING LEVEL OF GLOBAL USEFULNESS
DOUBLY COOPERATIVE
QUASI ADAPTIVE SYSTEM COMPLEXITY
FEEDBACK SUPPORT

Fig. 1. Relationship between specified (doubly-cooperative) links and arbitrary (uncooperative) links vis-a-vis overall system complexity, level of global usefulness, and feedback scenario.

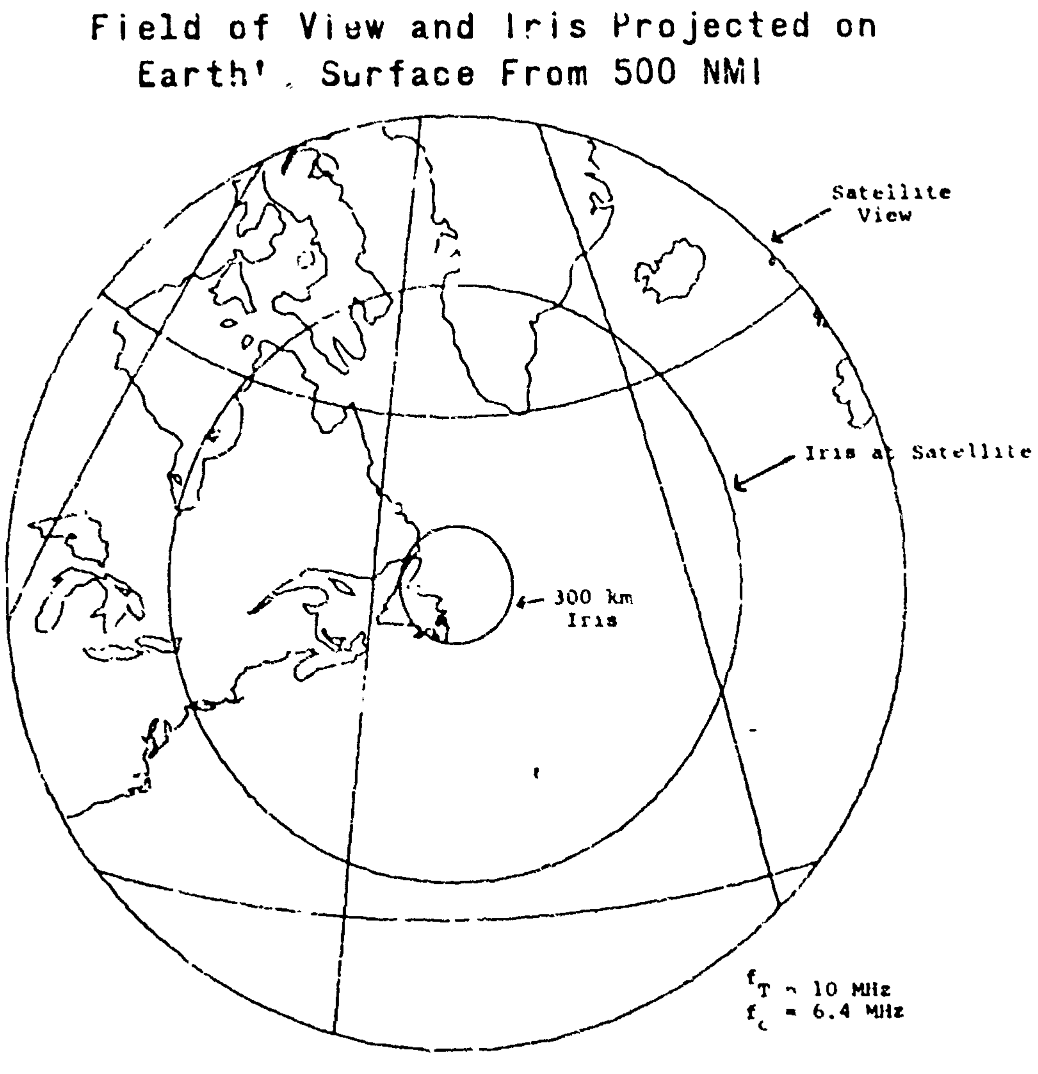

Field of View and Iris Projected on Earth's Surface From 500 NMI
Satellite View
Iris at Satellite
300 km Iris
f_T ~ 10 MHz
f_c = 6.4 MHz

Fig. 2. Projection of the field of view of a satellite orbiting at 834 km and the extent of the first-order ionospheric iris at 300 km and at the satellite height.

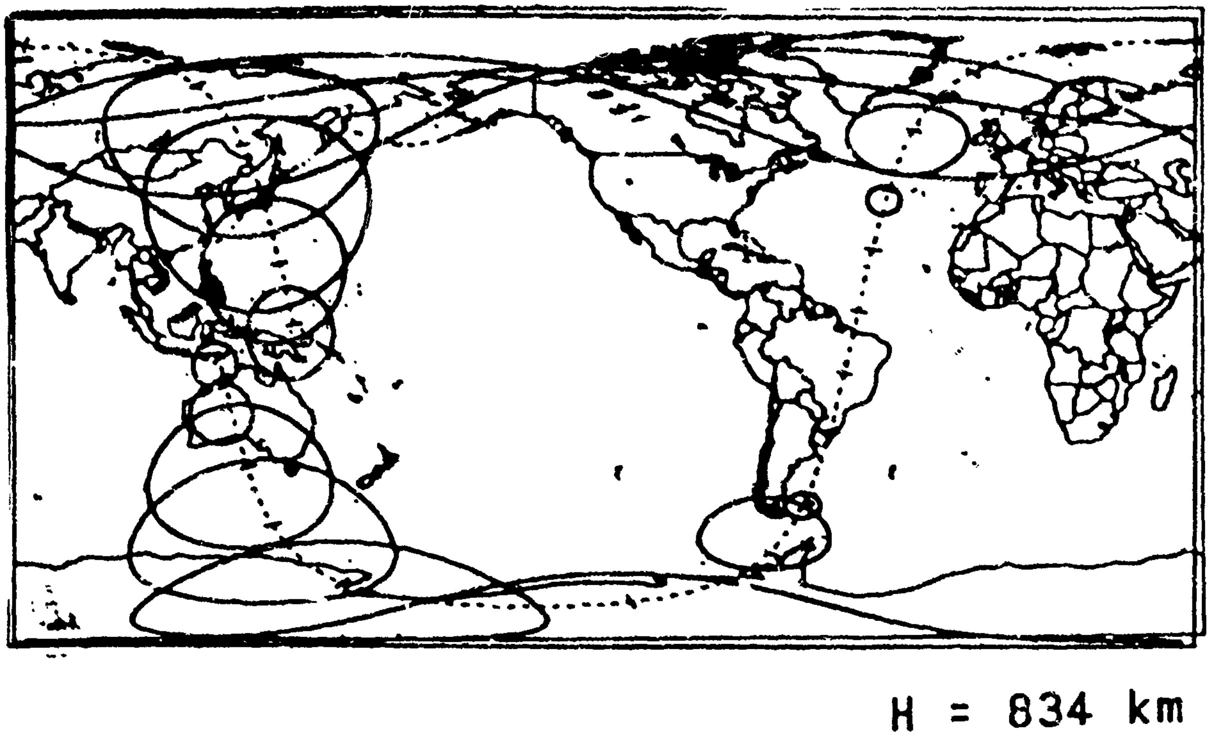


Fig. 3. Projection of individual first-order ionospheric irises for a DMSP orbit occurring during the month of March at 1800 GMT. The sunspot number is taken to be 100 and K_p = 3. The CCIR model (1966) is used in these calculations. Each tick mark is separated from its neighbor by 5 minutes. The iris distortion at high latitudes is a result of the projection used.

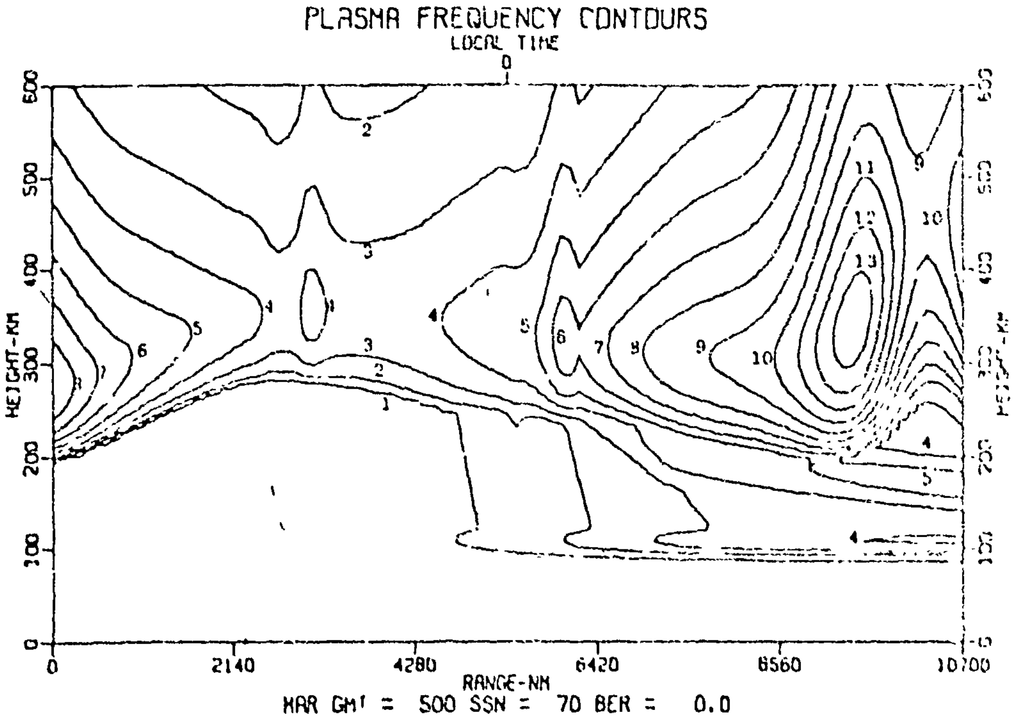


Fig. 4. Plasma frequency contours for March at 0500 GMT from the equator along the 69° W meridian, through the pole, and to the opposite equator along the 111° E meridian. A sunspot number of 70 is assumed and K_p = 3. The model is due to Thomason et al (1979).

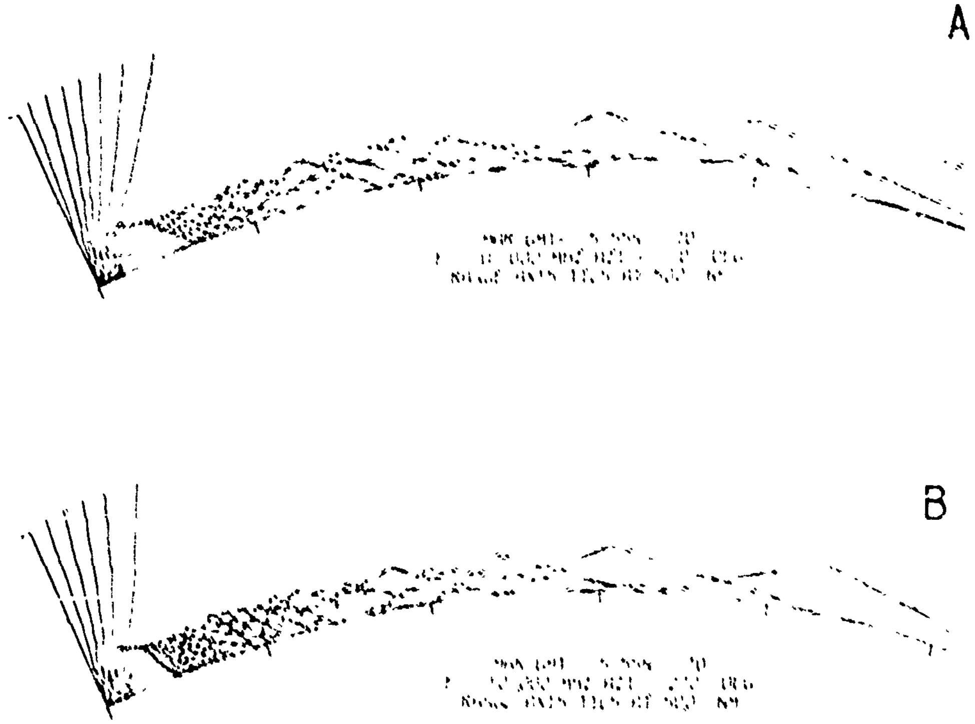


Fig. 5. (A) HF raytraces for March at 2400 GMT from the geographic equator at 60°E toward the North at steps of 5° elevation. The transmitter frequency is 10 MHz.

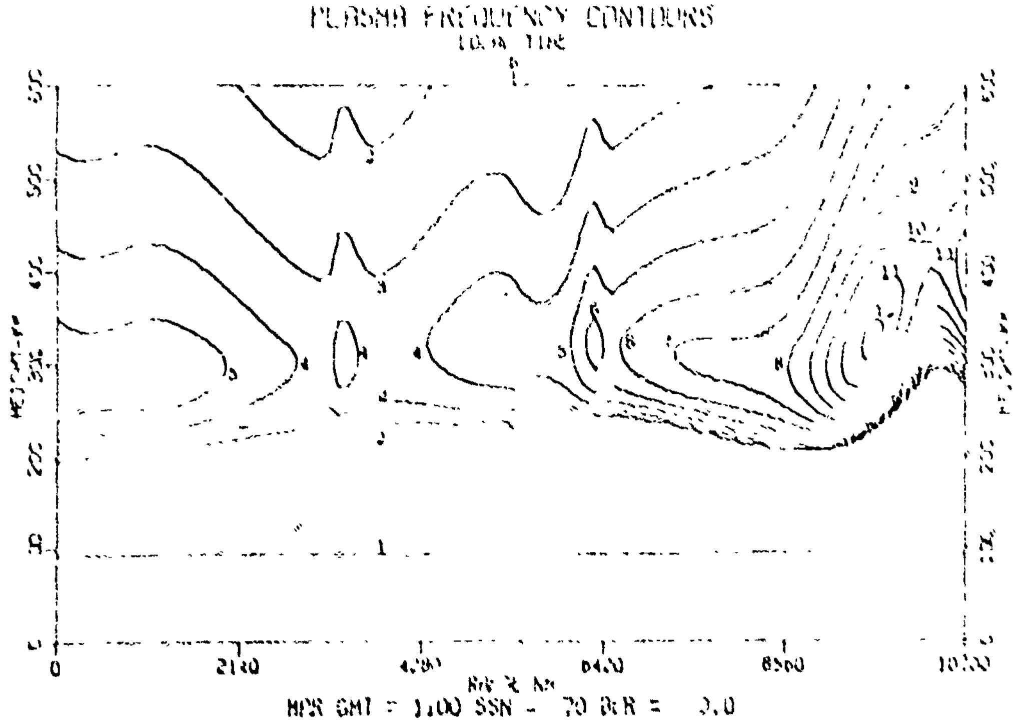


Fig. 6. Plot of plasma frequency contours for conditions similar to Figure 3 but for March at 1100 GMT.

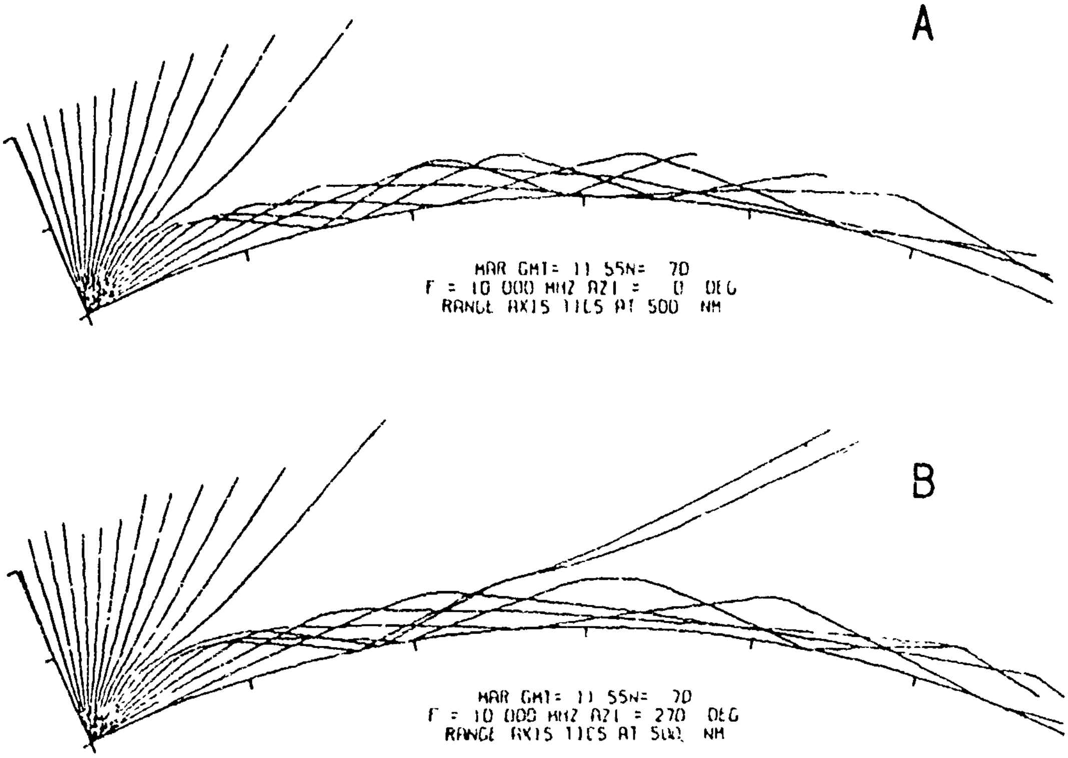


Fig. 7. (A) HF Raytraces for March 1100 GMT from the geographic equator at 69° W toward the North in steps of 5° elevation. The transmitter frequency is 10 MHz.

(B) HF raytraces for conditions similar to (A) above but for propagation toward the West.

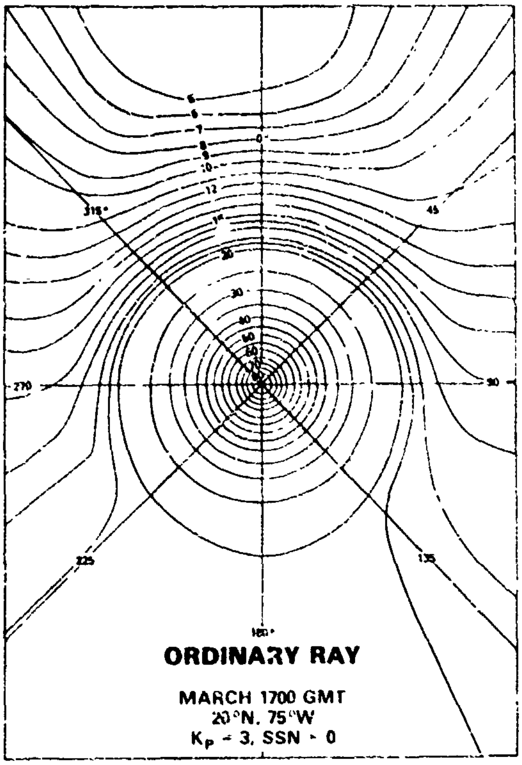


Fig. 8. Contours of HF breakthrough in a Range-Azimuth format with elevation angle as parameter. Ordinary ray. Range is linear on the plot, being 500 kilometers from the origin to the top (bottom) border, and 350 kilometers from the origin to the right-hand (left-hand) side border.

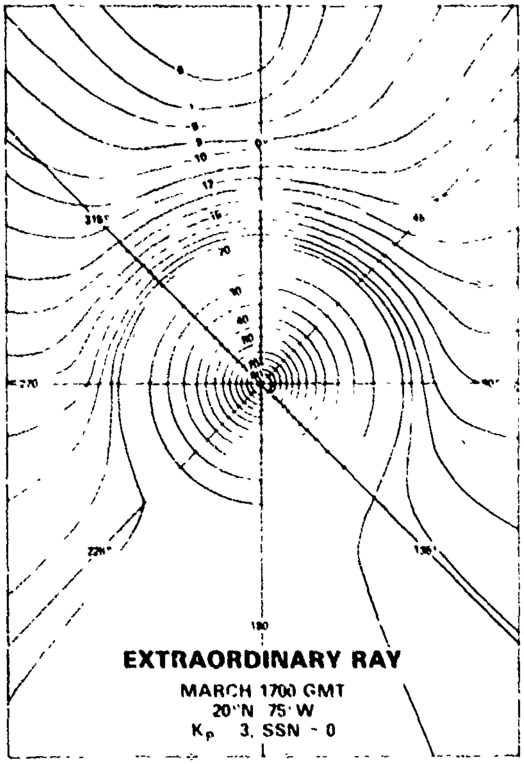


Fig. 9. Contours of HF breakthrough in a Range-Azimuth format with elevation angle as parameter. Extraordinary ray. Range is linear on the plot, being 500 kilometers from the origin to the top (bottom) border, and 350 kilometers from the origin to the right-hand (left-hand) side border.

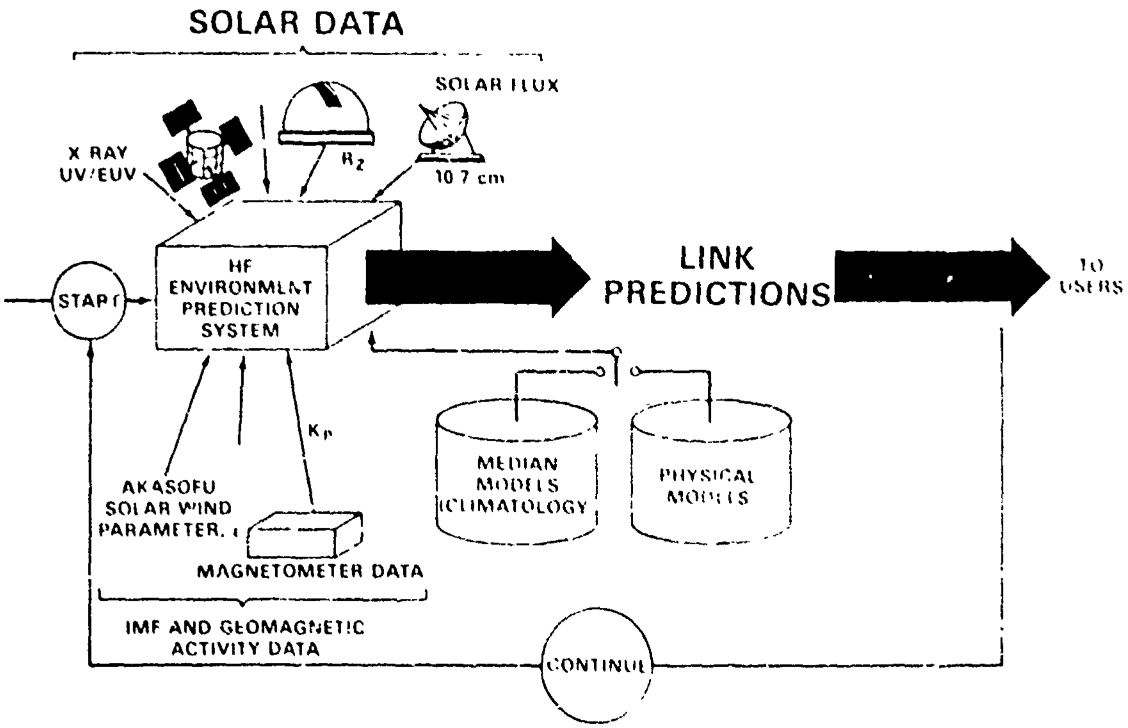

SOLAR DATA
SOLAR FLUX
X RAY
UV/EUV
R_Z
10.7 cm
START
HF ENVIRONMENT PREDICTION SYSTEM
LINK PREDICTIONS
TO USERS
K_p
AKASOFU SOLAR WIND PARAMETER, ε
MAGNETOMETER DATA
MEDIAN MODELS (CLIMATOLOGY)
PHYSICAL MODELS
IMF AND GEOMAGNETIC ACTIVITY DATA
CONTINUE

Fig. 10. Environmental Prediction System for HF link analysis.

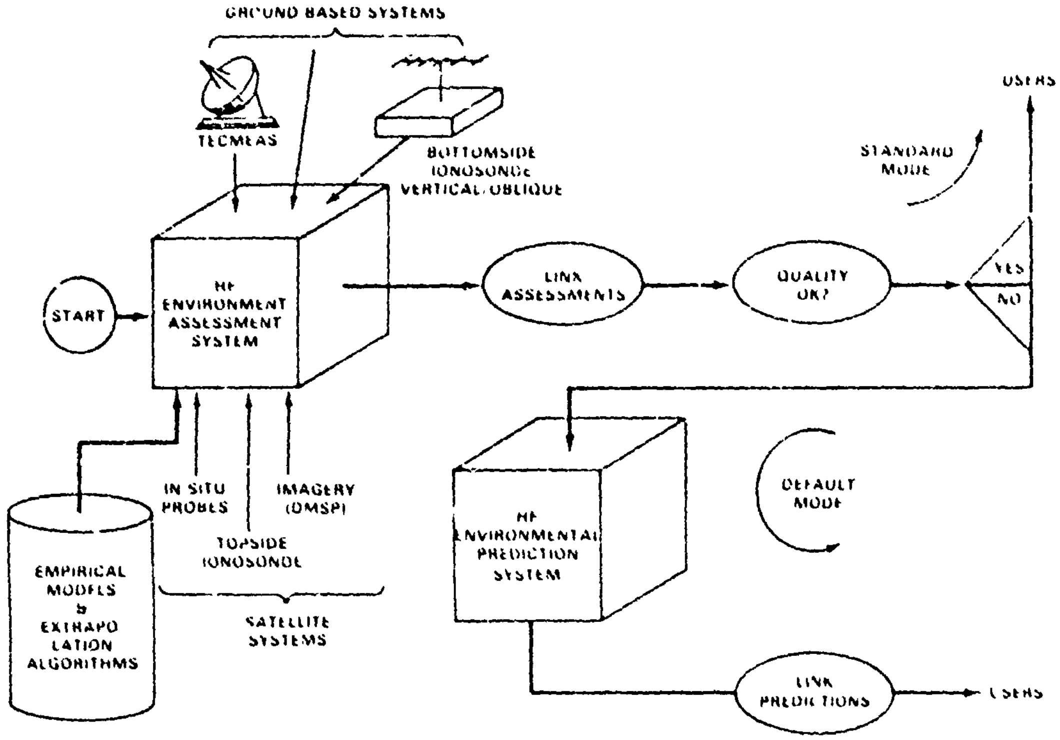

GROUND BASED SYSTEMS
TECMEAS
BOTTOMSIDE IONOSONDE VERTICAL/OBLIQUE
USERS
STANDARD MODE
START
HF ENVIRONMENT ASSESSMENT SYSTEM
LINK ASSESSMENTS
QUALITY OK?
YES
NO
IN SITU PROBES
IMAGERY (DMSP)
TOPSIDE IONOSONDE
SATELLITE SYSTEMS
EMPIRICAL MODELS & EXTRAPOLATION ALGORITHMS
HF ENVIRONMENTAL PREDICTION SYSTEM
DEFAULT MODE
LINK PREDICTIONS
USERS

Fig. 11. Environmental Assessment and Extrapolation System for HF link analysis.

DISCUSSION

Comment by B.W. Reinisch, US

You mentioned that a network of groundbased stations would supply the required information. In principle I agree, but a topside ionosonde will give information on the topside structure related to the occurance of scintillation. We are presently developing, together with RCA, a satellite born digisonde. By means of Doppler measurements we will obtain some indication of the direction of the ECHO motion and thus on the horizontal gradients of the topside ionosphere.

Author's Comment

I agree with you on your remark. By means of the final figure I showed, I was indicating that it would be possible to incorporate a wide range of environmental sensore to assist in solution of the Navy problem, that is obtaining adequate real-time updated assessments of the HF links over oceanic areas especially for the totally uncooperative situation. In view of the paper authored by Paul and Wright and presented by Dr Rush at this session I suggested that bottomside sounders of the advanced type proposed by those authors (provided they are optimally spaced) might be a partial solution if they were folded into a total system. I didn't mean to suggest that the bottomside sounder approach would solve the problem by itself. Indeed I was attempting to point out in my paper that topside data obtained either actively or passively using the HF breakthrough phenomenon would likely provide the best and most accurate coverage for the Navy. I concur that scintillation assessment is another feature that may in principle be addressed by the topside sounder approach and we are interested in this ourselves. To the extent that topside spread-F is related to scintillation phenomenon, a combined HF and SATCOM link assessment system on a single platform would certainly have a gread deal of charm. Also the Doppler capability provided by the satellite borne digisonde you are planning should be quite useful operationally and scientifically.

Comment by Prof Shearman, UK

The speaker has referred to the difficulty in using the HF breakthrough phenomenon in sea areas where there is a paucity of transmitters.

It should be possible to use a transmitter on land to illuminate a sea area by the conventional oblique ionosphere-reflected Hop mode. Energy scattered in a near-vertical direction from the sea surface and penetrating the ionosphere through the transparent "[illegible]" should then be detectable in a satellite vertically overhead. Propagation by this mode, but in the satellite-ground direction was identified by Japanese workers in Doppler observations of the earliest USSR Sputniks.

Author's Reply

Your suggestion is an interesting one but I am not familiar with early work you have cited. One of the difficulties I see in the proposition is that substantial transmitter power might be necessary to cover the vast ocean areas. Also, the extent of "side scatter" directed toward the satellite would be dependent upon the sea state and I suspect as a result that very unusual circumstances would have to obtain for the concept to be viable, say operationally. I must admit, however, that my counter-argument is only heuristic and not rigorously reasoned.

COUPLING BETWEEN THE NEUTRAL AND IONIZED UPPER ATMOSPHERE DURING DISTURBED CONDITIONS

G. W. Prölss
Institut für Astrophysik und extraterrestrische Forschung
Auf dem Hügel 71, D 5300 Bonn, F.R. Germany

SUMMARY

Using neutral composition data obtained by the ESRO 4 satellite and ionospheric data obtained from ground-based stations the storm-time interaction of the neutral and ionized upper atmosphere is reviewed. First we show that a close coupling exists between magnetic storm associated changes in the O/N_2 density ratio and corresponding changes in the critical frequency of the F2 layer indicating that negative ionospheric storms are caused by changes in the neutral composition. This experimental evidence is supported by a model calculation which demonstrates that the measured disturbance of the O/N_2 density ratio is indeed sufficient to explain the observed reduction in ionospheric plasma density. These data also show that positive storm effects are generated by a different disturbance mechanism. Given the close coupling between neutral atmospheric and negative ionospheric perturbations both phenomena have certain features in common. These include their variation with the intensity of the disturbance, their dependence on magnetic position, their variation with local time, and their variation with season. Presently available models of the neutral atmosphere are not yet sophisticated enough to reproduce these disturbance variations and are therefore of little use for modeling ionospheric storm effects.

1. INTRODUCTION

It is well known that communication circuits involving radio propagation through the ionosphere are materially affected during magnetic storms. Although this was discovered more than fifty years ago (Anderson, 1928) little progress has been made in modeling this important effect. The reasons for this are twofold. First of all our knowledge of the physical mechanisms responsible for ionospheric disturbances has been rather incomplete. And secondly the high degree of irregularity exhibited by these disturbance phenomena tended to mask any systematic variations.

Recently a number of satellite-born gas analyzers (neutral mass spectrometers) have provided extensive information on the neutral composition of the upper atmosphere. These data allowed for the first time a detailed study of magnetic storm associated changes in the neutral composition and their relation to ionospheric disturbance effects. As a result significant progress has been made in our understanding of both phenomena. It is the purpose of the present communication to summarize some of the findings obtained in these studies which are important for modeling ionospheric disturbance effects. First we show that negative ionospheric storms at middle and high latitudes are caused by changes in the neutral composition of the upper atmosphere (section 2). This section also presents evidence that positive storm effects are generated by a different mechanism. The second part of this paper describes some systematic variations which are common to both neutral atmosphere and negative ionospheric disturbance events (section 3). The progress made in modeling these disturbance effects is reviewed in the final part of this paper (section 4).

2. STORM-TIME COUPLING BETWEEN F2-LAYER PLASMA DENSITY AND NEUTRAL COMPOSITION

Theoretical studies (e.g. Cole, 1975; Rees, 1975) and recent observations (e.g. Banks, 1977; Brekke and Rino, 1978; Murphree and Anger, 1978) clearly demonstrate that magnetospheric storms deposit a large amount of energy in the high latitude region of the upper atmosphere. This energy injection leads to the development of disturbance zones in the neutral composition, as indicated schematically in Fig. 1 . Using gas analyzer data from polar-orbiting satellites the properties of these perturbations have recently been studied in some detail (e.g. Taeusch et al., 1971; Prölss and Fricke, 1976, Hedin et al., 1977c). Their basic latitudinal structure is illustrated in Fig. 2 using data obtained at about 9 LT over the North American continent during the strong geomagnetic storm of 29 October 1973. Magnetic storm induced changes in molecular nitrogen, atomic oxygen, helium and argon are plotted. Also shown for comparison is the latitudinal structure of the total mass density and of the molecular nitrogen to atomic oxygen concentration ratio. The departures from quiet-time conditions are presented in the form of relative changes. Thus R(n) is defined as the actually observed value of a density or parameter n divided by the corresponding quiet-time value; and R(n) = 1 serves as a reference, meaning no change with respect to quiet times. For further details on the form of data presentation see Prolss and Fricke (1976).

Prominent among the features illustrated in Fig. 2 is the well-defined disturbance zone in the neutral composition extending from high to middle latitudes. This disturbance zone is characterized by a large increase in argon, a smaller but still considerable increase in molecular nitrogen, and a significant decrease in helium. Atomic oxygen, which is the major constituent in the height range considered, shows a complex behavior with both moderate increases and decreases observed. Decreases are seen primarily at lower altitudes and increases at greater heights, indicating a height dependent behavior. Outside the well-defined higher latitude disturbance zone the perturbation effects are less conspicuous. Thus at low latitudes the geomagnetic activity effect consists of a moderate enhancement in all four constituents.

The lower part of Fig. 2 demonstrates that the molecular nitrogen to atomic oxygen density ratio is well-suited to study the latitudinal structure of an atmospheric disturbance. Whereas inside the disturbance zone this parameter shows a definite increase, indicating both the extent and magnitude of the composition disturbance, the lower latitude regime is characterized by a lack of change in this ratio. Additional advantages of this parameter are its large-scale height, which makes it less sensitive to changes in observation height, and its close connection to ionospheric perturbation effects. On the other hand, the total mass density ρ also shown in Fig. 2 is certainly ill-suited to study the disturbance morphology.

The measured changes in the neutral composition and specifically those in the atomic oxygen to molecular nitrogen (O/N_2) ratio are expected to materially affect the ionospheric plasma density and this is indeed observed. Figure 3 shows a comparison between the day-to-day variation of the atomic oxygen to molecular nitrogen ratio and that of the maximum electron density in the F2 layer. Plotted are data from six mid-latitude stations during a nine-day interval in February 1973. Hermanus (HE) is located in South Africa, Salisbury (SR), Townsville (TV), and Brisbane (BR) in Australia, and Port Stanley (PS) and South Georgia (SG) in the South American sector. Both data sets have been normalized to values observed on a reference day ($R(N_{max}) = R(O/N_2) = 1$) . In addition the O/N_2 ratios have been corrected for changes in observation height. Local solar time is approximately 11 hours. As is evident the variations of both parameters show a high degree of conformity. This is especially obvious during the magnetic-atmospheric-ionospheric disturbance starting on February 21 when a significant decrease in the O/N_2 ratio causes a similar drop in the plasma density.

Using the same form of data presentation, Fig. 4 shows the relative day-to-day variations in the O/N_2 ratio and in the maximum electron density for a six-day interval in October 1973. Point Arguello (PA), Boulder (BO), and Wallops Island (WP) are located in North America, Leningrad (LD) and Tomsk (TK) in Russia, and Brisbane (BR) in Australia. Local solar time of observation is approximately 9 hours. Again we note the close coupling between the two data sets, especially during the major magnetic storm of October 29/30, 1973.

From a theoretical point of view it is of great interest to know whether the magnitude of the observed composition changes is indeed sufficient to explain the observed variation in ionospheric plasma density. A detailed analysis of this problem has recently been performed by Jung and Prölss (1978). In their study the ionospheric behavior is simulated using a numerical model and actually observed neutral composition data. A result of this computation is shown in Fig. 5. The upper part shows the normalized day-to-day variation in the O/N_2 ratio as observed near Boulder during a six-day interval in October 73, a period which includes the strong magnetic storm of 29 October 73 (see also Fig. 4). These composition data have been introduced into the ionospheric model to reproduce the day-to-day variation in the maximum electron density observed at this station. The result of this calculation are presented in the lower part of Fig. 5 . Also shown for comparison is the observed variation of the maximum electron density. The good agreement between theoretical prediction and actual measurement is evident. This demonstrates that the observed composition changes can fully explain the observed reduction in the ionospheric plasma density, i.e. negative ionospheric storm effects.

Although in the past positive ionospheric storm effects also have been linked to changes in the neutral composition , it is now generally believed that storm associated increases in plasma density are caused by transport of ionization (e.g. Jones, 1973; Tanaka and Hirao, 1973; Anderson, 1976; Prolss and Jung, 1978). Recent gas analyzer measurements have provided direct evidence that there is indeed no correlation between positive disturbance effects and changes in the neutral composition. Figure 6 illustrates the distribution of positive ionospheric storm effects in relation to the latitudinal structure of an atmospheric disturbance. The upper panel shows changes in the N_2/O concentration ratio as observed during a disturbance event in October 1973. $R(N_2/O) = 1$ again serves as a reference, meaning no change with respect to quiet times. The lower part shows the time variation of the critical frequency as observed at three stations located along the satellite pass which provided the neutral composition data. Here the monthly median for October 1973 serves as the quiet-time reference. A comparison shows that both Yakutsk (YA) and Brisbane (BR), located inside the northern and southern disturbance zones, respectively, show a considerable depression in the critical frequency, i.e. negative disturbance effects. Positive ionospheric storm effects, on the other hand, are observed at Yamagawa (YG), which is situated outside the disturbance zones. Data presented in Fig. 7 corroborate this pattern whereas negative disturbance effects are associated with changes in the neutral composition positive storm effects are observed outside the N_2/O disturbance zones.

Further evidence for the lack of correlation between neutral composition data and positive ionospheric storm effects is presented in Fig. 8. Again a comparison is shown between the day-to-day variation in the O/N_2 concentration ratio and changes in the maximum electron density of the F2 layer. This time, however, only stations located outside the composition disturbance zone have been considered. As is evident, no correlation is found between the smoothly varying composition and the enhancements seen in the electron density. If we assume then that positive disturbance effects are caused by transport of ionization, and furthermore, that this transport is affected by winds and/o electric fields the block diagram in Fig. 9 may help to illustrate the principle interactions between magnetospheric, atmospheric and ionospheric disturbance events.

3. SYSTEMATIC VARIATIONS OF NEUTRAL-ATMOSPHERIC AND IONOSPHERIC DISTURBANCES

In spite of a considerable degree of irregularity in the ionospheric storm behavior certain systematic variations are clearly recognized. These include changes with the intensity of the disturbance, with geomagnetic position, with local time, and with season. Recent studies based on gas analyzer data show that these variations can be directly attributed to similar changes in the neutral composition. Thus they confirm the close coupling between atmospheric and ionospheric disturbance effects. This section summarizes the results of these investigations in a uniform manner.

3.1 Variation with magnetic storm intensity

Early studies of ionospheric storms revealed that at mid-latitudes there is a change-over from positive to negative disturbance effects with the growing intensity of a magnetic perturbation (Appleton and Ingram, 1935). This change-over is part of a more general pattern with both the magnitude and extension of negative storm effects increasing with the severity of the magnetic disturbance (e.g. Berkner and Seaton, 1940; Lange-Hesse, 1955, 1965; Matsushita, 1959). Recently gas analyzer data have shown that similar variations are also observed in the morphology of magnetic storm associated changes in the neutral composition (e.g. Prölss and Fricke, 1976). This is illustrated in Fig. 10 which shows the magnitude and extension of a neutral composition disturbance for two different levels of magnetic activity. Relative changes in the molecular nitrogen to atomic oxygen (N_2/O) concentration ratio are plotted with $R(N_2/O) = 1$ serving as the quiet time reference. To facilitate a comparison all data have been reduced to a common altitude of 280 km. Also to illustrate the disturbance morphology in a more comprehensive manner five latitudinal profiles have been superimposed for each disturbance level. As is evident, the magnitude of the N_2/O disturbance as well as its extension towards lower latitudes increase with the intensity of the magnetic storm in good agreement with corresponding changes in the morphology of ionospheric storm effects.

3.2 Variations with geomagnetic position

Systematic variations of ionospheric storm effects with geomagnetic position are not so well documented. Studying the global morphology of ionospheric perturbations, Appleton and Piggot (1952) noted that negative storm effects are more closely correlated with geomagnetic than with geographic coordinates. Subsequent studies often implicitly presumed this magnetic latitude control by using geomagnetic coordinates to organize their data (e.g. Matsushita, 1959; Rajaram and Rastogi, 1970).

Recent gas analyzer data suggest that this geomagnetic latitude dependence of ionospheric storm effects is caused by corresponding changes in the neutral composition. Thus very distinct longitudinal variations which are in phase with corresponding changes in the magnetic invariant latitude are frequently observed during disturbed conditions (e.g. Prölss and von Zahn, 1974a, 1977a; Jacchia et al., 1976). Figure 11 illustrates this effect. The lower panel shows a comparison between the longitudinal variation of the N_2/O density ratio and the variation of the magnetic invariant latitude at constant geographic latitude. All density data have been adjusted to a common altitude. As is evident, there is good agreement between the systematic variations of both parameters, clearly demonstrating the invariant latitude control of composition disturbances.

To show that these variations are not due to similar temporal changes in the magnetic activity the upper panel of Fig. 11 shows the magnetic AE index during the time interval of interest. The spatial character of this disturbance variation is further corroborated by ionospheric data. Figure 12 illustrates the ionospheric conditions near the maxima and minima of the N_2/O variation. Whereas Wallops Island (WP) (located near the center of the N_2/O disturbance) persistently shows considerable negative storm effects, Tokyo (TO) tends to exhibit only smaller depression effects in f_0F2.

3.3 Local time variations

One of the more prominent characteristics of ionospheric storms is their local time dependence. A first description of this effect was given by Kirby et al. (1936) who state that severe magnetic storms beginning during the daytime may show little correlation with radio data (i.e. negative storm effects), while a severe magnetic disturbance before sunrise is accompanied by disturbed radio conditions during the entire next day. Subsequent studies confirmed these findings and showed that at mid-latitudes negative storm effects commence most frequently in the early morning and very rarely in the noon and afternoon local time sectors (e.g. Appleton and Piggot, 1952; Thomas and Venables, 1966; Prölss and von Zahn, 1978). This is illustrated in Fig. 13. On the other hand, positive ionospheric storm effects are most frequently observed in the noon and afternoon sector of the first storm day (e.g. Mendillo, 1973; Hargreaves and Bagenal, 1977).

An explanation of these effects in terms of composition changes would involve a disturbance which is much larger in the nighttime/early morning sector than in the daytime sector (Prölss, 1976). Recent gas analyzer data are consistent with this concept. Figure 14 shows a comparison between the disturbance magnitude of two events which occurred during the same season and were associated with magnetic activity of comparable magnitude but were observed at different local times, namely in the morning and afternoon sectors, respectively. Relative changes in the N_2/O concentration ratio are plotted again as a function of magnetic latitude with $R(N_2/O)$ serving as the quiet-time reference. Ten of the most disturbed orbits have been superimposed for each event. As is evident, there is a significant local time asymmetry in the disturbance magnitude with larger disturbance effects in the morning sector. This supports our hypothesis that the major disturbance occurs in the nighttime/early morning sector. Unfortunately the data are not suited to document local time dependent changes in the extension of the disturbance since they were collected during all stages of the storm development.

Further evidence for a local time asymmetry comes from a statistical analysis of molecular nitrogen data obtained by the OGO 6 satellite (Taeusch, 1977). Although based on a limited data sample, this study also indicates that the major disturbance occurs in the night sector. However, in addition a clear local time asymmetry in the extension of the disturbance is evident with the N_2 density increase reaching down to much lower magnetic latitudes in the night sector.

3.4 Seasonal variations

Seasonal variations also belong to the well documented features of ionospheric storms. They were discovered early (Appleton et al., 1937; Kirby et al., 1937) and have since been investigated in numerous studies (e.g. Appleton and Piggot, 1952; Lange-Hesse, 1965; Duncan, 1969; Spurling and Jones, 1973; Prolss, 1977). The most prominent aspect of these variations is that during summer negative storm effects are observed to extend all the way from the polar region to the subtropics, whereas during winter they are restricted to the higher latitudes. This is illustrated in Fig. 15. Recent gas analyzer data show again that these changes can be directly attributed to similar variations in the neutral composition (Prolss and von Zahn, 1977b). This is illustrated in Fig. 16 which shows a comparison between the latitudinal structure of a summer and a winter disturbance. Again the N_2/O density ratio is used to describe the properties of the composition disturbance. $R(N_2/O) = 1$ serves as the quiet-time reference, and ten of the most disturbed orbits have been superimposed to describe the disturbance structure in a comprehensive way.

A comparison of the summer and winter data reveals that (1) whereas in summer the N_2/O disturbance extends all the way from the high to the low latitude regions, it is restricted to the higher latitudes in winter; this is in excellent agreement with ionospheric storm data; (2) whereas in summer the increase in the N_2/O disturbance at the lower latitude boundary is a gradual one, rather steep increases are frequently observed during winter; and (3) whereas the summer perturbations are of moderate magnitude, fairly large disturbance effects are commonly seen in winter.

4. MODELING ATMOSPHERIC DISTURBANCE EFFECTS

Given the close coupling between the neutral and ionized upper atmosphere any progress made in modeling neutral atmospheric disturbances should also be of use in describing ionospheric storm effects. The large amount of data provided by recent gas analyzer measurements has indeed stimulated the development of a number of emperical models of the thermospheric region which are of interest here (e.g. MSIS model: Hedin et al., 1977a, 1977b; ESRO 4 model: von Zahn et al., 1977; Jacchia '77 model: Jacchia et al., 1977). A comparison shows that all these models are well suited to reproduce the basic features of magnetic storm associated changes in the neutral composition. However none of these algorithms is sophisticated enough to reproduce the systematic variations discussed, for example, in the previous section. Thus all models consider only zonally averaged values and all are based on a fixed latitudinal profile. This means, for instance, that none of the systematic variations in the extension of a composition disturbance are taken into account.

Another major drawback of these models is, of course, that they do not provide any information on magnetic storm induced winds which are so important for modeling positive ionospheric storm effects.

Theoretical models do provide this information on the disturbed meridional wind component (Mayr and Volland, 1973). They also provide some insight into the mechanisms leading to the systematic variations discussed in the previous section (Volland, 1979). The disadvantage of this physical approach is that a number of simplifying assumptions must be introduced which limit the applicability of the theory. Also some of the basic input parameters needed for a quantitative treatment are incompletely known. Last, but not least, the formalism used is often too complex to be useful for practical applications. Given this situation neither the empirical nor the theoretical atmospheric models presently available are of much use in modeling ionospheric disturbance effects. It is, however, to be expected that this situation will improve in the near future with more sophisticated empirical models of the geomagnetic activity effect under development.

ACKNOWLEDGMENTS:

I am greatly indebted to U. von Zahn, who is the principle investigator of the ESRO 4 gas analyzer experiment. The ionospheric data used in this study were obtained from World Data Center A. This research was supported by the Deutsche Forschungsgemeinschaft through grant Pr 19/11.

5. REFERENCES

Anderson,C.N., 1928, Correlation of long wave transatlantic radio transmission with other factors affected by solar activity, Proc.Inst.Radio Eng., 16, 297 - 347 .

Anderson,D.N., 1976, Modeling the mid-latitude F-region ionospheric storm using east-west drift and a meridional wind, Planet.Space Sci., 24, 69 - 77 .

Appleton,E.V., and L.J. Ingram, 1935, Magnetic storms and upper-atmospheric ionisation, Nature, 136, 548-549.

Appleton,E.V., and W.R. Piggot, 1952, The morphology of storms in the F2 layer of the ionosphere I. Some statistical relationships, J.Atmos.Terr.Phys., 2, 236 - 252 .

Appleton,E.V., R. Naismith, and L.J. Ingram, 1937, British radio observations during the second international polar year 1932-33, Phil.Trans., Roy.Soc., A236, 191 - 259 .

Banks,P.M., 1977, Observations of joule and particle heating in the auroral zone, J.Atmos.Terr.Phys., 39, 179 - 193 .

Berkner,L.V., and S.L. Seaton, 1940, Systematic ionospheric changes associated with geomagnetic activity, Terr.Mag.Atmos.Elect., 45, 419 - 423 .

Brekke,A., and C.L. Rino, 1978, High-resolution altitude profiles of the auroral zone energy dissipation due to ionospheric currents, J. Geophys. Res., 83, 2517 - 2524 .

Cole,K.D., 1975, Energy deposition in the thermosphere caused by the solar wind, J.Atmos.Terr.Phys., 37, 939 - 949 .

Duncan,R.A., 1969, F-region seasonal and magnetic-storm behaviour, J.Atmos.Terr.Phys., 31, 59 - 70 .

Hargreaves,J.K., and F. Bagenal, 1977, The behavior of the electron content during ionospheric storms: A new method of presentation and comments on the positive phase, J.Geophys.Res., 82, 731 - 733 .

Hedin,A.E., P. Bauer, H.G. Mayr, G.R. Carignan, L.H. Brace, H.C. Brinton, A.D. Parks, and D.T. Pelz, 1977a, Observations of neutral composition and related ionospheric variations during a magnetic storm in February 1974, J.Geophys.Res., 82, 3183 - 3189 .

Hedin,A.E., J.E. Salah, J.V. Evans, C.A. Reber, G.P. Newton, N.W. Spencer, D.C. Kayser, D. Alcaydé, P. Bauer, L. Cogger, and J.P. McClure, 1977b, A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 1. N_2 density and temperature, J. Geophys.Res., 82, 2139 - 2147 .

Hedin,A.E., C.A. Reber, G.P. Newton, N.W. Spencer, H.C. Brinton, H.G. Mayr, and W.E. Potter, 1977c, A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 2. Composition, J.Geophys.Res.,82, 2148 - 2156 .

Jacchia.L.G., J.W. Slowey, and U. von Zahn, 1976, Latitudinal changes of composition in the disturbed thermosphere from Esro 4 measurements, J. Geophys.Res., 81, 36 - 42 .

Jacchia,L.G., J.W. Slowey, and U. von Zahn, 1977, Temperature, density, and composition in the disturbed thermosphere from Esro 4 gas analyzer measurements: A global model, J.Geophys.Res., 82, 684 - 688 .

Jones,K.L., 1973, Wind, electric field and composition perturbations of the mid-latitude F-region during during magnetic storms, J.Atmos.Terr.Phys., 35, 1515 - 1527 .

Jung, M.J., and G.W. Prölss, 1978, Numerical simulation of negative ionospheric storms using observed neutral composition data, J.Atmos.Terr.Phys., 40, 1347 - 1350 .

Kirby,S.S., T.R. Gilliland, N. Smith, and S.E. Reymer, 1936, The ionosphere, solar eclipse and magnetic storm, Phys.Rev., 50, 258 - 259 .

Kirby,S.S., N. Smith, T.R. Gilliland, and S.E. Reymer, 1937, The ionosphere and magnetic storms, Phys.Rev., 51, 992 - 993 .

Lange-Hesse, G., 1955, Geomagnetic activity and average deviations of daytime F_2-layer critical frequencies in different geomagnetic latitudes, J.Atmos.Terr.Phys., 7, 49 - 60 .

Lange-Hesse, G., 1965, Ionisationsschwankungen in der Ionosphere und Variation der brauchbaren Frequenzen im Kurzwellen-Weltverkehr entlang eines Langengrades in Abhängigkeit vom erdmagnetischen Störungsgrad, A.E.Ü., 19, 326 - 336 .

Matsushita,S., 1959, A study of the morphology of ionospheric storms, J.Geophys.Res., 64, 305 - 321.

Mendillo,M., 1973, A study of the relationship between geomagnetic storms and ionospheric disturbances at mid-latitudes, Planet.Space Sci., 21, 349 - 358 .

Murphree,J.S., and C.D. Anger, 1978, Instantaneous auroral particle energy deposition as determined by optical emissions, Geophys.Res.Lett., 5, 551 - 554 .

Prolss,G.W., 1976, On explaining the negative phase of ionospheric storms, Planet.Space Sci., 24, 607 - 609 .

Prolss,G.W., 1977, Seasonal variations of atmospheric-ionospheric disturbances, J.Geophys. Res., 82, 1635 - 1640 .

Prolss,G.W., and K.H. Fricke, 1976, Neutral composition changes during a period of increasing magnetic activity, Planet.Space Sci., 24, 61 - 67 .

Prolss,G.W., and M.J. Jung, 1978, Traveling atmospheric disturbances as a possible explanation for daytime positive storm effects of moderate duration at middle latitudes, J.Atmos.Terr.Phys., 40, 1351 - 1354 .

Prolss,G.W., and U. von Zahn, 1974a, Esro 4 gas analyzer results 2. Direct measurements of changes in the neutral composition during an ionospheric storm, J.Geophys.Res., 79, 2535 - 2539 .

Prolss,G.W., and U. von Zahn, 1977a, On the global morphology of negative ionospheric storms, Space Res., 17, 433 - 438 .

Prolss,G.W., and U. von Zahn, 1977b, Seasonal variations in the latitudinal structure of atmospheric disturbances, J.Geophys.Res., 82, 5629 - 5632

Prolss,G.W., and U. von Zahn, 1978, On the local time variation of atmospheric-ionospheric disturbances, Space Res., 18, 159 - 162 .

Rajaram,G., and R.G. Rastogi, 1970, North-south asymmetry of ionospheric storm-dependence on longitude and season, J.Atmos.Terr.Phys., 32, 113 - 118 .

Rees,M.H., 1975, Magnetospheric substorm energy dissipation in the atmosphere, Planet.Space Sci., 23, 1589 - 1596 .

Spurling,P.H., and K.L. Jones, 1973, The nature of seasonal changes in the effects of magnetic storms on mid-latitude F-layer electron concentration, J.Atmos.Terr.Phys., 35, 921 - 927 .

Taeusch,D.R., 1977, Structure of electrodynamic and particle heating in the disturbed polar thermosphere, J.Geophys.Res., 82, 455 - 460 .

Taeusch,D.R., G.R. Carignan, and C.A. Reber, 1971, Neutral composition variation above 400 kilometers during a magnetic storm, J.Geophys.Res., 76, 8318 - 8325 .

Tanaka,T., and K. Hirao, 1973, Effects of an electric field [illegible] namical behavior of the ionospheres and its application to the storm time [illegible] e of the F-layer, J.Atmos.Terr.Phys., 35, 1443 - 1452 .

Thomas,L., and F.H. Venables, 1966, The onset of the F-regio [illegible] ance at middle latitudes during magnetic storms, J.Atmos.Terr.Phys., 28, 599 - 605 .

Volland,H., 1979, Magnetospheric electric fields and currents and their influence on large scale thermospheric circulation and composition, Rev.Geophys.Space Phys., 17, (in press).

Von Zahn, U., W. Kohnlein, K.H. Fricke, U. Laux, H. Trinks, and H. Volland, 1977, ESRO 4 model of global thermospheric composition and temperatures during times of low solar activity, Geophys.Res.Lett., 4, 33 - 36 .

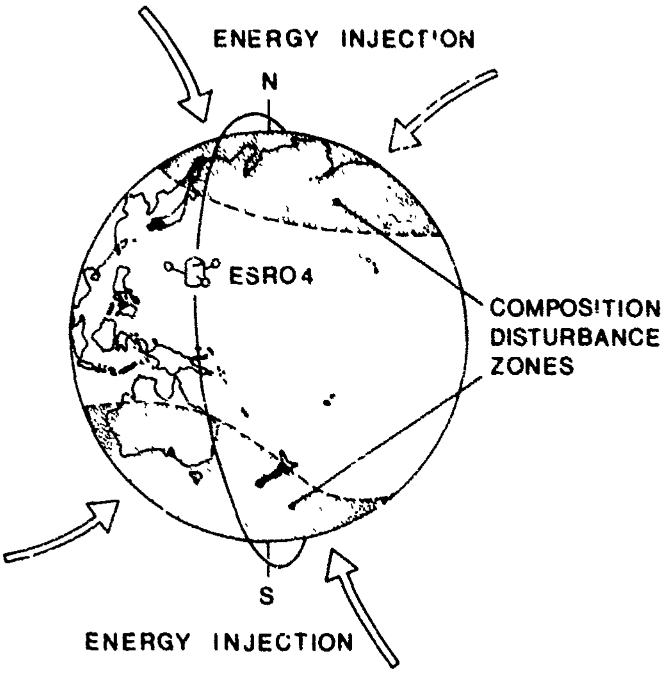

ENERGY INJECTION
N
ESRO 4
COMPOSITION DISTURBANCE ZONES
S
ENERGY INJECTION

FIGURE CAPTIONS

Fig. 1 Energy injection and formation of neutral atmospheric composition disturbance zones during a magnetospheric storm event. Also shown is a sample orbit of the ESRO 4 satellite which provided all neutral composition data shown in the following figures (Prölss and von Zahn, 1977a).

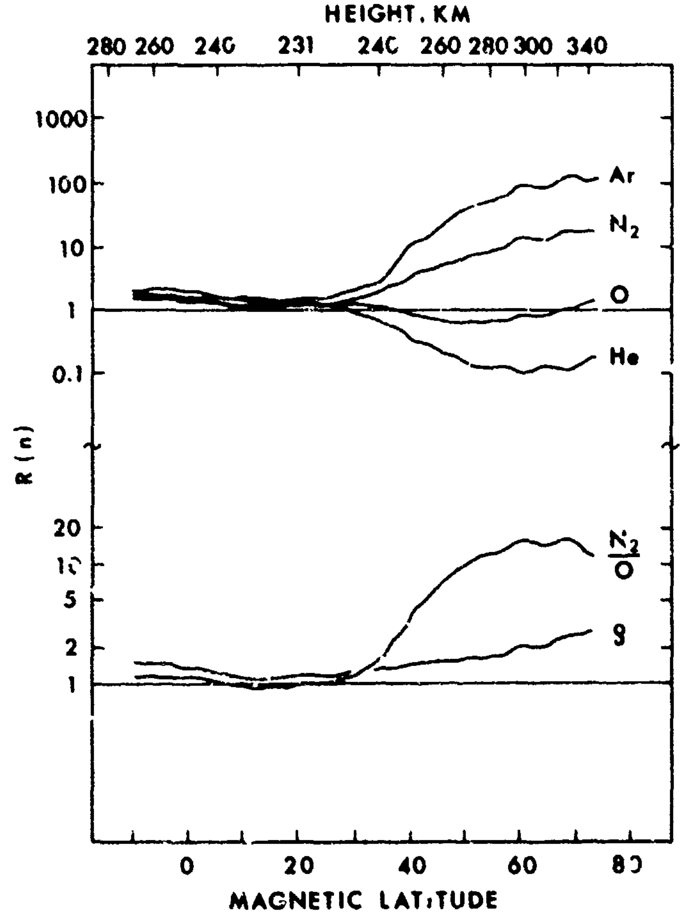


Fig. 2 Basic latitudinal structure of magnetic storm induced changes in the number density of argon (Ar), molecular nitrogen (N_2), atomic oxygen (O), and helium (He). In addition the lower part shows the corresponding variations of the atomic oxygen to molecular nitrogen concentration ratio (O/N_2) and of the total mass density (ρ).

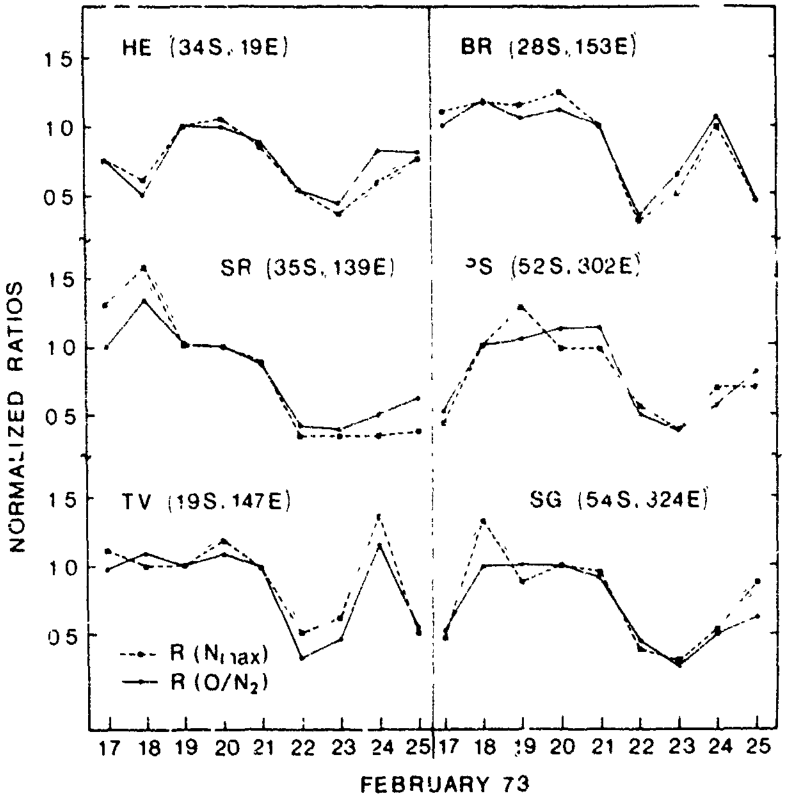


Fig. 3 Relative day-to-day variations in the F2-layer maximum electron density (N_{max}) and in the atomic oxygen to molecular nitrogen (O/N_2) concentration ratio above six mid-latitude stations (positions in geographic coordinates) during a nine-day interval in February 1973.

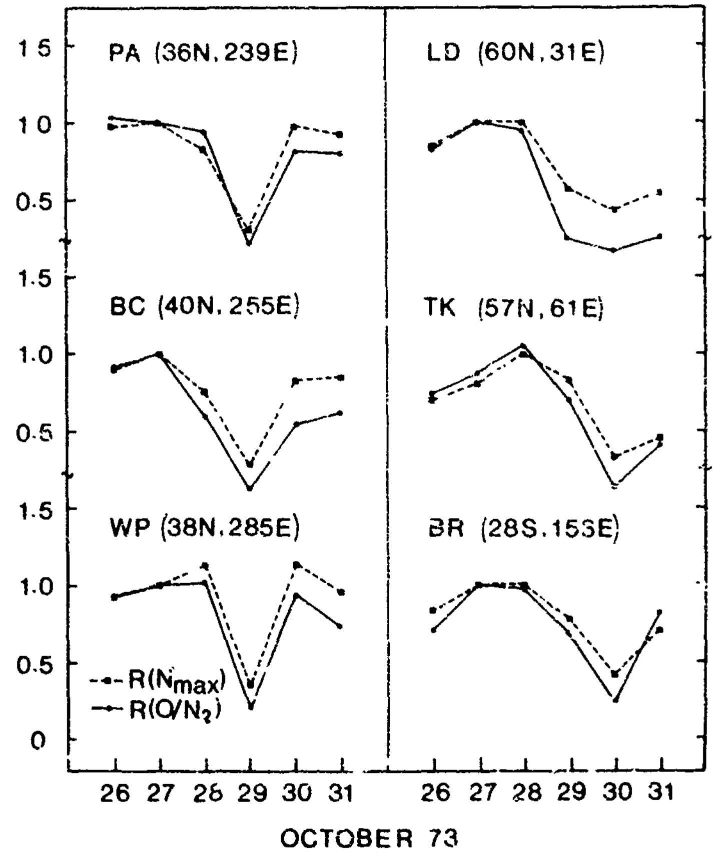


Fig. 4 Some as Fig. 3 for a six-day interval in October 1973.

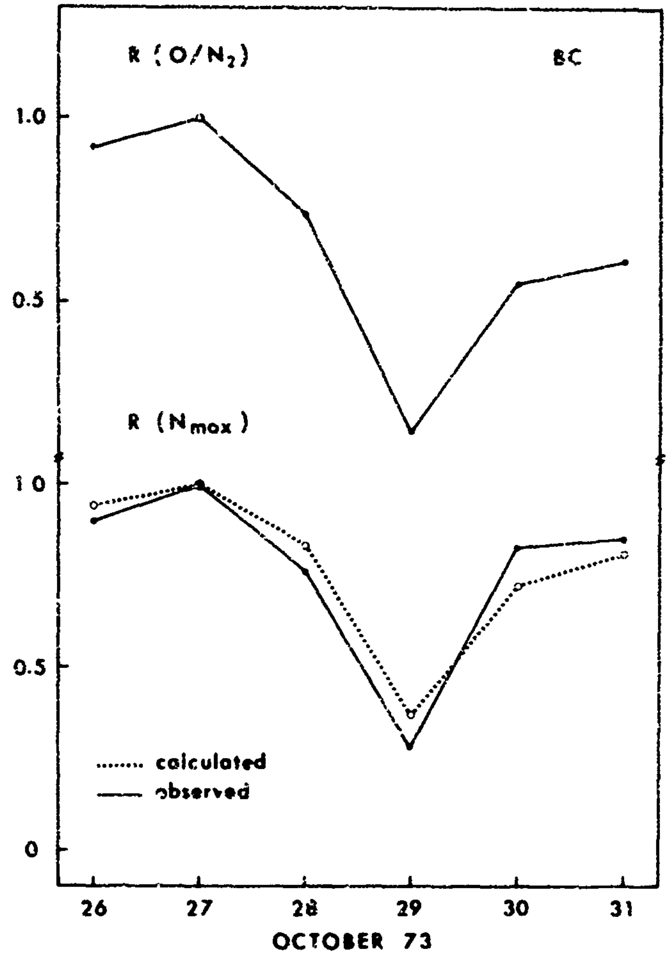


Fig. 5 Comparison between calculated and observed day-to-day variations in the F2-layer maximum electron density N_{max} at Boulder/USA (lower part). The upper panel shows the corresponding changes in the measured O/N_2 concentration ratio used in the model calculation (Jung and Prölss, 1978).

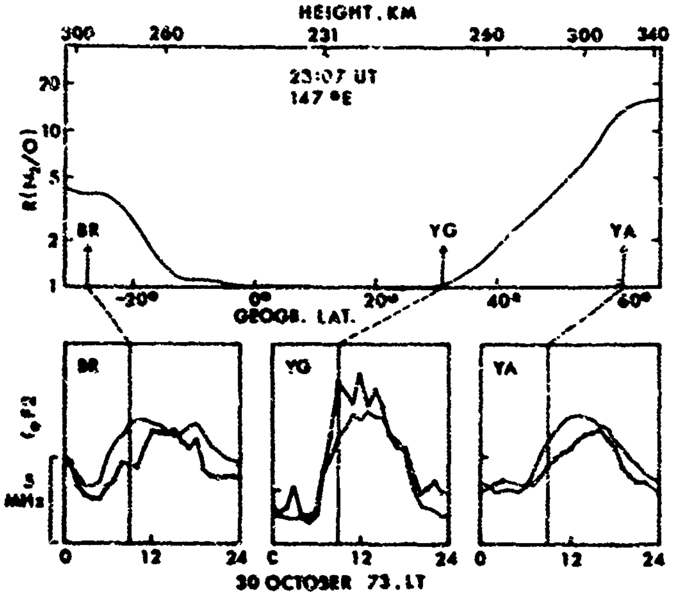


Fig. 6 Comparison between the latitudinal structure of neutral atmospheric and ionospheric disturbance effects. The upper panel shows relative changes in the O/N_2 concentration ratio as observed close to 147°E on October 30, 1973. The lower part shows the local time variation of the F2-layer critical frequency (fat dotted line) as observed at three ionosonde stations whose relative positions with respect to the atmospheric disturbance zone are indicated by arrows. Also shown for comparison is the monthly median of f_oF2 for October 73 (thin line) (Prölss, 1977).

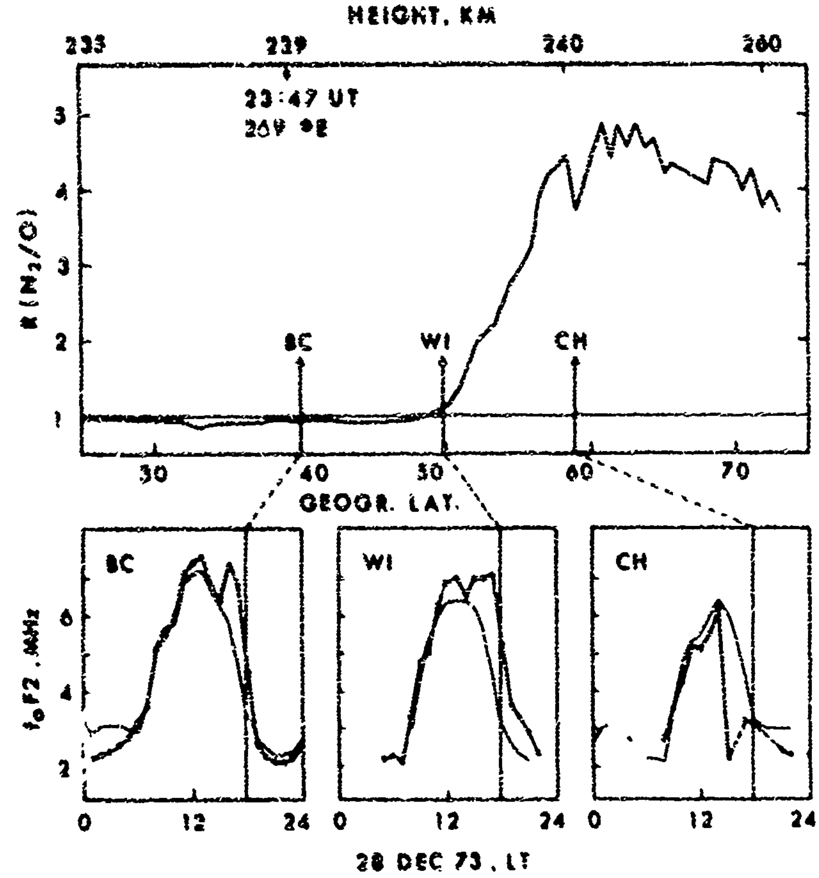


Fig. 7 Same as Fig. 6 for magnetic storm event of 28 December 1973.

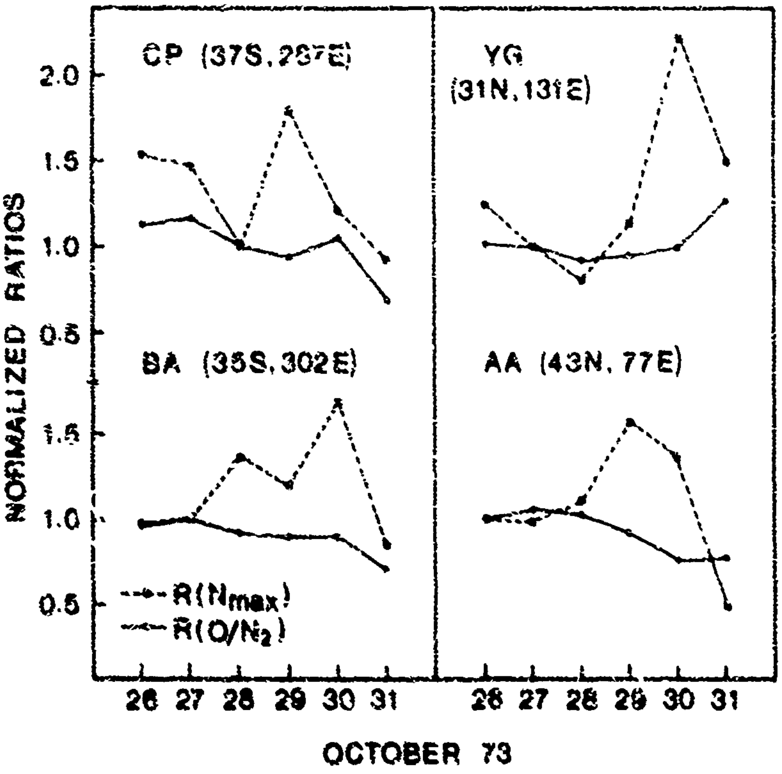


Fig. 8 Relative day-to-day variations in the F2-layer maximum electron density (N_{max}) and in the atomic oxygen to molecular nitrogen (O/N_2) concentration ratio above four mid-latitude stations (positions in geographic coordinates) during a six-day interval in October 1973. Note that in contrast to Fig. 4 all stations were located outside the neutral composition disturbance zone.

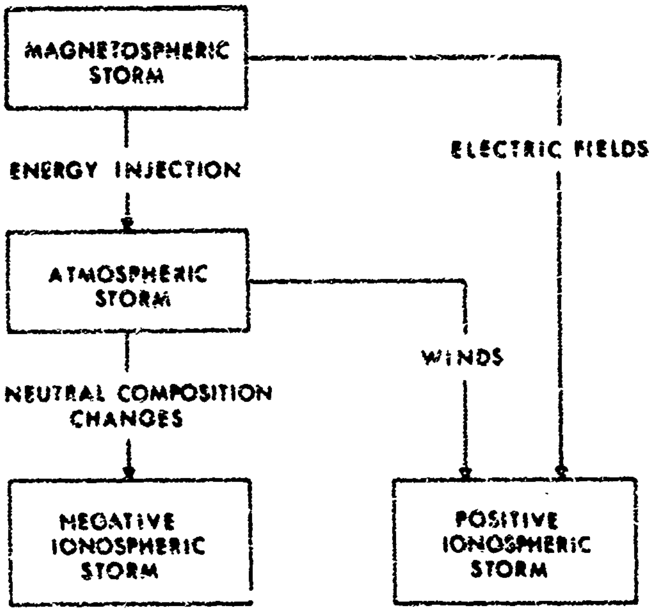

MAGNETOSPHERIC STORM
ENERGY INJECTION
ELECTRIC FIELDS
ATMOSPHERIC STORM
WINDS
NEUTRAL COMPOSITION CHANGES
NEGATIVE IONOSPHERIC STORM
POSITIVE IONOSPHERIC STORM

Fig. 9 Causal relation between magnetospheric, neutral-atmospheric and ionospheric disturbance effects.

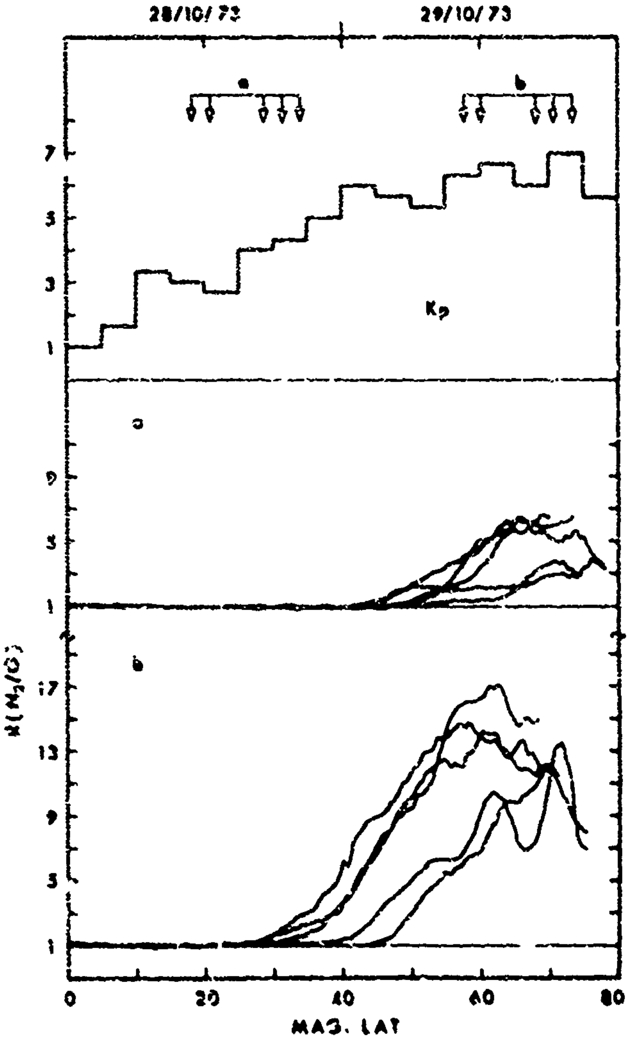


Fig.10 Variation of the latitudinal structure of an atmospheric disturbance with magnetic storm intensity. Shown is a comparison between latitudinal profiles of relative changes in the N_2/O concentration ratio (lower part) observed during two different phases (a and b) of a magnetic storm whose time variation is indicated in the upper part of the figure.

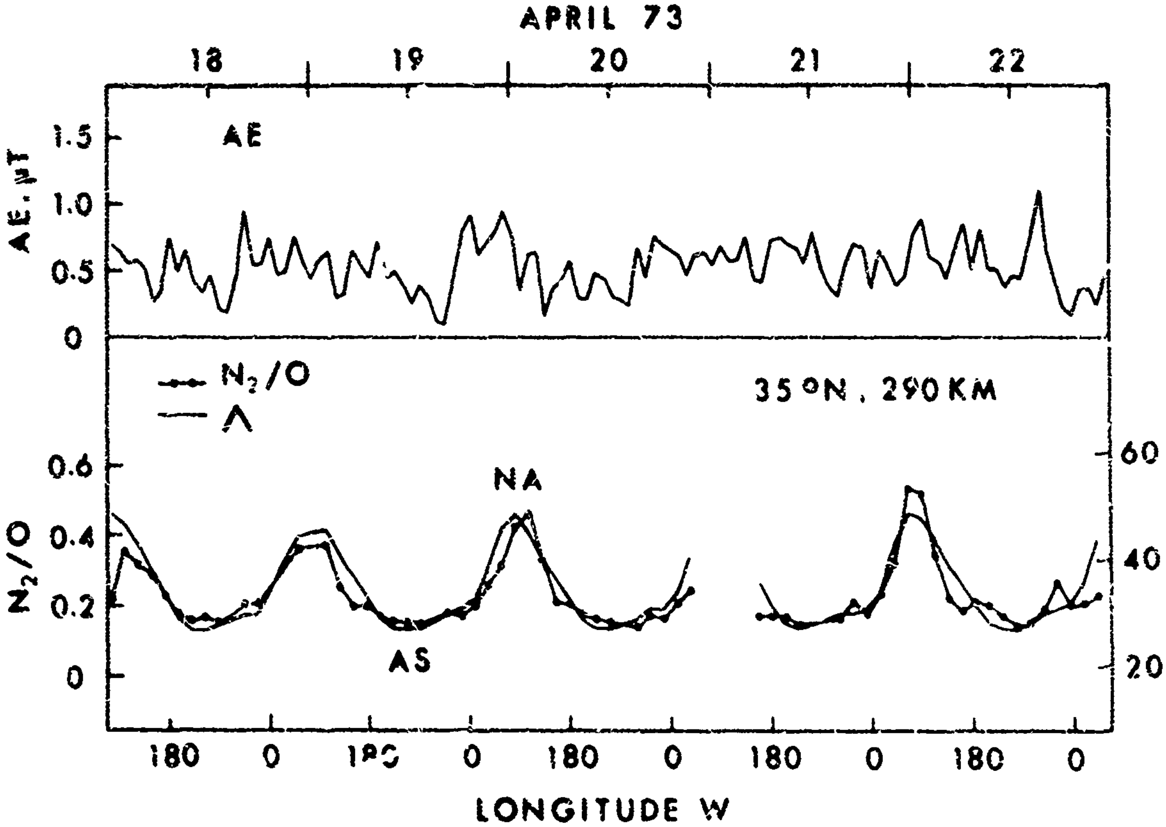


Fig.11 Comparison between the longitudinal variations of the N_2/O density ratio and the magnetic invariant latitude Λ along a constant geographic latitude (lower panel). The level of magnetic activity (AE index) during the time interval of interest is shown in the upper panel.

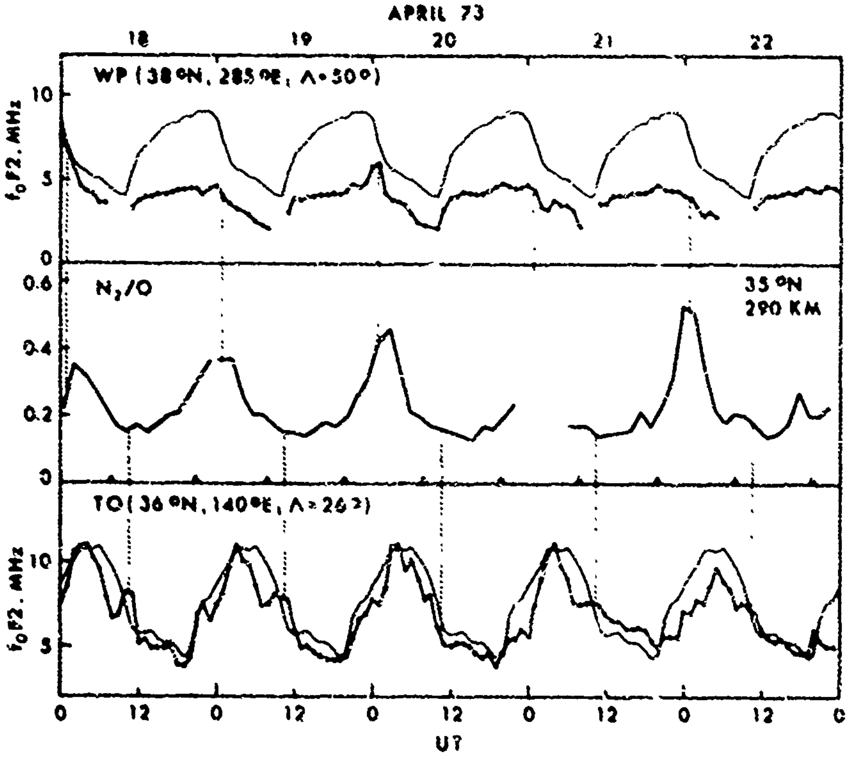


Fig.12 Comparison between the longitudinal variation of the N_2/O density ratio and the behavior of the critical frequency as observed at two stations located near the maximum and near the minimum of the composition disturbance, respectively. The thin line in the upper and lower panel indicates the mean variation of f_oF2 during five undisturbed days.

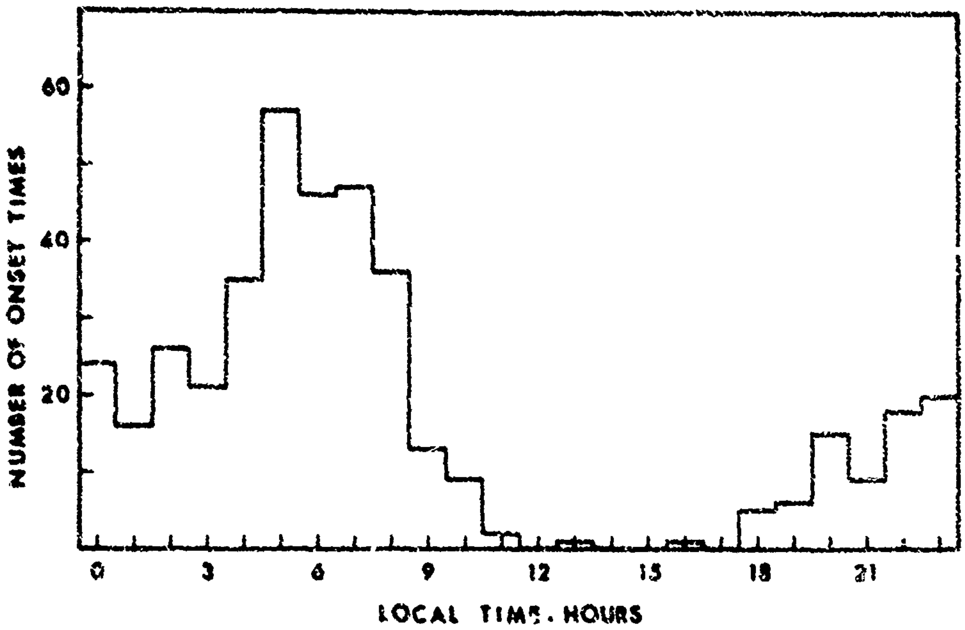


Fig.13 Local time distribution of onset times of negative ionospheric storm effects as observed at some 24 mid-latitude stations during 1973 (Prölss and von Zahn, 1978).

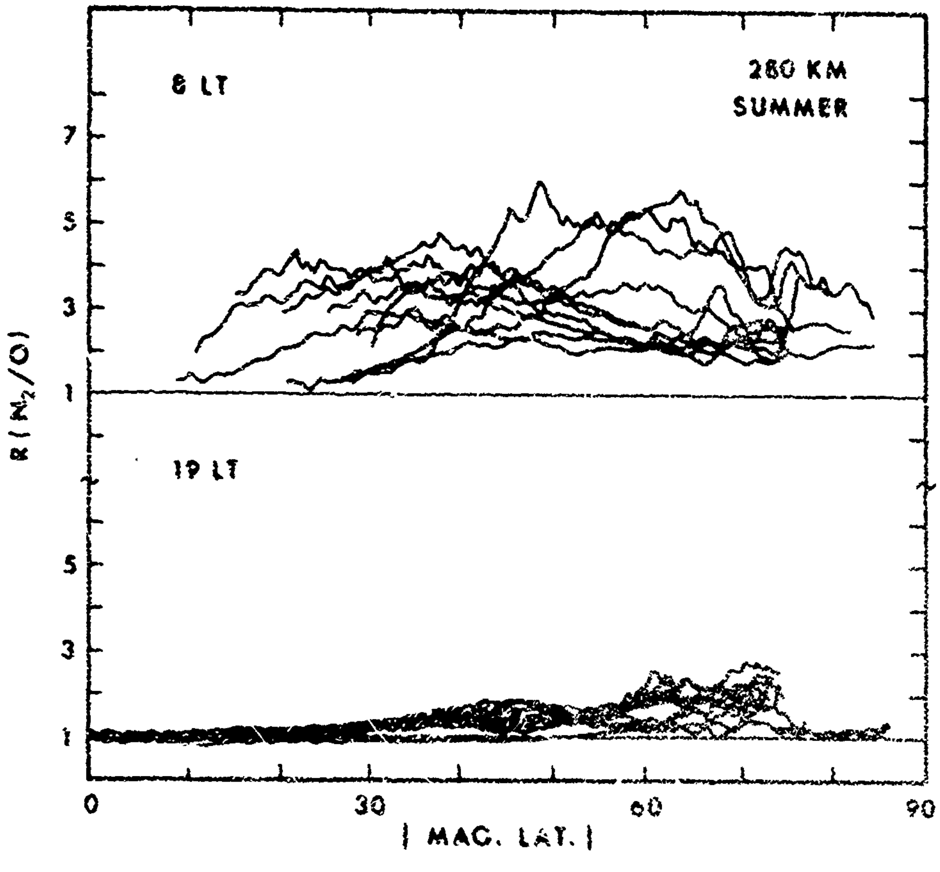


Fig.14 Comparison between the magnitude of two atmospheric disturbances (indicated by changes in the N_2/O density ratio), one being observed in the morning (8LT, 21/22 Nov 73, Kp_{max}=6-) and one being observed in the evening sector (19LT, 4/5 Dec 73, Kp_{max}=6). (Prölss and von Zahn, 1978).

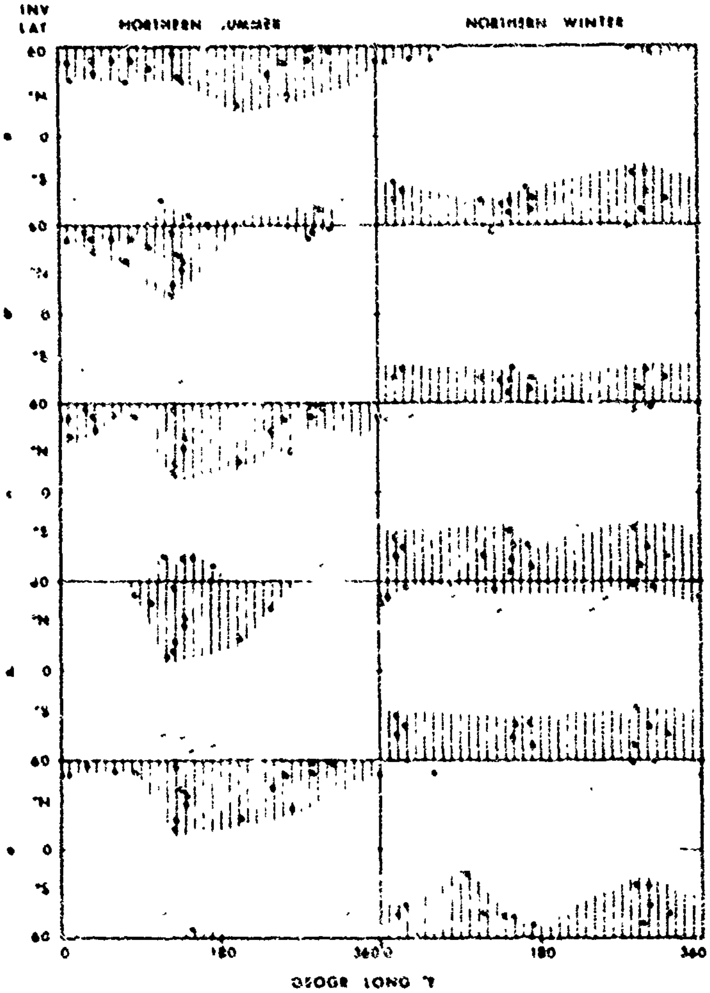


Fig.15 Extension of ionospheric disturbance zones (hatched areas) in a geographic longitude/invariant latitude coordinate system during five northern summer and five northern winter events. The solid, semisolid, and open circles indicate that significant, small and no negative storm effects are observed at these locations, respectively (Prölss, 1977).

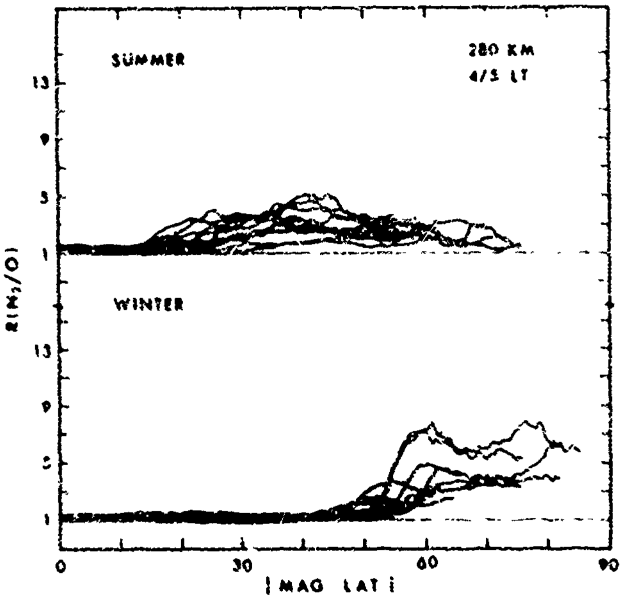


Fig.16 Comparison between the latitudinal structure of a summer (June 24, 1973; $Kp_{max}=5$) and a winter disturbance event (June 2-3, 1973; $Kp_{max}=4$). (Prölss and von Zahn, 1977).

DISCUSSION

E.R. Schmerling, US

I congratulate the author on emphasizing the important point that ionospheric behavior cannot be understood without understanding the behavior of the neutral ionosphere. I question, however, his conclusion that electric fields are responsible for magnetic storm effects at low latitudes, whereas, the neutral composition governs changes at high latitudes, since the strongest electric fields have been observed just in the high latitude regions.

Author's Reply

There can be no doubt that electric fields are strongest in the high-latitude regions. In fact it is the dissipation of these electric fields in the polar regions which causes most of the atmospheric-ionospheric disturbance effects, e.g. neutral composition changes, traveling atmospheric disturbances and waves. However, a small fraction of these substorm electric fields will also penetrate to middle latitudes where it may cause positive storm effects especially in the afternoon/premidnight local time sector.

H. Soicher, US

Observations of the effects of magnetic storms on total electron content (TEC) have shown that depending on location and local time shortly after sudden commencement (SC) there is a positive phase (i.e. TEC increases sharply). This is followed by a negative phase which may last up to a few days.

How are these explained in view of the author's neutral composition observations? Is there interaction between neutral composition changes and ionospheric transport processes?

Author's Reply

Before answering your question I would like to point out that the sudden commencement (SC) is not related to any of the large-scale atmospheric-ionospheric disturbance effects and its use as a reference point should certainly be discontinued.

Within the framework of the model presented in this talk, positive storm effects primarily observed in the afternoon sector during the initial phase of major substorm activity are caused by large scale traveling atmospheric disturbances (TADs) and to a lesser degree by substorm electric fields which transport ionization to higher altitudes and therefore into regions of reduced losses. During this time composition changes are also generated, however, they are restricted to the high-latitude region on the dayside and only in the night sector especially in the early morning sector do they extend to middle latitudes. Ionosonde stations which are located in the early morning sector while a magnetic storm is in progress will therefore immediately observe negative storm effects. On the other hand stations which are initially located in the afternoon sector will first have to rotate through the night sector before negative storm effects are observed.

BASIC FINDINGS HELPFUL FOR IONOSPHERIC PREDICTIONS

K. Harnischmacher/K. Rawer
Ionosphären-Institut
D-7814 Breisach, F.R.G.

SUMMARY

A large part of the day-by-day fluctuation of ionospheric parameters seems to be due to lunar influences. These are much larger than revealed by long term harmonic analysis since they appear preferentially at certain solar hours and depend on the season. Data from Freiburg (summer periods) show that different years compare well provided Jupiter and Venus are far from Earth; the lunar effect is deformed during other periods. Such effects were found for the critical frequency foF2 of the F2-layer, but also - to a smaller extent - for foEs are another cause of fluctuations during certain periods. Also, the efficiency of a magnetic disturbance upon foF2 seems to be smaller, when Jupiter and Venus are far from Earth.

1. INTRODUCTION

For many decades ionospheric predictions had monthly median conditions for aim. Since ionospheric stations used to observe in a day-by-day, hour-by-hour schedule the large variability of most ionospheric parameters, in particular foF2 and foEs, is well known. The monthly median diurnal variation used to be taken as a systematic effect, at a given location still depending on season and (average) solar activity. Efforts to link the day-by-day fluctuations with that of indices of solar activity were in vain - with the exception of foE, the critical frequency of the normal E-layer (BIBL, K., 1951). Therefore the fluctuations were interpreted as 'noise' (RAWER, K., 1944 and 1947), i.e. as being provoked by not-identifiable, thus unpredictable causes.

These fluctuations are quite large, the range of variability being ± 15...20 % for foF2 in the temperate zone; it is even larger for foEs, the top-frequency of the sporadic layer Es (which is important for realistic predictions of the maximum usable frequency, at least at distances up to 2000 km).

We should note here that all influences which are synchronized with the Sun do not produce 'fluctuations' in our sense since the average diurnal variation is determined by purely empirical methods. Thus, solar tides, may they be produced by heating or by gravity, cannot produce more spread in monthly massplots than is due to the systematic variation of the Sun's apparent orbit from the beginning to the end of a month. Lunar tides, however, which can be supposed as to be of gravitational origin, are not included. These cause a part of the fluctuations the importance of which shall be discussed in the following.

2. LUNAR TIDES IN THE F-REGION

After D.F. MARTYN's (1947) initial investigations of lunar tides at the magnetic equator, J. BARTELS (1950) analyzed data from Huancayo (12°S, 75° W, magnetic dip + 2°) and found a quite important effect in the critical frequency foF2 of about 1 MHz in October and November, but a much smaller effect in June and July. Thus, at the dip equator, a sometimes considerable part of the fluctuations might be due to lunar tides. Using the S. CHAPMAN and J. BARTELS (1940) equation of the luni-solar variation O. BURKARD (1951) gave an idealized description of the lunar effect at Huancayo in terms of four harmonics. He was first to see that lunar effects tend to appear preferentially at certain (solar) hours. A comprehensive compilation of consecutive work on lunar tides at many stations is due to S. MATSUSHITA (1967). He found 'remarkable differences of the behaviour between the magnetic equatorial zone and the region away from this zone', average amplitudes in the first being about twice as large than in the second. His survey paper considers different ionospheric parameters while we shall concentrate in the following mainly on the critical frequency foF2. For this characteristic, M. BOSSOLASCO and A. ELENA (1960) harmonically analyzed data from stations of particular concern to our study: Freiburg (48.1°N), Genova (44.6°N) and Léopoldville (now Kinshasa, 4°S). They obtained amplitudes of the semidiurnal lunar tide of foF2 of 0.04, 0.06 and .08 MHz, respectively. This is quite small compared with the fluctuation shown in monthly mass plots of foF2.

With respect to such effects, classical Fourier analysis over longer time intervals is not really helpful. This is due to the fact that the Sun-Moon configuration is continuously varying. Also the diurnal period of illumination varies with the season and excites itself a season-dependent semidiurnal variation. In a correct analysis the Sun-dependent influences should be hold constant. We do this by following the seasonal variation in intervals of two weeks only. On the other hand, in order to avoid statistical destruction of lunar effects (as applied voluntarily in monthly medians) we adjust these intervals to the lunar period and phase. Beginning and end of each basic interval is always at a well-determined lunar phase. This has for consequence that by allowing for a slight shift in season (a few days only) we obtain "lunar synchronization" enabling us to compare lunar effects at a given season in different years.

2.1. Handling of Peak Electron Density Data

When looking for possible causes of these fluctuations at our mid-latitude station Freiburg, we selected summer periods from May through August, the average (solar) diurnal variation being smallest during that season. Further, in order to avoid the remaining seasonal variations, we rejected the usual methods tending to determine an average lunar variation over long periods. Considering BARTELS' (1950) results for Huancayo (at the magnetic equator), we grouped data, for only one month from different years however, synchronizing, systematically by the lunar phase. When doing so we found that consistent results were obtained when combining certain years, while other combinations were much less satisfying.

We are now rather certain that planetary influences are the cause of this queer finding. When taking account of the two planets which should have the largest tidal effect, namely Jupiter and Venus, we obtain a consistent picture from years with similar planetary constellation. Since the tidal acceleration provoked by these planets is much smaller than that of the Moon, it remains difficult to understand this finding with a linear theory. We presume that planetary waves and oscillations play their part in the game such that it is not indifferent whether Jupiter has its culmination (or anti-culmination) with the Sun, or at another solar hour. In fact we shall see below that these effects get some influence always when one of the two planets is near Earth.

After this 'planetary selection' we obtain consistent behaviour for summer months, the lunar effect appearing most clearly when both planets are far away or are aligned with the Sun. We feel that, with this knowledge, a large part of the so-called fluctuation of foF2 is in fact becoming predictable.

2.2. Filtering by Appropriate Smoothing

It is known that short period events sometimes influence the ionosphere. In order to eliminate such undesired features, we applied a cautious, two step smoothing procedure to the measured foF2-data: (1) smoothing over the date by averaging over three consecutive days (with weights 1:4:1); (2) smoothing over the season by averaging over corresponding days of opposed lunar phase, i.e. the day 14d before and 14d after the considered day (with weights 1:2:1). After this filtering the individual curves are quite smooth and - except for a few exceptions - show systematic changes with the semi-monthly lunar phase. The semi-diurnal component uses to appear more distinctly for new and full Moon, i.e. when the Moon is in phase or antiphase to the Sun.

In order to demonstrate the lunar influence better we applied one more filtering procedure by averaging over corresponding days of two consecutive years, taking account of the lunar phase as indicated above. It appeared then very clearly that the so-called irregular fluctuation is to a large extent a systematic influence of the Moon. Examples are shown in Fig. 1a, where the so filtered diurnal curves of foF2 are shown depending on the lunar phase. In order to show that magnetic disturbances could not have seriously influenced the picture we show daily Ap-indices (averaged after the same schedule) at the right hand side. The two peak type of foF2 appears preferentially around full and new Moon, but tends to disappear around the half Moon phase. Similar behaviour is besides shown by ocean tides under certain conditions (see Fig. 1b).

Exceptions occured for strong magnetic disturbances and, quite generally, for the year 1969. At first, we attributed this to the high solar activity of this particular year. Meanwhile we have another explanation.

2.3 Planetary Influences

Realizing that the orbital period of Jupiter (about 12 years) is quite near to that of the sunspot cycle, we decided to check this planets influence. This is not easy because the influence is probably weak against those of Sun and Moon. First of all, we identified years where, due to the configuration, the jovian influence should not be disturbing the lunar one. These are conditions where Jupiter and Sun are in phase (see class (1) below) or antiphase (see class (2) below) such that the tidal game is essentially reduced to the two main participants only, namely Sun and Moon. Note that conditions in Fig.1a were so chosen.

At least during periods when Venus is near Earth this planets effect must also be taken into account. We deal with this in class (3) below.

Under conditions where Jupiter and Venus are of comparable effect and out of phase, our procedure breaks down. This situation occurred in 1969, see class (4) below.

We shall now present the four typical classes. We illustrate them by a few examples embracing only a small part of all results we have obtained from 24 summer seasons.

Class (1) Jupiter and Venus far away

Jupiter culminates between 10 and 15 h local time (LT) such that the jovian influence is almost synchroneous with that of the Sun. This condition occurred in 1954 and again in 1966. In Figs. 2 data from summer months in these years are shown. In Figs. 2c and d Jupiter and Sun were culminating rather exactly at the same hour, the effect of lunar phase was much more pronounced, and so was the importance of the half-day wave, than in Figs. 2a, b and e where Jupiter was out of phase by 2 h and more. This shows that the half-day wave depends critically on the phase of the planet, even when this latter is far away.

Class (2) Jupiter near Earth, Venus far away

Seen from Earth, Jupiter is in opposition to the Sun, the culmination occuring between 17 and 08 h LT. These conditions are fulfilled in the following summer seasons: 1959 (except July, August), 1960, 1961 (except May, June), 1962 and in the next orbital period: 1971 and 1973. (Summer 1972 was a comparable case: Though Venus was near Earth, Jupiter and Venus culminated about at the same hour.) When considering the same examples shown in Figs. 3, we see again that the lunar effect is clearer for culmination with the Sun (Fig. 3a) than when the phase difference is large. In the latter case the lunar effect is quite poor (Fig. 3b), and so is the importance of the half-day wave. Under such conditions we have sometimes seen a 6 h modulation (the first harmonic of the 12 h period), too.

Class (3) Jupiter far away, Venus near Earth

In this case Venus is prevailing. Since it is an inner planet, the culmination uses to occur around noon (10...14 h LT). Whenever it differs by less than 1 h from that of the Sun, the lunar modulation is very clearly appearing as shown in Fig. 4, quite similar to the undisturbed lunar effect as shown in Fig. 1a.

Class (4) Odd conditions. In summer 1969 Venus and Jupiter had comparable influence and culminated at different hours, away from noon. In these conditions the regular lunar effect is seriously disturbed (Fig. 5). These are still effects of the lunar phase. However, the pattern around full Moon is more like that found near half-Moon in the other classes. Higher harmonics are more important here.

3. LUNAR TIDES IN THE E-REGION

This subject is also delt with by S. MATSUSHITA (1967) with results of classical harmonic analysis.

3.1 The Critical Frequency of Normal E, foE

"Normal E" is known to be the most regular ionospheric layer. Except for small deviations in the morning and evening the critical frequency, foE, follows true local solar time so accurately that the equation of time can empirically be derived from the symmetry hour of the diurnal curve (HARNISCHMACHER, E., 1950). E.V. APPLETON and A.J. LYON (1961) have shown that there is a small noon-shift at most stations which can be positive or negative, but R. EYFRIG (1962) was able to identify certain positions on the globe where this shift is zero. Where measurements are made correctly, the total dispersion range appearing in monthly mass-plots is about ±6% (RAWER, K. and SUCHY, K., 1967). Slightly larger deviations, always positive, occur during solar flares (BIBL, K., 1951). Anyway, the fluctuations of foE remain so small that account of them is usually not taken in practical applications. The diurnal variation can therefore be described accurately enough with a simple law depending on the solar zenith angle, e.g. R. EYFRIG's (1960).

3.2 The Sporadic E-Layer

This layer is a particular case. It is much thinner than a normal layer and occurs irregularly so that only statistical forecasts can be obtained (RAWER, K., 1955, 1962b). It is often structured in clouds so that a critical frequency in the usual sense is not defined. One identifies the range of transparency by two characteristics, fbEs and foEs (RAWER, K., 1962a; PIGGOTT, W.R. and RAWER, K., 1961, 1972).

3.2.1 The Blanketing Frequency, fbEs,

is defined as the lowest sounding frequency on which transparency of the Es-layer became apparent. It characterizes the conditions in the first Fresnel-zone above the station and corresponds, roughly speaking, to the minimum of the peak electron density in that zone. In many cases, where Es appears on ionograms, fbEs is identic with foE; this means that Es appears as isolated clouds without a connecting 'stratum'. We shall not consider this characteristic further.

3.2.2 The Top-Frequency, foEs,

is defined as the highest sounding frequency on which an essentially continuous Es-trace (ordinary or ionic component) was observed. It characterizes the quasi-maximum of the peak electron density occuring somewhere in a large cone around the vertical. Therefore, in-situ measurements use to compare better with fbEs than with foEs (BROSS et al., 1967). Nevertheless foEs is a very important characteristic for two reasons: (1) geo physically it demonstrates the horizontal inhomogeneity of the Es 'cloud-cover'; (2) it often controls the maximum usable frequency for radio wave propagation (since the waves are reflected or scattered from isolated Es-clouds).

At temperate latitude Es occurs preferentially in summer. Its statistics reveals that it mustbe aquasi-meteorological phenomenon (RAWER, K., 1955). Though experimental proofs are not yet final present theories try to explain the phenomenon as being due to vertical plasma redistribution under the influence of wind shears. J.D. WHITEHEAD's (1961, 1962) theory requires a shear in the zonal wind of a particular sign while W.I.AXFORD's (1963) theory admits any sign of the shear.

At our latitude Es occurs in summer on almost all days rather regularly, at least at day-light hours, but the values of the top frequency are largely and sometimes quickly variable. Taking account of our experience with foF2 we analyzed data from five summer periods for which the planetary conditions were easy to characterize. The same smoothing procedures were applied as described in Sect. 2.2 above.

The results of our preliminary analysis are shown in Figs. 6a and b. Fig. 6a is for the most easy case where the planetary influence can be neglected since Venus was far away and Jupiter was culminating with the Sun. Our synchronized analysis superposes two-week periods from May to July in the years 1965, 1966, 1971 and 1972 thus covering about 40 weeks alltogether. In this 'package' a lunar modulation, similar to that found for foF2, is clearly apparent: an afternoon maximum appears around full and new Moon only. The second maximum is well visible on four days. There is only one (the forenoon) maximum for half-Moon conditions. It shows a shift in time corresponding rougnly to constant lunar hour. An amplitude of more than 1 MHz is reached but only around 15 h. The appearing of an half-day wave is quite similar to the features shown in Figs. 1.

Another result is obtained when one of the planets is culminating off from noon. This result does not compare with those for foF2. In Fig. 6b we have the planetary culmination in the middle of forenoon, between 8 and 9 h LT. The total set of data is about of same importance as that used in Fig. 6a, namely from June to August in 1965, 1971, 1972 and 1976. A weak second maximum appears only at three days near full and new Moon. The first (forenoon) maximum, however, is now modulated by about .5 MHz in amplitude, the greatest values aoccuring at half Moon, quite different from Fig. 1a. We suppose that this is also a planetary influence.

4. EFFICIENCY OF MAGNETIC DISTURBANCES

At temperate latitude magnetic disturbances are known to have a very pronouced effect on the F2-layer, considerably decreasing foF2. During such periods the lunar effect is overcome by that of the disturbance. In the previous chapters we have eliminated this undesired interference by superposing several intervals of two weeks each. When considering many magnetic disturbances, we found that their influence on the ionosphere was apparently different from year to year. The 'efficiency' of a magnetic disturbance was small under conditions where Jupiter and Venus were far away and much larger when they were near Earth. Restricting our analysis to summer again, we found that in addition to the sunspot cycle, the probability of a magnetic disturbance to appear seems to be larger when Sun, Moon and planet Jupiter and/or Venus are nealy aligned.

5. CONCLUSIONS

We have no satisfying theory allowing to deduce the observed phenomena. The tidal acceleration varies with mass divided by the cube of the distance. Compared with that of the Moon, the contribution of Venus varies between about $0.2\cdot10^{-6}$ and $55\cdot10^{-6}$, the corresponding values being $1.8\cdot10^{-6}$ and $6\cdot10^{-6}$ for Jupiter. In a straight forward (linear) theory the corresponding acceleration vectors are just to be added. On the other hand, one might consider indirect effects like the disturbance of the lunar orbit provoked by the planets, obeying a law with the inverse square of the distance. Anyway, we cannot see at present how so small contributions might have a visible effect.

On the other hand, since we feel that we have now empirically established planetary effects upon foF2, even on foEs and potentially on the efficiency of magnetic disturbances, we suppose that some specific mechanism is in the game. This could for example be a resonance which might be perturbed by small additional forces attacking out of the phase. It is probable that there exist complicated wave patterns in the upper atmosphere which might be more easily triggered by the lunar forces when planetary disturbances are absent.

There are, of course, other causes of fluctuations for which no long term prediction is feasible: (1) the changes occuring during magnetic disturbances, which last in most cases only two days; (2) the quick (20 min) fluctuations also known as 'travelling ionospheric disturbances' stem probably from internal gravity waves in the upper atmosphere. Our investigation shows, however, that the regular tidal variations which occur mainly with a period of two weeks, are not only quite important (at least at certain hours) but can also be included in daily forecasts. This has been shown for summertime foF2 at Freiburg, to some extent also for foEs.

REFERENCES

APPLETON, E.V. and LYON, A.J., 1961, "Studies of the E-layer of the Ionosphere II", J. Atmosph. Terr. Phys. 21, 73-99.

AXFORD, W.I., 1963, "The formation and vertical movement of dense ionized layers in the ionosphere due to neutral wind shears", J. Geophys. Res. 68, 769-779.

BARTELS, J., 1950, "Ebbe und Flut in der Ionosphäre", Ber.dtsch.Wetterdienstes US-Zone, Nr. 12, 30-33.

BIBL, K., 1951, "L'ionisation de la couche E, sa mesure et sa relation avec les éruptions solaires", Ann. Géophys. 7, 208-214.

BOSSOLASCO, M. and ELENA, A., 1960, "On the lunar semidiurnal variation of the D- and F2-layers", Geofis. pura e appl. 46, 167-172.

BROSS, P., JACOBS, K.-G. and RAWER, K., 1967, "Direct electron density measurements in the ionosphere below 200 km", Space Res. VII/1, 467-476.

BURKARD, O., 1951, "Studie zum ionosphärischen Gezeiteneffekt", J. Atmosph.Terr. Phys. 1, 349-352.

CHAPMAN, S. and BARTELS, J., 1940, "Geomagnetism I", Oxford: Claredon Press (Chapt. VIII, pp. 247 ff.).

EYFRIG, R., 1960, "Kritische Bemerkungen zur Beschreibung der Ionisation der ionosphärischen E-Schicht", Geofisica pura e appl., 45, 62-64.

EYFRIG, R., 1962, "Non-seasonal variation in the E-layer ionization", Nature (London) 196, 750-760.

HARNISCHMACHER, E., 1950, "L'influence solaire sur la couche E normale de l'ionosphère", R.C. Acad.Sci. Paris 230, 1302.

MATSUSHITA, S., 1967, "Lunar tides in the ionosphere", Handbuch der Physik 49/2, 547-602, Berlin: Springer-Verlag.

MARTYN, D.F., 1947, "Atmospheric tides in the ionosphere. II. Lunar tidal variations in the F-region near the magnetic equator". Proc. Roy. Soc. London A 190, 273-288.

PIGGOTT, W.R. and RAWER, K., 1961, "URSI Handbook of Ionogram Interpretation and Reduction", Amsterdam: Elsevier. Second edition 1972, Asheville (N.C.): World Data Center A for Solar-Terrestrial Physics, Report UAG-23.

RAWER, K., 1944, "Eine Kennwert-Methode zur Vorhersage ionosphärischer Grenzfrequenzen", Dtsch.Luftfahrtforschung F.B.Nr. 1953.

RAWER, K., 1947, "La prévision des fréquences critiques de l'ionosphère", Revue Sci. 85, 234-235.

RAWER, K., 1955, "Unregelmäßigkeit und Regelmäßigkeit der sporadischen E-Schicht, Geofisica pura e applicata 32,170-224.

RAWER, K., 1962, "Definition of frequency parameters of Es-layers and their accuracy", [1], 151-165.

RAWER, K., 1962, "Structure of Es at temperate latitudes", [1], 292-243.

RAWER, K. and SUCHY, K., 1967, "Radio-observations of the ionosphere", Handbuch der Physik 49/2, 1-546.

WHITEHEAD, J.D., 1961, "The formation of the sporadic E-layer in the temperate zones", J. Atm.Terr. Phys. 20, 49-58 [change of sign required for y in eqs. (1) and (2)].

WHITEHEAD, J.D., 1962, "The formation of a sporadic-E layer from a vertical gradient in horizontal wind", [1], 176-291.

[1] SMITH, E. and MATSUCHITA, S. (eds.), 1962, "Ionospheric sporadic E", London: Pergamon Press.

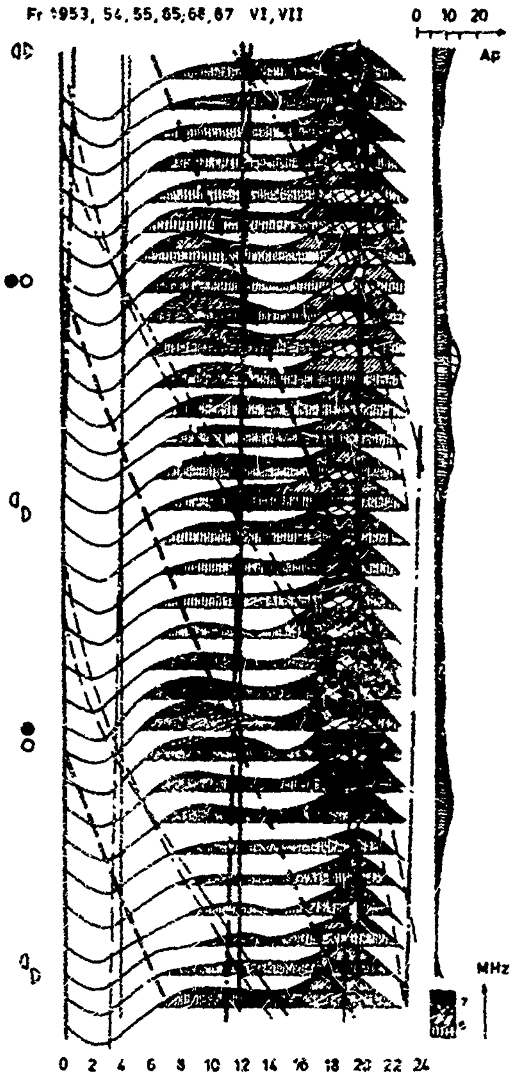


Fig. 1a Daily curves of foF2 (smoothed as explained in Section 2.2), compiled with lunar synchronization from comparable periods. Jupiter aligned with Sun. Double peak appears at full and new, mono-peak at half Moon.

Legend: (also for following figures):

- thick lines: culmination hour for Sun (full), Moon (broken) and planet (Jupiter, dash-dotted);

- thin lines : rise and set hour. Right hand curve: smoothed Ap-value. vertical scale: one hatched range equals .5 MHz.

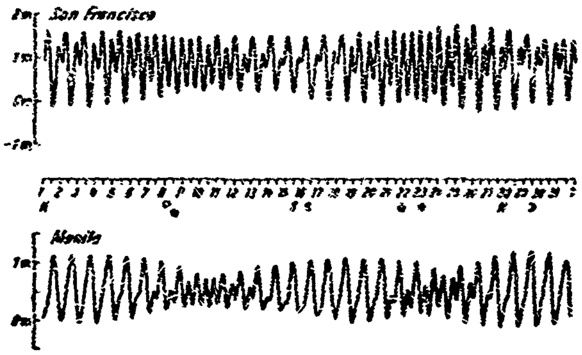


Fig. 1b Water level at San Francisco (top) and Manila (bottom) in March 1936. Source: DEFANT, A., 1957, "Flutwellen und Gezeiten des Wassers", Handbuch der Physik 48, 846-927, Fig. 12.

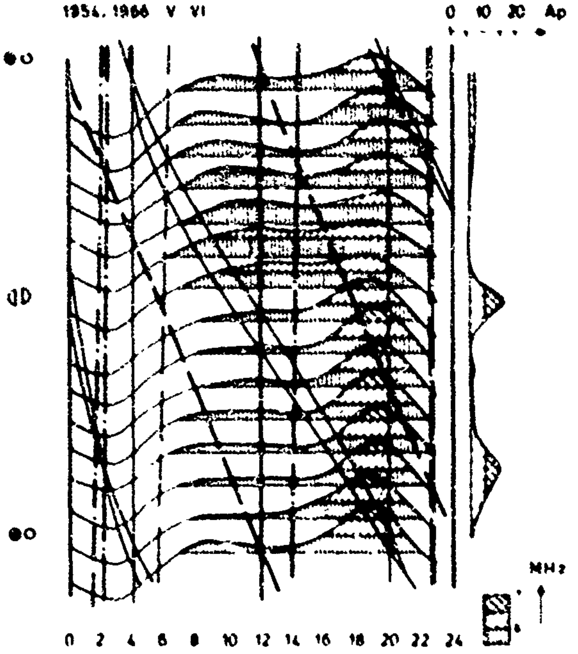


Fig. 2a Jovian phase + 2 h

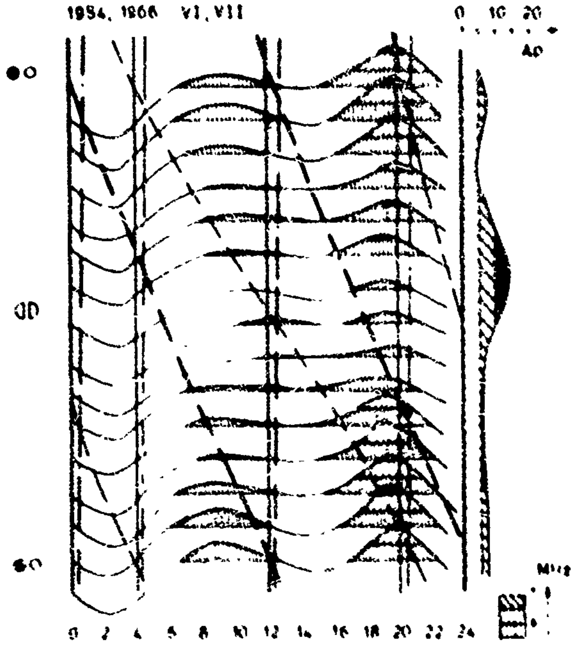


Fig. 2c Jovian phase 0...1 h

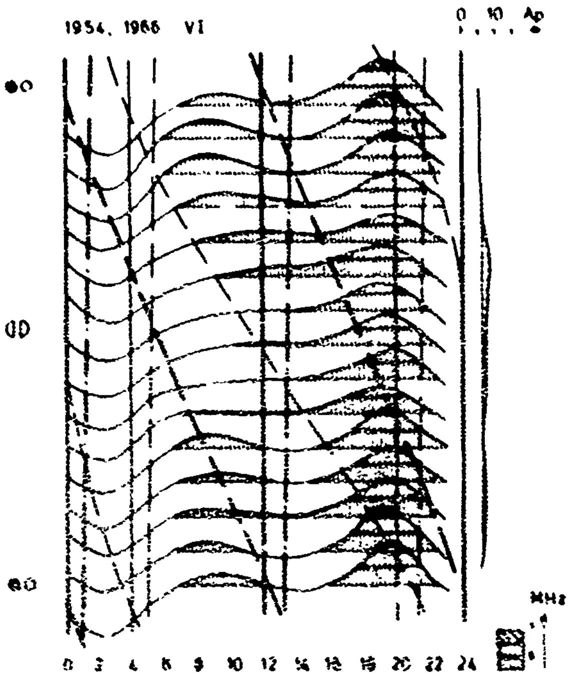


Fig. 2b Jovian phase 1...2 h

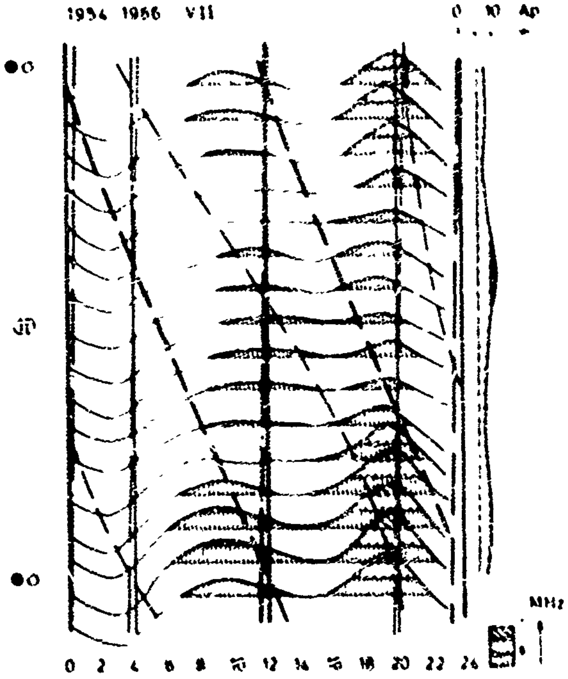


Fig. 2d Jovian phase about 0

Fig. 3 Daily smoothed curves of foF2 as in Fig. 1a. Jupiter and Venus far from Earth, phase difference between Jupiter and Sun variable.

Continued

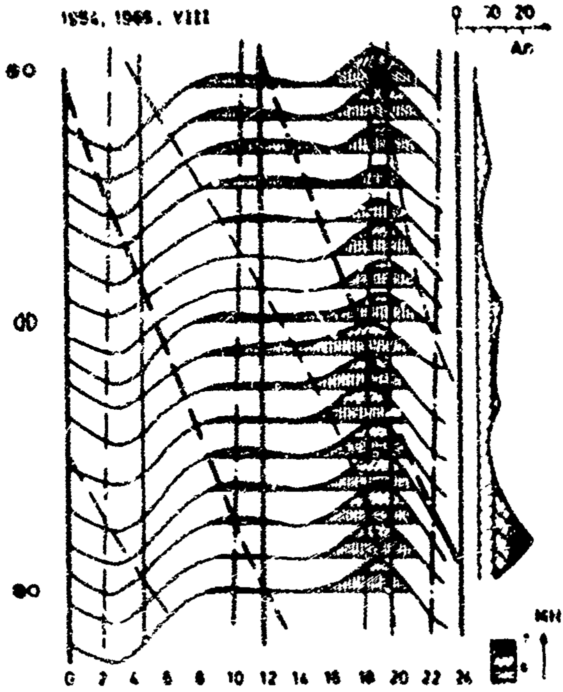


Fig. 2e Jovian phase -1...-2 h

Fig. 2 Daily smoothed curves of foF2 as in Fig. 1a. Jupiter and Venus far from Earth, phase difference between Jupiter and Sun variable.

Concluded

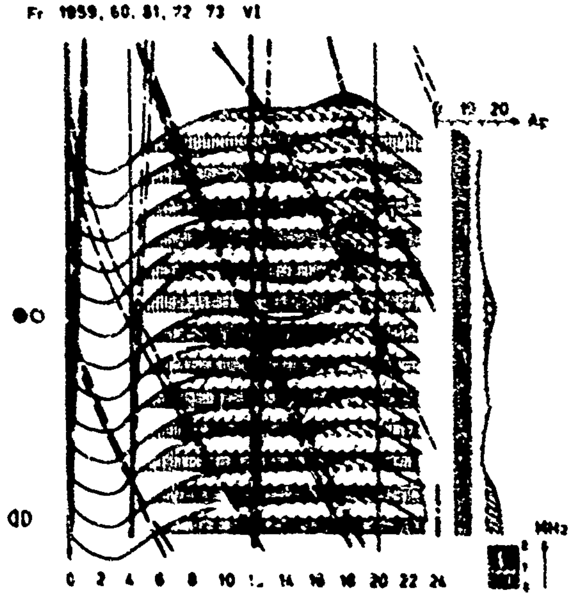


Fig. 3A Jovian phase 0...1 h

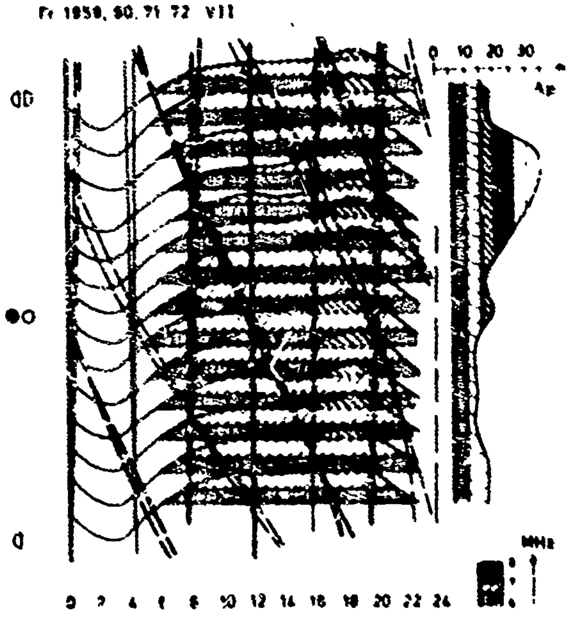


Fig. 3b Jovian phase about -4 h

Fig. 3 Daily smoothed curves of foF2 as in Fig. 2, but during decrease of solar cycle. (Hatching code different). Jupiter near Earth.

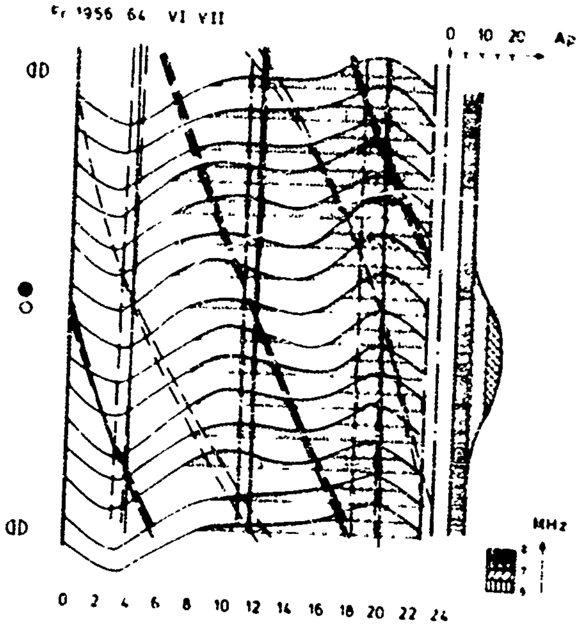


Fig. 4 Daily smoothed curves of foF2 as in Fig. 3 (same hatching code). Jupiter far from Venus near Earth. Planetary position means Venus here. Venusian phase about -1 h.

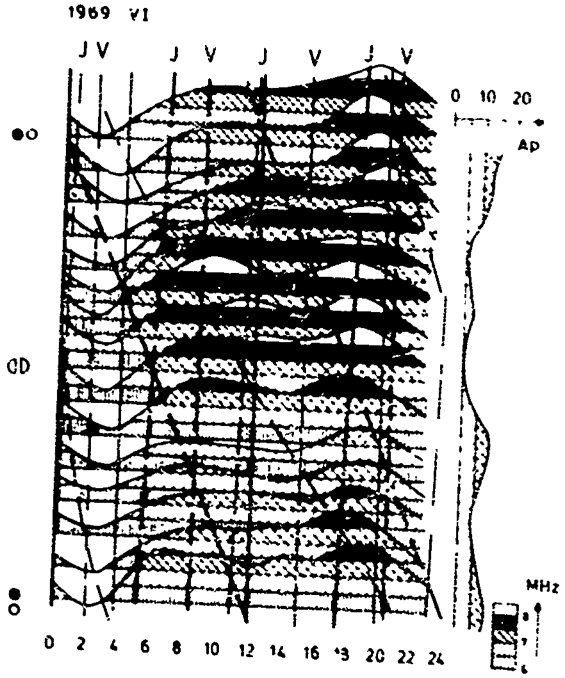


Fig. 5 Daily smoothed curves of foF2 as in Fig. 4 (same hatching code). May/June 1969. Planetary disturbed conditions: Jupiter and Venus near Earth and at odd phases.

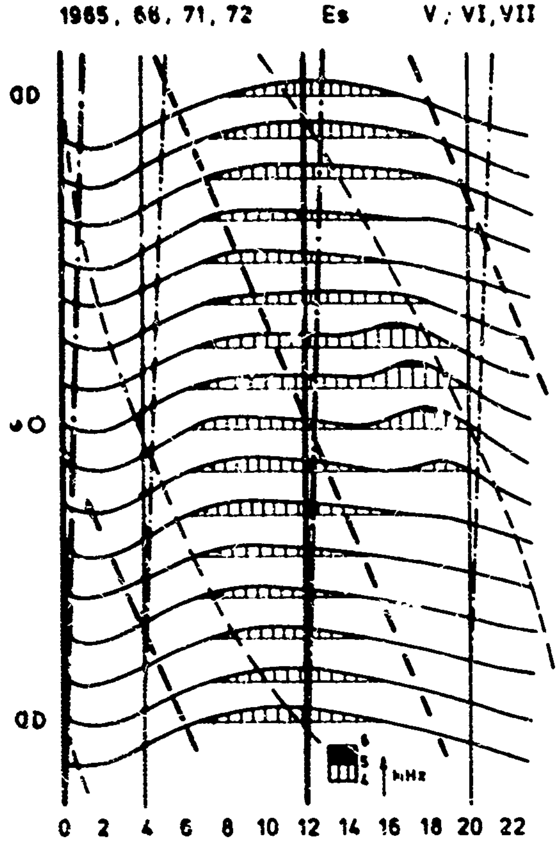


Fig. 6a Jovian phase 0...1 h

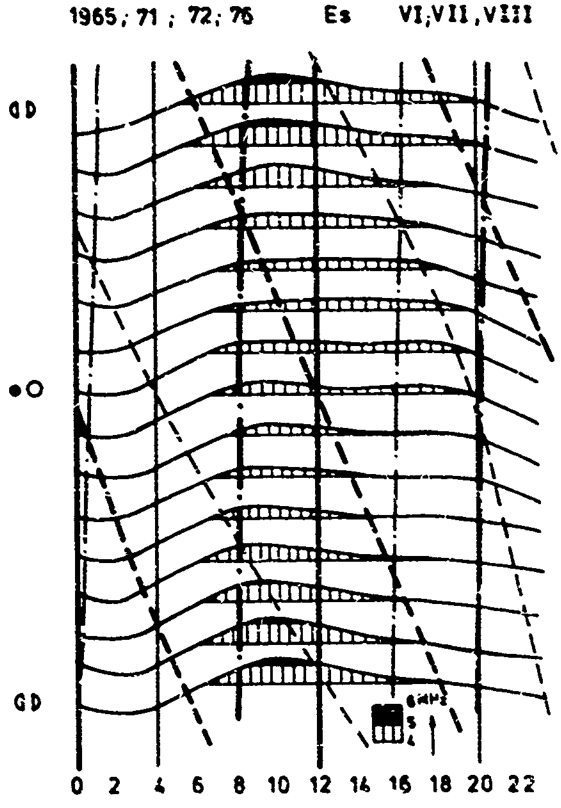


Fig. 6b Jovian phase about -4 h.

Fig. 6 Daily smoothed curves of foEs (compressed vertical scale: one hatched range equals 1 MHz).

SUR LA DETERMINATION DE LA MUF CLASSIQUE

Par R. HANBABA
Centre National d'Etudes des Télécommunications
Département M. I. R.
22301 LANNION (FRANCE)

RESUME

Le calcul de la fréquence la plus élevée à laquelle des ondes radioélectriques peuvent se propager par réflexion ionosphérique entre deux points donnés est généralement basé sur une représentation des modes de propagation par des trajets rectilignes entre le sol et des points de réflexion dans l'ionosphère. En adoptant le modèle quasi parabolique de variation avec l'altitude de la concentration électronique de BRADLEY et DUDENEY, des expressions analytiques de la longueur d'un bond ionosphérique, de la MUF et de l'angle d'élévation du parcours radioélectrique peuvent être obtenues.

Toutefois, pour des applications pratiques, il est nécessaire d'utiliser une méthode qui soit aussi simple que possible. Une méthode "manuelle" de détermination de la MUF et des angles d'élévation du parcours radioélectrique est présentée.

1 - INTRODUCTION -

Le problème pratique à résoudre est celui qui consiste à calculer la fréquence la plus élevée à laquelle des ondes radioélectriques peuvent se propager par réflexion ionosphérique entre deux points donnés au sol.

Deux méthodes d'évaluation de cette limite supérieure de fréquence (MUF) peuvent être envisagées. En effet, les trajectoires des ondes peuvent être déterminées, par exemple, pour une suite de fréquences croissantes jusqu'à ce que le groupe d'ondes ne puisse plus être réfracté par l'ionosphère et parvenir au lieu de réception. Pour se faire, un modèle de distribution verticale de la concentration électronique doit être préalablement défini. Les coordonnées géographiques des points pour lesquels ce modèle doit être établi ainsi que l'angle d'élévation et l'azimut de départ du parcours radioélectrique se déduisent d'algorithmes itératifs. Cette procédure implique des calculs très longs. Elle ne peut être employée que pour des études particulières de propagation.

La seconde méthode fait appel à des expressions analytiques reliant la longueur d'un bond ionosphérique, l'angle d'élévation de ce parcours et la fréquence. Pour des modèles ionosphériques simples, des solutions ont été présentées par CROFT et HOOGASIAN, WESTOVER, FOLKESTAD et NIELSON dans un numéro spécial de la revue "Radio Science" (1968).

Des modèles ionosphériques plus réalistes ont été développés depuis cette date. Le profil de distribution verticale de la concentration électronique que nous avons adopté a été élaboré par BRADLEY et DUDENEY (1973). Il se compose des éléments suivants (figure 1) :

a) Une couche E parabolique située en dessous de la hauteur h_mE de sa fréquence maximale de plasma foE et ayant une demi-épaisseur y_mE : les valeurs de h_mE et y_mE sont supposées constantes et respectivement égales à 110 km et 20 km.

b) Une couche F2 également parabolique située en dessous de la hauteur h_mF2 de la fréquence maximale de plasma foF2 et ayant une demi-épaisseur y_mF2 ; les valeurs des paramètres h_mF2 et y_mF2 se déduisent d'équations empiriques.

c) Une augmentation linéaire de la densité électronique avec la hauteur ; cette région s'étend de la hauteur h_mE au point de la couche F2 parabolique où la fréquence de plasma fj est égale à 1,7.foE.

Ce modèle présente un grand intérêt pour les calculs de propagation ionosphérique :

d) Les hypothèses simplificatrices introduites en permettent une utilisation peu onéreuse.

e) La précision de cette représentation est suffisante.

f) Les paramètres de ce modèle se déduisent des valeurs des caractéristiques ionosphériques qui sont régulièrement mesurées en un grand nombre de stations de sondage (PIGGOTT et RAWER, 1972) et pour les quelles on dispose de méthodes de prévisions pour un lieu géographique et une période quelconques.

Des expressions analytiques reliant la longueur d'un bond, l'angle d'élévation du parcours radioélectrique et la fréquence ont été obtenues pour ce modèle (MILSOM, 1977 ; CCIR, 1978). Bien que les effets du champ géomagnétique soient négligés, les calculs sont encore très longs et il est coûteux d'établir ainsi régulièrement des prévisions ionosphériques.

Une méthode aussi simple que possible est donc donnée. La MUF et les angles d'élévation du parcours radioélectrique se déduisent d'expressions analytiques très simples et de tableaux de valeurs numériques calculées une fois pour toutes.

2 - DETERMINATION DES TRAJECTOIRES DES ONDES RADIOELECTRIQUES

Les trajectoires des ondes radioélectriques sont obtenues par intégration numérique du système d'équations aux dérivées partielles de Haselgrove.

Dans ce type de méthode, l'indice de réfraction de l'ionosphère est calculé en chaque point d'intégration. Il faut donc connaître, en tout point de l'espace, la densité électronique, le champ géomagnétique et la fréquence de collisions électroniques. Si la fréquence de collisions peut être raisonnablement négligée, il n'en est pas de même du champ géomagnétique. Il n'est toutefois pas nécessaire d'introduire un programme très élaboré de calculs des composantes de ce champ et l'hypothèse d'un champ dipolaire est suffisante.

Nous n'avons pas introduit de gradients horizontaux d'ionisation. Des profils de distribution verticale de la densité électronique ont été définis à partir des valeurs médianes mensuelles de paramètres ionosphériques obtenues aux stations de sondage de Dakar, Fort-Archambault, Kerguelen, Ouagadougou, Poitiers, Tahiti et Terre Adélie pour un grand nombre de périodes comprises entre 1957 et 1976.

La réfraction troposphérique peut être négligée. En effet, en se limitant à des angles d'élévation supérieure à 2 degrés, nous montrons que l'erreur sur la distance franchie avec une seule réflexion ionosphérique reste inférieure à 23 km ; l'erreur sur l'angle d'élévation est inférieure à 0,5 degré.

Il est bien connu, que, pour un modèle de densité électronique et une fréquence f donnée, la distance franchie en un seul bond par un rayon émis dans un azimut déterminé passe par une valeur minimale d_{min} pour une valeur particulière β_M de l'angle d'élévation β du rayon à son départ du sol.

On a représenté sur la figure 2 l'erreur que l'on commettrait sur le paramètre d_{min} en négligeant le champ géomagnétique, pour 3 positions géographiques de l'émetteur, en fonction de l'azimut dans lequel se fait cette émission. Lorsque la latitude géomagnétique reste sensiblement constante sur toute la longueur de la liaison, l'accroissement correspondant de d_{min} est très faible. Par contre, lorsque l'émission est faite vers le Nord ou vers le Sud, l'accroissement de d_{min} ne peut plus être négligé. Cet accroissement décroit généralement avec la latitude géomagnétique (figure 3). En négligeant le champ géomagnétique, l'erreur sur la valeur calculée de l'angle d'élévation du parcours radioélectrique est très faible : elle atteint une valeur maximale de 0,5 degré pour des trajectoires parallèles aux lignes de force du champ géomagnétique.

Les variations du rapport f/f_{oF2}, appelé "facteur de distance M_d" en fonction de d_{min} ont été analysées pour le mode ordinaire. La limite supérieure des fréquences propagées par ce mode peut alors s'écrire :

$$MUF = fOF2 \cdot M_d$$

Les prévisions pour la propagation ionosphérique étant établies pour des probabilités d'assurer une liaison allant jusqu'à 95 % du temps, les calculs ont été effectués en choisissant des coordonnées et un azimut pour lesquels la valeur calculée de M_d est minimale.

3 - CALCUL de la MUF et de l'ANGLE D'ELEVATION DU PARCOURS RADIOELECTRIQUE A LA MUF

3.1 - Limite supérieure des fréquences réfléchies par la couche E normale.

Aux distances inférieures à 2250 km, la couche E peut occulter complètement les couches ionisées de la région F et imposer à elle seule la limite supérieure de la bande des fréquences utilisables. Il suffit alors d'une série de valeurs du paramètre M_d pour calculer la MUF (figure 4). L'angle d'élévation β_M du rayonnement permettant alors d'assurer la liaison est petit et nettement inférieur à l'angle que présenterait une propagation par F (figure 5).

3.2 - Limite supérieure des fréquences réfléchies par la région F.

La forme de la région F étant variable, une série de valeurs des paramètres M_d et β_M pour différentes distances n'est plus suffisante. Dans une première approximation, il est possible de réduire la dispersion des valeurs de ces paramètres en introduisant le facteur M3000F2 mais ces valeurs dépendent également des autres caractéristiques ionosphériques définissant le profil d'ionisation.

Des équations empiriques reliant M_d et β_M à la distance d et à ces caractéristiques, à savoir fOE, foF2, M3000F2 et h'F (ou h'F2), ont été établies (on choisit la valeur h'F ou h'F2 qui représente la hauteur de réflexion virtuelle minimale sur la couche F2).

Pratiquement, les relations suivantes sont utilisées :

$$M_d = M_0 - (\alpha_0 (d) \cdot x_M + \Delta M_d)$$

Avec :

$$x_M = y_{mF2} \cdot \left(1 - \frac{1,22 \cdot 10^5}{h_{mF2} - y_{mF2}} \cdot \left(\frac{f_{OE}}{f_{OF2}}\right)^2\right)$$

Les valeurs des paramètres h_{mF2} et y_{mF2} sont données par les équations empiriques (BRADLEY et DUDENEY, 1973) :

$$h_{mF2} = \frac{1490}{M3000F2 + \Delta M} - 176$$

où :

$$\Delta M = \frac{0,18}{X - 1,4} + \frac{0,096\ (R12 - 25)}{150}$$

Avec :

$$X = \frac{foF2}{fOE}$$

Lorsque $foF2 < 1,7 \cdot fOE$, $X = 1,7$; R12 est la moyenne glissante sur 12 mois du nombre de taches solaires,

et :

$$y_{mF2} = h_{mF2} - h' + \Delta h'$$

où :

$$\Delta h' = \left(\frac{0,163}{X-1,35}\right)^{0,86} \cdot (h_{mF2} - 104)$$

$$h' = h'_F \text{ ou } h'_{F2}$$

Avec :

$$\begin{cases} 35 \leqslant y_{mF2} \leqslant h_{mF2} - 110 \\ y_{mF2} = 0 \quad \text{pour } X = 1,7 \end{cases}$$

M_o et $a_o(d)$ sont respectivement donnés par les figures 6 et 7. ΔM_d traduit la diminution du "facteur de distance" pour les faibles valeurs de foF2 :

$$\Delta M_d = 18,182 \cdot a'_o(d) \cdot a\ (foF2\ ,\ M3000F2)$$

Avec :

$$a'_o(d) = \begin{cases} a_o(d) & \text{pour } d \leqslant 3000 \text{ km} \\ a_o(3000) & \text{pour } d > 3000 \text{ km} \end{cases}$$

$a(foF2,\ M3000F2)$ est la plus grande des deux valeurs $a_F\ (foF2,\ M3000F2)$ et 0 :

$$a_F\ (foF2,\ M3000F2) = 56,575 - 14,712.foF2 + 1,587.foF2^2 + 1,094.M3000F2^2 - 10,187.M3000F2$$

pour $foF2 < 4$ MHz et :

$$a_F(foF2, M3000F2) = \frac{(7 - foF2)\ .\ (M3000F2 + 0,533\ foF2 - 5,931)}{0,267\ .\ foF2 - 1,866}$$

pour $foF2 \geqslant 4$ MHz.

Dans le cas particulier où y_{mF2} est égal à 35 km, on pose :

$$x_M = \frac{1490}{M3000F2} - 176 - h'$$

L'angle d'élévation à la MUF peut se déduire de relations similaires :

$$\beta_M = \beta_o + (5 \cdot 10^{-3} \cdot x_\beta + \Delta\beta_{Md})$$

Avec :

$$x = y_{mF2} \cdot \left(1 - \frac{1,70 \cdot 10^5}{h_{mF2} - y_{mF2}} \cdot \left(\frac{f_{OE}}{f_{OF2}}\right)^7\right)$$

La figure 8 donne la relation qui existe entre β_o, d et M3000F2. $\Delta\beta_{Md}$ désigne l'augmentation de l'angle d'élévation pour les faibles valeurs de foF2 :

$$\Delta\beta_{Md} = \begin{cases} 1,2 - 0,24.foF2 & \text{pour } foF2 \leqslant 5 \text{ MHz} \\ 0 & \text{pour } foF2 > 5 \text{ MHz} \end{cases}$$

Lorsque y_{mF2} est égal à 35 km, on pose :

$$x_\beta = x_M$$

4 - LIMITE SUPERIEURE DE LA DISTANCE AU SOL FRANCHIE AVEC UNE SEULE REFLEXION IONOSPHERIQUE

Pour que l'angle d'élévation du parcours radioélectrique soit supérieur à 2 degrés, la distance maximale qu'il est possible de franchir avec une seule réflexion sur la couche E est égale à 2250 km.

Lorsque le rayon est réfracté dans la région F, nous admettrons que cette limite n'est fonction que du facteur M3000F2 :

$$d_{max} = 6585 - (875.M3000F2)$$

5 - CONCLUSION

Les hypothèses simplificatrices décrites sont très utiles aux calculs de prévisions pour la propagation ionosphérique des ondes radioélectriques. Bien que cette méthode est avant tout destinée aux prévisions à long terme, elle est également applicable en considérant les variations d'un jour à l'autre des caractéristiques ionosphériques.

Ces calculs feront l'objet de vérifications expérimentales exploitant les enregistrements en cours du champ de l'onde ionosphérique sur des "liaisons tests". On s'attachera à étudier plus particulièrement le rôle des gradients horizontaux de densité électronique.

BIBLIOGRAPHIE -

BRADLEY, P.A., DUDENEY, J.R., 1973, "A simple model of the vertical distribution of electron concentration in the ionosphere", J. Atmos. Terr. Phys., 35, 2131 - 2146.

CCIR, 1978, "Deuxième méthode informatique provisoire du C.C.I.R. pour l'évaluation du champ et de l'affaiblissement de transmission de l'onde ionosphérique pour les fréquences comprises entre 2 et 30 MHz", Document CCIR, XIVe Assemblée plénière, Kyoto.

CROFT, T.A., HOOGASIAN, H., 1968, " Exact ray calculations in a quasi-parabolic ionosphere with no magnetic field", Radio Science, 3, n.1, 69-74.

FOLKESTAD, K., 1968, "Exact ray computations in a tilted ionosphere with no magnetic field", Radio Science, 3, n. 1, 81-84.

JONES, R.M., STEPHENSON, J.J., 1975, "A versatile three dimensional ray tracing computer program for radio waves in the ionosphere", Tech. Rep. OT, ITS.

MILSOM, J.D., 1975, "Exact ray tracing through the Bradley-Dudeney model ionosphere", The Marconi Rev., 40, 172-195.

NIELSON, D.L., 1968, "Ray-path equations for an ionized layer with a horizontal gradient", Radio Science, 3, n. 1, 101-109.

PIGGOTT, W.R., RAWER, K., 1972, "URSI handbook of ionogram interpretation and reduction", Rep. UAG-23, WDCA, NOAA, Boulder, Colorado.

WESTOVER, D.E., 1968, "Exact ray-path solutions in a quasi-linear ionosphere", Radio Science, 3, n. 1, 75 - 79.

LEGENDES DES FIGURES -

Figure 1 - Modèle ionosphérique de BRADLEY-DUDENEY.

Figure 2 - Erreur, en omettant le champ géomagnétique, sur l'évaluation de la valeur minimale d_{min} de la distance franchie avec une seule réflexion ionosphérique en fonction de l'azimut dans lequel une onde radioélectrique de fréquence 10 MHz est émise en un point de latitude géomagnétique :

1 0°

2 30°

3 60°

(paramètres caractéristiques du modèle ionosphérique :
foF2 = 4 MHz, fOE = 0,5 MHz, h_{mF2} = 330 km, y_{mF2} = 81 km)

Figure 3 - Erreur, en omettant le champ géomagnétique, sur l'évaluation de la valeur minimale d_{min} de la distance franchie avec une seule réflexion ionosphérique en fonction de la latitude géomagnétique du point d'où une onde radioélectrique est émise dans l'azimut 180°.

(fréquence d'émission : 1 13 MHz
2 10 MHz
3 7 MHz
4 5 MHz

paramètres caractéristiques du modèle ionosphérique : foF2 = 4 MHz, fOE = 0,5 MHz, h_{mF2} = 330 km, y_{mF2} = 81 km)

Figure 4 - Variation du paramètre M_{β} avec la distance pour une onde radioélectrique réfléchie par la couche E normale.

Figure 5 - Angle d'élévation β_M d'une onde radioélectrique réfléchie par la couche E normale.

Figure 6 - Variation du paramètre M_0 avec la distance pour une onde radioélectrique réfractée dans la région F pour différentes valeurs du paramètre M3000F2.

Figure 7 - Variation du paramètre $\alpha_0(d)$ avec la distance.

Figure 8 - Variation du paramètre β_0 avec la distance pour une onde radioélectrique réfractée dans la région F pour différentes valeurs du paramètre M3000F2.

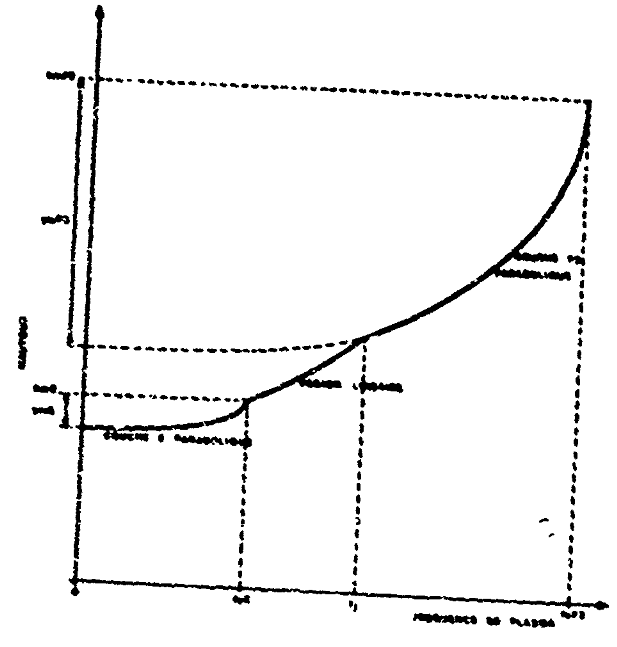


FIGURE 1 -

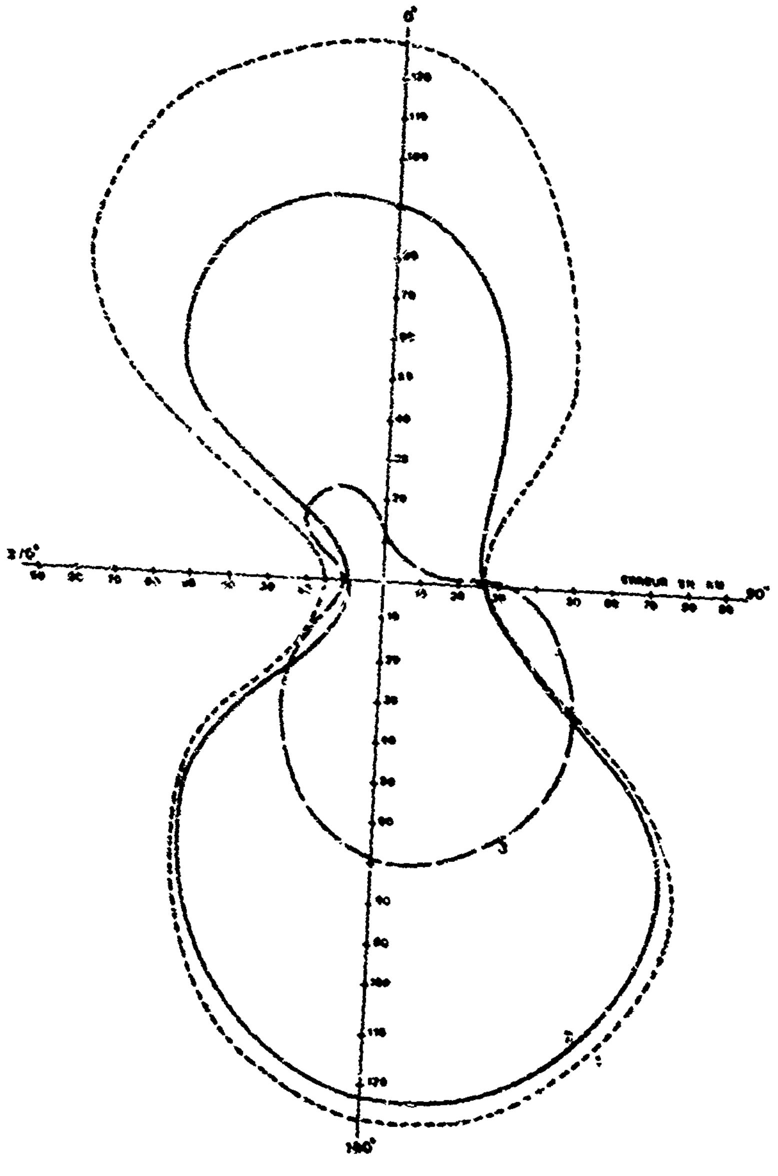


FIGURE 2

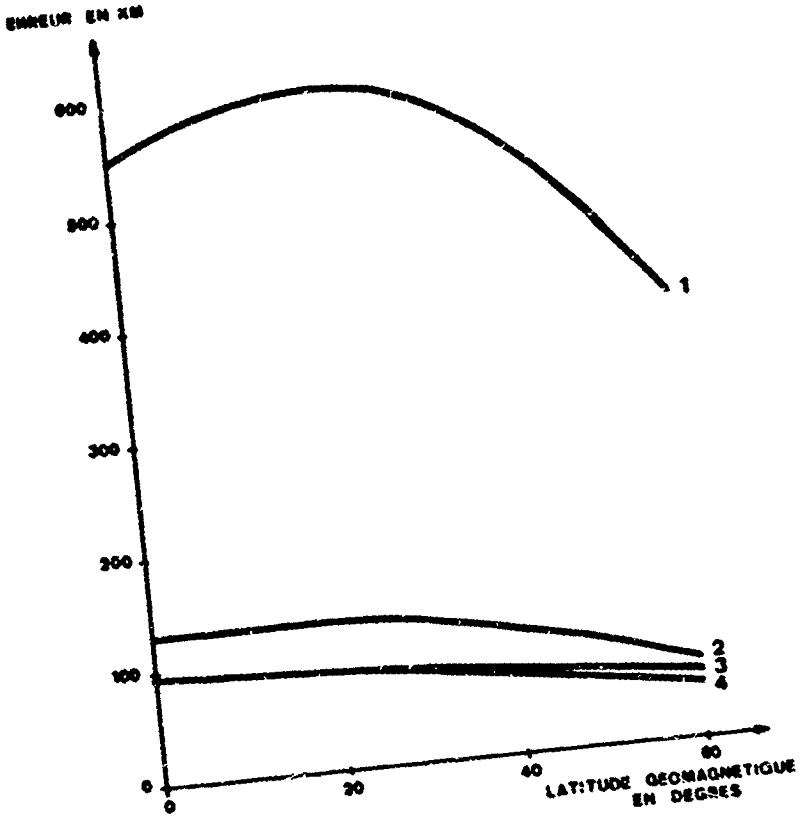


FIGURE 3 -

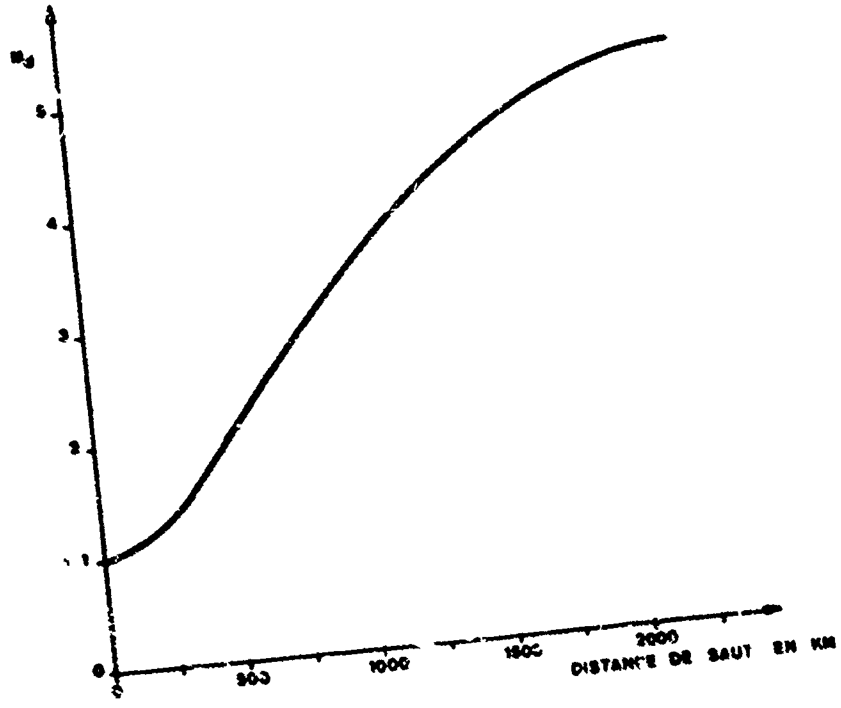


FIGURE 4 -

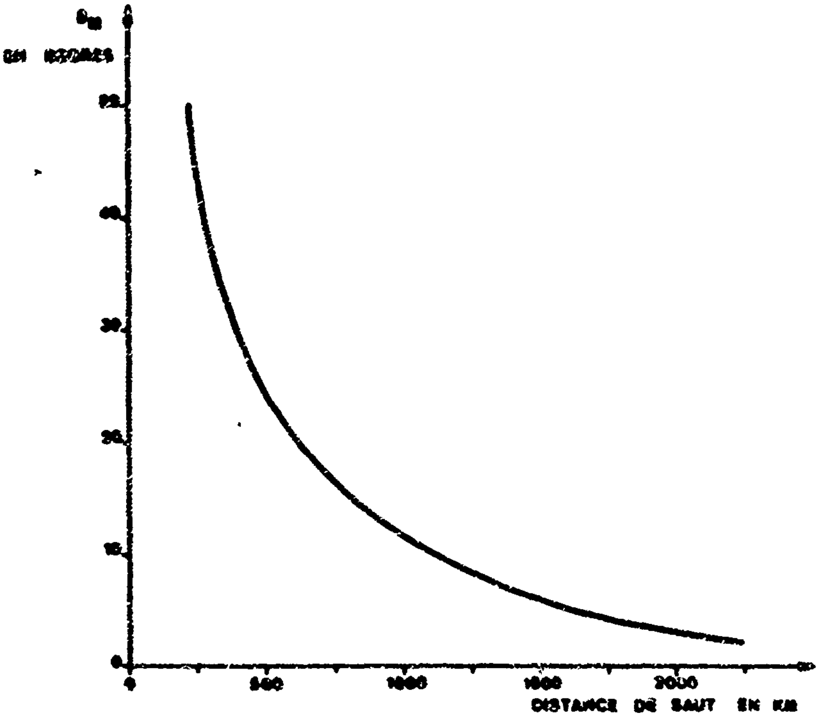


FIGURE 5 -

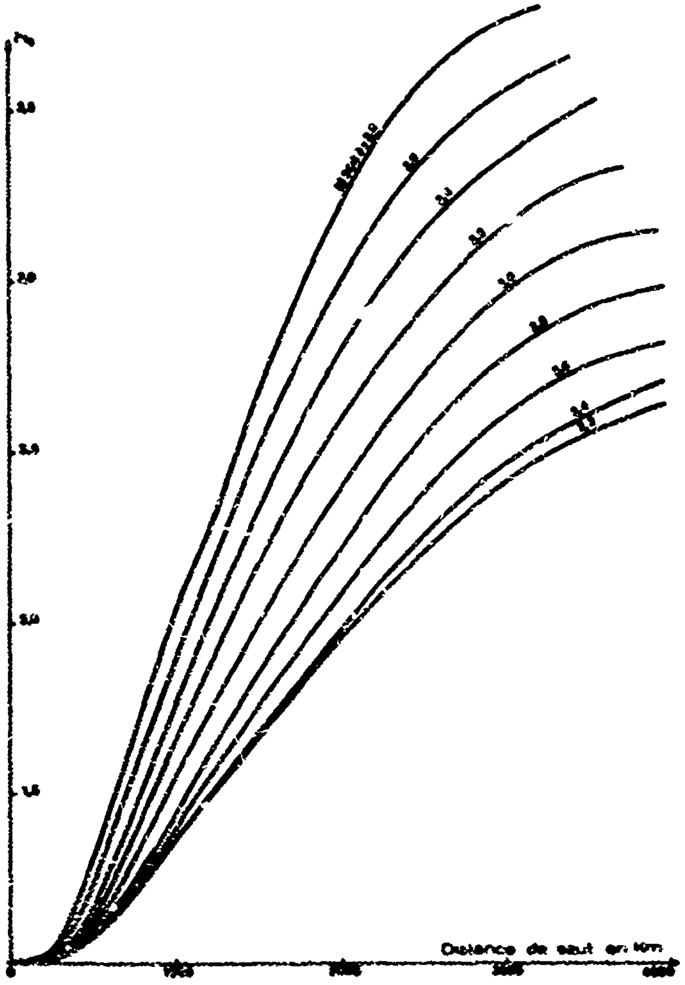


FIGURE 6 -

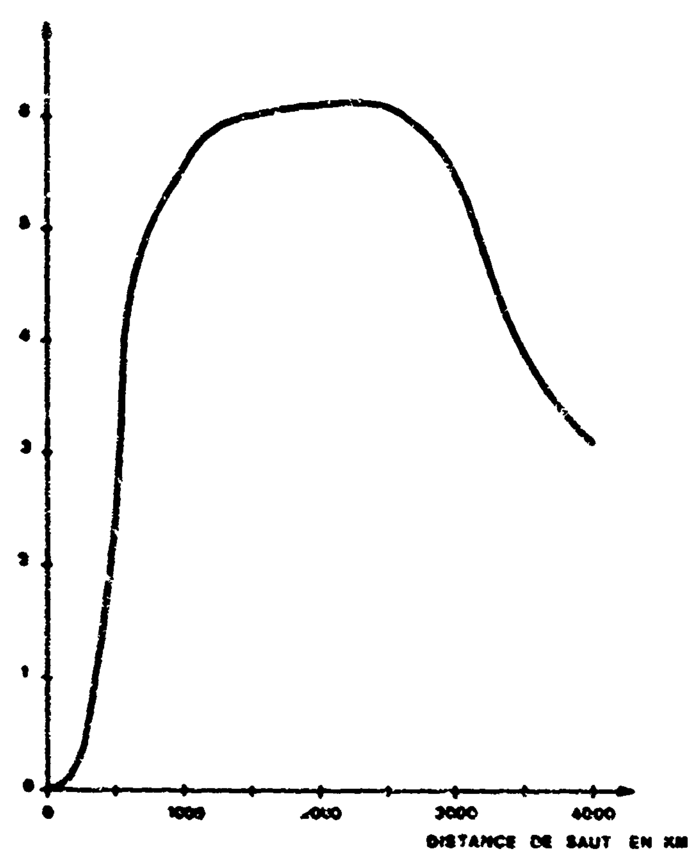


FIGURE 7 -

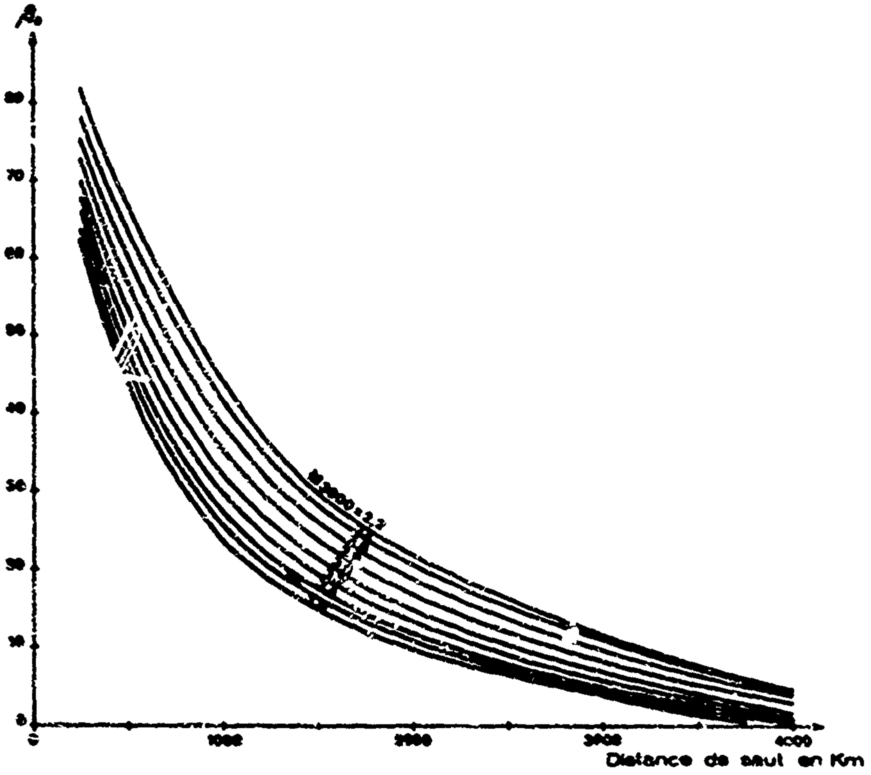


FIGURE 8 -

SUMMARY OF SESSION III

HIGH LATITUDE IONOSPHERE EFFECTS

The ionosphere in middle and low latitudes is predominantly controlled by solar UV radiation. The polar ionosphere, however, is controlled besides UV radiation - mainly by solar particle radiation. The intensity of the UV radiation except during solar flares only shows little variation from day to day. The particle radiation, however, is very irregular in intensity. Therefore in polar regions the HF propagation conditions show complete other features compared to lower latitudes.

The solar particles precipitating in the polar ionosphere are also called auroral electrons and protons. These auroral particles are produced as the result of the interaction between the earth's magnetosphere and the solar wind.

A useful classification of auroral particles from the HF propagation standpoint is in terms of the altitude regimes in which the precipitating particles produce maximum ionization. Thus electrons of (a) ~0.1 keV energy produce ionization principally in the F-region, (b) 1-10 keV in the E-region and (c) ≥40 keV in the D-region. These three classes of auroral particles have their principal sources in three fundamental domains of the magnetosphere. (a) F (~0.1 keV) particles arises from the magnetosheath (the thermalized solar wind), (b) E (1-10 keV) particles from the plasma sheet, and (c) D (≥40 keV) particles from the so-called trapped radiation or Van Allen belts. Therefore the morphology of the magnetospheric domains and where they map down into the ionosphere determines the spatial and temporal characteristics of the auroral ionospheric regions. The auroral particles producing ionization in the E- and F-layer influence the MUF and those producing D-region ionization the LUF in HF communication.

In addition to particle energy a further necessary distinction is spatial structure of the precipitation. This is particularly evident in the E-layer (1-10 keV) ionization: The structured component gives rise to the discrete auroral arcs, rays, etc., which produce a variety of types of sporadic E; the unstructured component gives rise to the continuous aurora the major component of the mantle and diffuse auroras - which produces the auroral E-layer (also known as particle E and night E) with a definable critical frequency. The F and D auroras have similar structured and unstructured components.

The auroral particles (≥40 keV) producing ionization in the D-region have a strong influence on the LUF in HF communication. In former years it was nearly impossible to carry out quantitative measurements of the often occurring extreme absorption in polar regions. In 1956, the riometer was shown to be a powerful tool for use in the study of high latitude absorption. In particular, it allowed for a distinction to be recognized between "polar cap" absorption caused by solar protons of 10-100 MeV energy ejected during very strong flares and what now came to be called "auroral" absorption caused by auroral particles ≥40 keV. There are good resaons for considering the riometer the best hope for establishing an adequate data base for the prediction of auroral absorption effects on HF communication circuits.

Studies given the phenomenon known variously as: "the no-echo condition" "polar blackout", "auroral zone blackout", and now "auroral absorption" have been fruitful, and there have been "breakthroughs," but much more remains to be done. Questions remain about the geographic distributions, the temporal variations, and the relationship with polar cap absorption on one hand, and with magnetospheric substorms on the other.

Recent experiments carried out in the field of "Direction and Doppler Characteristics of Medium and Long Path HF Signals within the Night-Time Sub-Auroral Region" suggests the persistent existence of two relatively weak propagation modes in the evening subauroral region which may be used to extend the range of operating frequencies normally used for communications and surveillance well above the very low F-mode MUF's normally present for that time and region:

(a) A sporadic-E mode which is characterised by a narrow Doppler spread. It is centered close to great-circle, and it possesses an amplitude-dependent spread in direction of the order of 8 degrees for moderately strong propagation. This mode shows appreciable variation in strength from evening to evening, as well as moderate variations within the course of an evening.

(b) Skip-distance-focussed ground sidescatter. This mode occurs in directions well to the south of great-circles, and involves propagation times much in excess of typical great circle modes. It is more constant in strength than the E_s mode.

The relatively weak nature of these 2 described modes would have to be considered in planning for their use.

In the field of "HF Propagation Simulation" advances in ionospheric modelling and numerical ray tracing, over the past few years, have been mainly directed toward the solution of specific applied problems in HF propagation. Ionospheric models have achieved greater versatility in computational terms and earlier models have been modified to better present polar ionospheric features. The calculation of radio absorption has been improved and deviate absorption can now be estimated accurately. Numerical ray tracing techniques developed in the 1960s have been given a similarly expanded versatility permitting their application to a wide range of problems.

Presented examples illustrating recent applications of current HF propagation simulation techniques include the estimation of HF auroral clutter impact on an OTH radar, the investigation of the ionospheric iris effect in ground-satellite propagation, the effect of TiDs on backscatter ionograms and the real time updating of an ionospherical model and computation of ionospheric ducting.

CHARACTERISTICS OF THE HIGH LATITUDE IONOSPHERE
PRODUCED BY AURORAL PARTICLE PRECIPITATION

BY

James A. Whalen

Air Force Geophysics Laboratory

Hanscom AFB, MA 01731 USA

Summary

HF propagation at high latitudes is controlled by the high latitude ionosphere which is turn controlled in large part by precipitating auroral particles, electrons and protons. These auroral particles are produced as the result of the interaction between the earth's magnetosphere and the solar wind. It is the purpose of this paper to describe the general domains of the high latitude ionosphere; to show how these domains relate in space and time to general magnetospheric domains; and to show how these relations aid the understanding of HF propagation in the high latitudes.

In addition the particle-produced ionospheric E region near midnight depends quantitatively on plasma sheet properties: occurrence of auroral sporadic E of about 15 minute duration results from plasma flows in the plasma sheet; auroral E layer height and latitude profiles depend on plasma sheet temperatures and other properties, the definition of which will require further investigation both by satellite and by ionospheric techniques.

Classification of Auroral Particles by Energy, by Magnetospheric Regime

A useful classification of auroral particles on an ionospheric, hence propagation, sense is in terms of the altitude regime in which the particles produce maximum ionization. Thus electrons with kinetic energy of ~0.1 keV produce ionization principally in the F-region; 1-10 keV, the E region; ≳40 keV, the D region (Rees, 1964). (Protons occur in similar categories although the energies necessary to reach these regions are higher by a factor of 4; for example, 4-40 keV protons produce maximum ionization in the E region, Eather and Burrows, 1966).

These same 3 particle energy classes also have precipitation patterns with separate spatial and temporal identities. These patterns result from the fact that the energy classes have their principal sources in three fundamental domains of the magnetosphere. Thus F (~0.1 keV) particles arise from the magnetosheath and are solar wind particles which are thermalized in passing through the bow shock; E (1-10 keV), the plasma sheet; D (≳40 keV), the Van Allen or trapped radiation belts.

Figure 1 shows schematically a model of the magnetosphere in noon-midnight cross-section in which these domains and their appropriate energy populations are identified.

It might be well to note that the F (~0.1 keV) magnetosheath particles represent the part of the aurora which is best understood as to the source: thermalized solar wind particles which gain entry to the dayside ionosphere through the so-called magnetospheric cusps. How these solar wind particles gain entry to the plasma sheet (if indeed they do) and further how they are energized by a factor of 10 to 100 is not understood. Furthermore how the plasma sheet particles are further energized so as to populate the Van Allen belts (if they are) is an additional mystery.

These questions of where the auroral particles come from and how they are accelerated are the fundamental unresolved questions of auroral-magnetospheric physics and are the subject of much active on-going research. It is not the purpose of this paper to treat these questions; instead it is to address the general domains of the high-latitude ionosphere as illuminated by their association with the magnetospheric domains - the morphology and temporal state of the auroral ionospheric domains are a consequence of where and when these magnetospheric domains map down to the earth's upper atmosphere.

Spatial Structure

In addition to particle energy a further general distinction is necessary in order to classify the spatial structure of auroral precipitation, hence the spatial structure of the auroral ionosphere. This distinction is particularly evident in the E (1-10 keV) auroras which are the dominant energy sources for the auroral ionosphere. The structured component gives rise to the discrete auroral arcs, bands, rays, and other visible sometimes dazzling displays. Discrete aurora has historically received the most attention and is popularly described as 'the aurora'; it is photographable hence forms the basis of the Feldstein auroral oval (Feldstein and Starkov, 1967).

The discrete auroral forms can be as narrow as 100 m but can fill latitudinal bands hundreds of km wide.

The ionospheric response to discrete (E) auroras are forms of sporadic E, usually auroral Es. However other forms of Es appear at times to be associated with discrete (E) aurora (e.g. Wagner et al, 1973).

The other general type of E (1-10 keV) aurora is the unstructured component, the continuous aurora, (Whalen et a 1971) which is also known as the mantle (Sandford, 1968) or diffuse aurora (Lui and Anger, 1973). This is also known as the proton aurora since it includes protons - how much of the ionization is attributable to protons and how much to electrons is not well understood and appears to be quite variable.

The continuous aurora gives rise to the auroral E-layer (also known as night E and particle E), is continuous in latitudinal width being typically 200 to 300 km wide (Whalen et al 1971, 1977); and is continuous in longitude where it forms a circumpolar band not only statistically (Wagner et al 1973; Nesprozvannaya and Schuka, 1976) but also instantaneously (Wagner and Pike, 1972, Buchau et al, 1972). Finally the continuous aurora is slowly varying in time.

These two types of aurorae are distinguishable in large scale satellite photographs of the aurora by the DMSP systems. City lights show the location and scale of the photographs (Fig 2). The bright structured forms at high latitude are the discrete auroras; the relatively faint glow at lower latitudes, the continuous aurora.

Morphology of the E Region

The most common and generally accepted model of the aurora is the Feldstein (1967) auroral oval, a statistical compilation of the discrete auroras as determined from auroral all-sky camera photographs taken during the IGY. The oval is a band encircling the magnetic pole but displaced eccentrically at higher latitude in the noon sector as a consequence of the distortion of the earth's magnetic field by the solar wind. Figure 3 shows the Q=3 oval projected on a map of the northern hemisphere in magnetic coordinates (Whalen, 1970). The pattern fixed with respect to the sun, is shown in 4 different orientations with respect to the earth at the indicated values of UT. As a function of the magnetic index Q, the oval expands and moves equatorward, the highest values relating to extreme disturbances as, e.g., solar flare effects.

The typical moderate distribution of discrete aurora, the Q=3 oval, will be used to form a reference system for the other types of auroras. The coordinate system of the oval, Corrected Geomagnetic Latitude and Corrected Geomagnetic Local Time, is shown in 6 different plots in Figure 4. The Q=3 oval is drawn in each plot bounded by the heavy lines. The shaded areas schematically map the types of auroras differentiated as to structured (top row) and unstructured (bottom row); and to particle energy/ionospheric regime by columns.

The E type auroras appear in the center column of Figure 4, the discrete auroral oval itself (above); the continuous aurora (below). The continuous aurora tends to overlap the oval in the night sector (Whalen et al, 1977), but to locate equatorward of the oval in the day sector (Whalen et al, 1971). In quiet times the discrete aurora disappears entirely and the continuous aurora remains (Wagner et al, 1973).

Although the continuous aurora forms a band which encircles the pole for all levels of activity its location can change and the total energy fluxes can vary by several orders of magnitude.

In terms of the magnetosphere the continuous aurora appears to be precipitation from the plasma sheet, both having an energy spectrum which can be described as Maxwellain (Heerwin and Hones, 1974).

The discrete auroral particles also apparently arise from the plasma sheet but undergo additional accelerations which will be discussed later in terms of the temporal characteristics. A further distinction is that continuos aurora results from both electron and proton precipitation. Discrete aurora results from electrons alone since protons effectively diffuse throughout distances of a few hundred km in the atmosphere.

These two E-auroras are not only the most important component of the auroras in terms of energy input into the atmosphere, but also epitomise the minimum structural classes of all the auroras, structured and unstructured.

Morphology of the F-Region

The F-precipitation appears to have structured and unstructured components analogous to the E-region. These are observed in the all-sky photometric observations of Mende and Eather (1976). They seem both to be confined to the region of the oval - whether there are differences in the morphology of the two is not known. The distinction, however, is not important as regards HF propagation because the ionosphere at F layer heights exhibits irregular and spread conditions throughout these regions of soft precipitation. In addition, the polar cap - the circular area poleward of the oval - also is a region of spread F.

An additonal ionospheric region of importance is the F-layer trough, a region of low ionization, which is located equatorward of the oval in the night and evening sectors, is 5°-10° wide and has foF2≤2 MHz (Muldrew, 1965). The poleward edge or wall of the F-layer trough is a region of large electron density gradients which can cause large refractions and thus large departures from great circle propagation paths for HF. Buchau et al (1973) treat a specific case. Pike (1976) has developed an analytical model of the F-layer trough wall.

These F region precipitation patterns are shown in the left column of Figure 4. Thermalized solar wind from the magnetosheath produces the dayside region as noted earlier but apparently not the night sector regions, the field lines for which apparently map down from the plasma sheet. In any case the regions are continuous in local time, as shown.

Morphology of the D Region

The D-region precipitation unstructured component exists as a band of generally a few degrees wide which extends from midnight to morning and sometimes is entirely circumpolar (e.g. Hook, 1968). In quiet times its locus near midnight is the equatorward edge of the continuous aurora; in active times the distribution can be quite wide. The temporal characteristics have been the subject of many studies (e.g. Hartz and Brice, 1967). The occurrence of this precipitation causes D-region HF absorption of relatively short duration (15 minutes) near midnight; longer (1-2 hours), in the morning sector. Structured D-region auroras seem to be associated with activity in the night sector as, e.g. with poleward expanding arcs during substorms. Thus they are transient in nature and occupy the midnight sector of the Feldstein oval.

The patterns for the D-region are shown schematically in the righthand column of Fig. 4.

The unstructured pattern marks the region in which Van Allen belt particles 'un-trap' so as to precipitate into the atmosphere. Berkey et al (1974) have described a number of events in which the dynamics of this precipitation are detailed.

Lifetimes of Discrete E Auroras Near Midnight

As noted earlier the discrete E aurora produces various types of sporadic E which interfere with HF propagation - HF communications and OTH-B surveillance. In addition the optical emissions from the discrete aurora interfere with certain optical surveillance systems. Accordingly the duration of auroral displays gives important information as to the duration of the relevant interference. The temporal lifetimes of discrete auroras have been studied using all sky camera photographs taken on 12 extended local midnight flights by the AFGL Airborne Ionospheric Observatory. Photographs were taken at 1 minute intervals during the total of 93 hours, continuous coverage per flight ranging from 5 to 10 hours (Krukonis and Whalen, 1979). At the present time a high-speed aircraft is unique in its ability to observe the aurora at a given local time, continuously for a duration of many hours. This is possible at high latitudes because the speed of the aircraft permits it to 'keep up' with the rotation of the earth and therefore to reside at a given local time for extended periods of time.

A typical flight is that of 21 January 74 which departed from Goose AB, Labrador, at 0149 Universal Time (UT) and arrived at Eielson AFB, Alaska at 1150 UT.

The aircraft remained within the midnight sector its entire duration, and between 67° and 74° CGL for nearly all of the data portion of the flight.

Figure 5 is a montage of all-sky camera photographs taken on this flight and is a typical example of the data. Each horizontal strip consists of 1/2 hour of data so that the montage displays the complete minute by minute record of the flight, approximately 9 hours duration. UT times are listed to the left of the photographs.

The all sky camera has a 160° diameter field of view so that each photograph records the aurora at E region heights within the area of a circle of 800 km diameter.

The montage reveals temporal patterns not otherwise apparent. In particular, auroral activations can be identified as individual events with definable lifetimes since the sequences of photographs within these events exhibit temporal coherence when viewed in the montage format. For example, the uppermost 5 strips of Figure 8 are a minute by minute record of what appears to be a single event, similar in structure to the classic auroral substorm. Throughout the remainder of the montage however, are many examples of active periods of much shorter duration indicating a much more complex organization than earlier apparent substorm; as illustration, note the period from approximately 0935 to 0945 UT.

The auroral photographs have been categorized by a modified auroral activity index (Carovillano et al., 1975) as Active, Moderate, Quiet and No visible discrete auroras.

A, M and Q events taken as a whole have a Poisson-like lifetime distribution which maximizes near 15 minutes; another lesser group appears at about 120 minutes (Figure 6). Thus the auroral lifetime can be treated as a discrete random event with most probable duration near 15 minutes.

The implication of this temporal variation on HF propagation near the auroral oval is that propagation conditions which are affected by auroral sporadic E are likely to change within about 15 minutes. Whether the change will be for the better or for the worse is not definable at this time.

Reilly (1977) has described very detailed propagation measurements in relation to the oval as defined by the 15 minute magnetic Q index. Apparently the lifetime distribution described here relates closely to the permissability of using a 15 minute temporal resolution.

Plasma flows which have been observed in the plasma sheet (e.g. Pytte et al 1976, Coroniti et al, 1978) are the apparent cause of the auroral activation events reported here: duration of the flows are the same as the 10-20 minute auroral lifetimes; both occur in substorm and non-substorm conditions; a possible mechanism exists for the acceleration of the electrons responsible for the discrete aurora across electrostatic shock waves produced by such plasma flows.

Thus plasma sheet processes result in interference with HF propagation. Accordingly ignorance of plasma sheet dynamics translates into ignorance of the predictability of this interferance.

Spectral and Latitudinal Properties of the Continuous (E) Aurora Near Midnight

Recent studies have shown that the electron precipitation responsible for the continuous aurora is well-represented by a Maxwellian spectrum when the precipitation is integrated over the loss cone (Sharber and Whalen, 1979). Furthermore, the Maxwellian characteristic energy (or 'temperature') and the total energy flux for the precipitated particles are both well represented by a Gaussian latitudinal distribution; the fact that it is nearly the same latitudinal distribution for both these parameters implies further that the particle number flux is approximately constant across the latitudinal distribution (Whalen and Sharber, 1979).

These properties are observed in a detailed study of a single case, directly by the ISIS-2 satellite electron spectrometer; and via their effects on the atmosphere by coinciding airborne ionospheric and photometric measurements.

The satellite measurements of this energy flux, Q, and characteristic energy, α, are shown in Figure 7. Gaussian functions of 3.2° and 3.5° (Full Width at Half Maximum) have been fit to the measured values of Q and α, respectively. (For Q, the Gaussian fits the points well except at the maximum and at higher latitudes in which regions the Gaussian is shown as a dashed curve.) The ionospheric significance of the particle parameters Q and α is that they translate into two essential ionospheric parameters, maximum electron density (nemax) and the height of maximum ion production ($h\eta$max), respectively.

The former is a consequence of the existence of equilibrium in the E region; i.e. ion production rate equals ion recombination rate which is proportional to nemax^2 or foE^4 (e.g. Omholt, 1955). The relative variation of these parameters with latitude is shown in Figure 8a for a 3.7° FWHM Gaussian for Q. The low latitude half of the distributions are plotted, the origin of latitude is taken as the maximum of Q, and all three are normalized to unity at the maximum.

The characteristic electron kinetic energy translates into the height of the maximum rate of ion production (e.g. Rees, 1964) from which the height of maximum electron density can be inferred. A Gaussian latitudinal distribution for α (which is the same as that for Q in Figure 8a) produces the latitudinal variation in $h\eta$max shown in Figure 8b. Here αmax is taken to be 1 keV.

The spectral properties of this auroral precipitation have been measured to have a large range of variation near midnight. Qmax has been observed to have extreme values of 0.3 and 20 $\text{ergs/cm}^2\text{sec}$ which correspond to values of foE of about 2.3 and 6.7 MHz respectively. α max can range between 0.5 and 7 keV with corresponding values of $h\eta$max of 150 and 100 km.

The latitudinal distributions also vary: The Gaussian width can have values between 2.5° and 5° FWHM; the location of the maximum, 65° to 71° CG latitude.

An effective plasma sheet temperature relates directly to the value of αmax in this precipitation (Rearwin and Jones, 1974) hence to the relative ionization height profile and to $h\eta$max in particular. The value of Qmax and the location and width of the Gaussian latitudinal distribution apparently relate no less directly to the plasma sheet, but to properties not yet understood. It is clear that the predictability of HF propagation by auroral E layer modes awaits further understanding of these plasma sheet properties.

Conclusions

The tracing of regions of auroral precipitation from magnetospheric sources to ionospheric sinks helps conceptually to unscramble some of the complexities of the auroral particle produced ionosphere and the resulting complexities in HF propagation at high latitudes. In addition quantitative relations exist between the E (1-10 keV) auroras and plasma sheet properties: discrete (E) auroras and resulting sporadic E are the result of plasma flows in the plasma sheet; the spectral and spatial properties of the continuous (E) auroras and resulting auroral E layer height and latitudinal profiles are directly related to plasma sheet temperature and to other parameters not yet understood. Further quantifying of the E ionosphere awaits better understanding of the behavior of the plasma sheet. This is not to say, however, that this better understanding will be achieved solely by satellite measurements. Ionospheric parameters themselves relate so closely to particle and plasma sheet parameters that ionospheric measurements can serve as important probes of the plasma sheet.

REFERENCES

Berkey, F.T., V.M. Driatskiy, K. Henriksen, B. Hultqvist, D.H. Jelly, T.I. Shchuka, A. Theander, and J. Yliniemi, A Synoptic Investigation of Particle Precipitation Dynamics for 60 Substorms in IQSY (1964-1965) and IASY (1969), Planet. Space Sci., 22, 255, 1974.

Besprozvannaya, A.S. and T.I. Shchuka, Distribution of Anomalous Ionization in the High Latitude E Region According to Ground-Based Sounding, Geomag. and Aeron., 16, 430, 1976.

Buchau, J., G.J. Gassmann, C.P. Pike, R.A. Wagner, and J.A. Whalen, Precipitation Patterns in the Arctic Ionosphere Determined from Airborne Observations, Ann. Geophys., 28, 443, 1972.

Buchau, J., C.P. Pike and M. Wong, Detailed Specification of the Arctic Ionosphere and an Application to Three-Dimensional Ray Tracing, AFCRL-TR-73-0726, AF Surveys in Geophysics, No. 279, 1973.

Carovillano, R.L., R.E. Sheehan, M.S. Gussenhoven, and G.L. Siscoe, Unified Model of Auroral Substorm Development, AFCRL-TR-75-0558, 1975.

Coroniti, F.V., L.A. Frank, R.P. Lepping, F.L. Scarf and K.L. Ackerson, Plasma Flow Pulsations in Earth's Magnetic Tail, J. Geophys. Res., 83, p. 2162, 1978.

Eather, R.H. and K.M. Burrows, Excitation and Ionization by Auroral Protons, Aust. J. Phys., 19, 309, 1966.

Feldstein, Y.I. and G.V. Starkov, Dynamics of Auroral Belt and Polar Geomagnetic Disturbances, Planet. Space Sci., 15, p. 209, 1967.

Hartz, T.R. and N.M. Brice, The General Pattern of Auroral Particle Precipitation, Planetary Space Sci., 15, 1967.

Hook, J.L., Morphology of Auroral Zone Radiowave Absorption in the Alaska Sector, J. Atmos. Terr. Phys., 30, 1341, 1968.

Krukonis, A.P. and J.A. Whalen, Occurrence and Lifetimes of Discrete Auroras Near Midnight, J. Geophys. Res., in press, 1979.

Lui, A.T.Y. and C.D. Anger, A uniform belt of diffuse aurora and airglow from ISIS 2 spacecraft, Planet. Space Sci., 21, 819, 1973.

Mende, S.B. and R.H. Eather, Monochromatic All Sky Observations and Auroral Precipitation Patterns, J. Geophys. Res., 81, 3771, 1976.

Muldrew, D.B., F-Layer Ionization Troughs Deduced from Alouette Data, J. Geophys. Res., 70, 2635, 1965.

Oguti, T. and K. Marubashi, Enhanced Ionization in the Ionospheric F_2 Region Around Geomagnetic Noon in High Latitudes, Rep. Ionos. Space Res. Japan, 20, 96, 1966.

Omholt, A., The Auroral E-layer Ionization and the Auroral Luminosity, J. Atmosph. Terr. Phys., 7, 73, 1955.

Pike, C.P., An Analytical Model of the Main F Layer Trough, AFGL-TR-76-0098, Air Force Surveys in Geophysics, No. 343, 1976.

Pike, C.P., J.A. Whalen, and J. Buchau, a 12 Hour Case Study of Auroral Phenomena in the Midnight Sector: F-layer and 6300 A Measurements, J. Geophys. Res., 82, 3547, 1977.

Pytte, T., R.L. McPherron, M.G. Kivelson, H.I. West, Jr., and E.W. Hones, Jr., Multiple-Satellite Studies of Magnetospheric Substorms: Radial Dynamics of the Plasma Sheet, J. Geophys. Res., 81, p. 5921, 1976.

Rearwin, S. and E.W. Hones, Near Simultaneous Measurement of Low-Energy Electrons by Sounding Rocket and Satellite, J. Geophys. Res., 79, 4322, 1974.

Rees, M.H., Note on the Penetration of Energetic Electrons into the Earth's Atmosphere, Planet. Space Sci., 12, 722, 1964.

Reilly, A.E., Analysis of Sweep Frequency Oblique Polar Region HF Radio Propagation Measurements, AFGL-TR-77-0102, Environmental Research Paper, No. 596, 1977.

Sandford, B.P., Variations of Auroral Emissions with Time, Magnetic Activity and the Solar Cycle, J. Atmos. Terr. Phys., 30, 1921, 1968.

Sharber, J.R. and J.A. Whalen, Integration over the Loss Cone of Electrons in the Continuous (Diffuse) aurora: The Maxwellian Character of the Spectrum, (Abstract), EOS Trans. AGU, 60, 355, 1979.

Thomas, J.O. and M.K. Andrews, The Transpolar Exospheric Plasma, 3, A Unified Picture, Planet. Space Sci., 17, 433, 1969

Wagner, R.A. and C.P. Pike, A Discussion of Arctic Ionograms, in Radar Propagation in the Arctic, edited by J. Frihagen, Chap. 4, pp. 1-20, Technical Editing and Reproduction, London, 1972.

Wagner, R.A., A.L. Snyder, and S.-I. Akasofu, The Structure of the Polar Ionosphere During Exceptionally Quiet Periods, Planet. Space Sci., 21, 1911, 1973.

Whalen, J.A., R.A. Wagner and J. Buchau, A 12 Hour Case Study of Auroral Phenomena in the Midnight Sector: Oval, Polar Cap and Continuous Auroras, J. Geophys. Res., 82, p. 3529, 1977.

Whalen, J.A., Auroral Oval Plotter and Nomograph for Determining Corrected Geomagnetic Local Time, Latitude and Longitude for High Latitudes in Northern Hemipshere, AFCRL 70-0422, Environmental Research Paper #327, Bedford, Mass., 1970.

Whalen, J.A. and J.R. Sharber, Integration over the Loss Cone of Electrons in the Continuous (diffuse) Aurora: Gaussian Latitudinal Profile, (Abstract). EOS Trans. AGU, 60, 355, 1979.

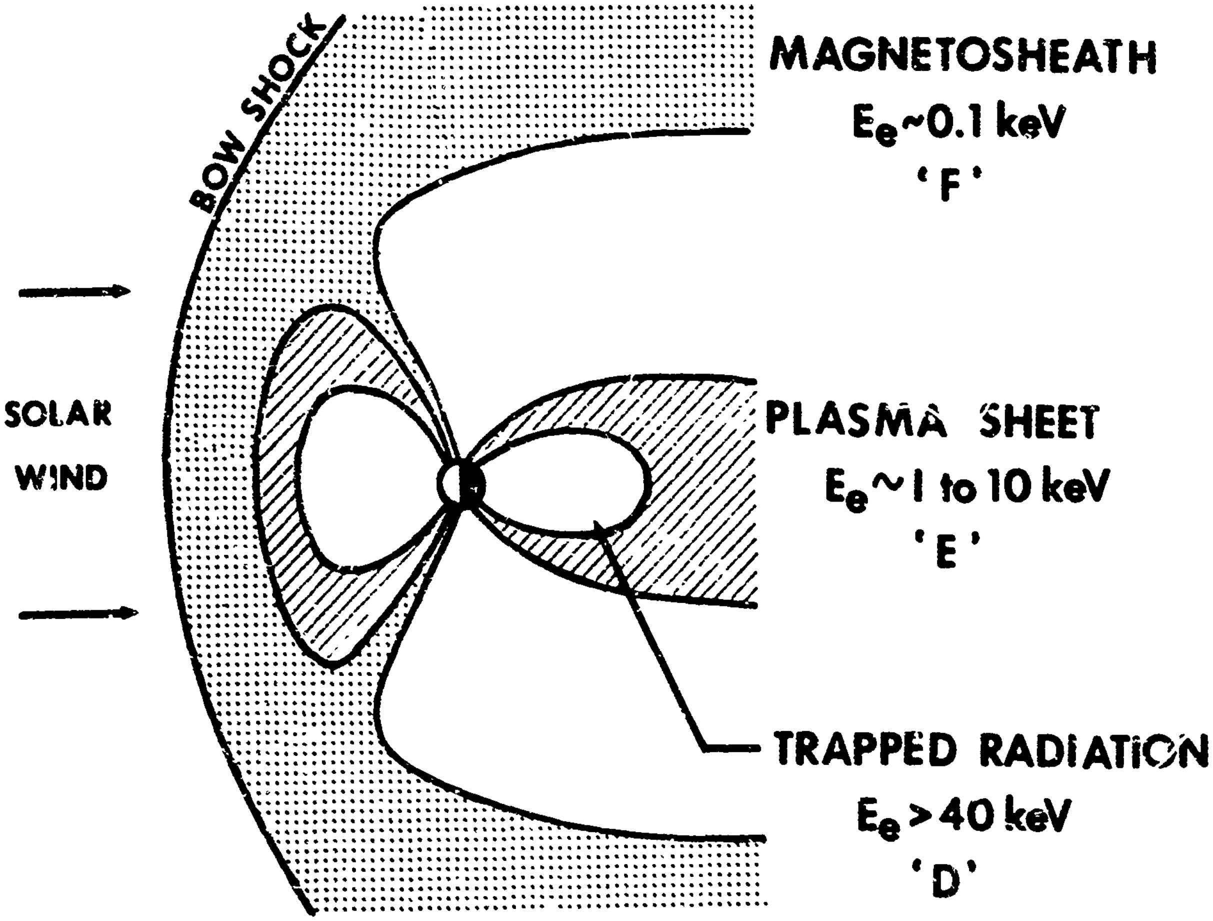

BOW SHOCK
MAGNETOSHEATH
E_e~0.1 keV
'F'
SOLAR
WIND
PLASMA SHEET
E_e~1 to 10 keV
'E'
TRAPPED RADIATION
E_e>40 keV
'D'

Figure 1   Schematic of the magnetosphere of the earth in noon-midnight cross section showing: principal magnetospheric domains; the associated particle populations via representative electron kinetic energy, Ee; and the ionospheric region (F, E or D) in which these particles produce maximum ionization.

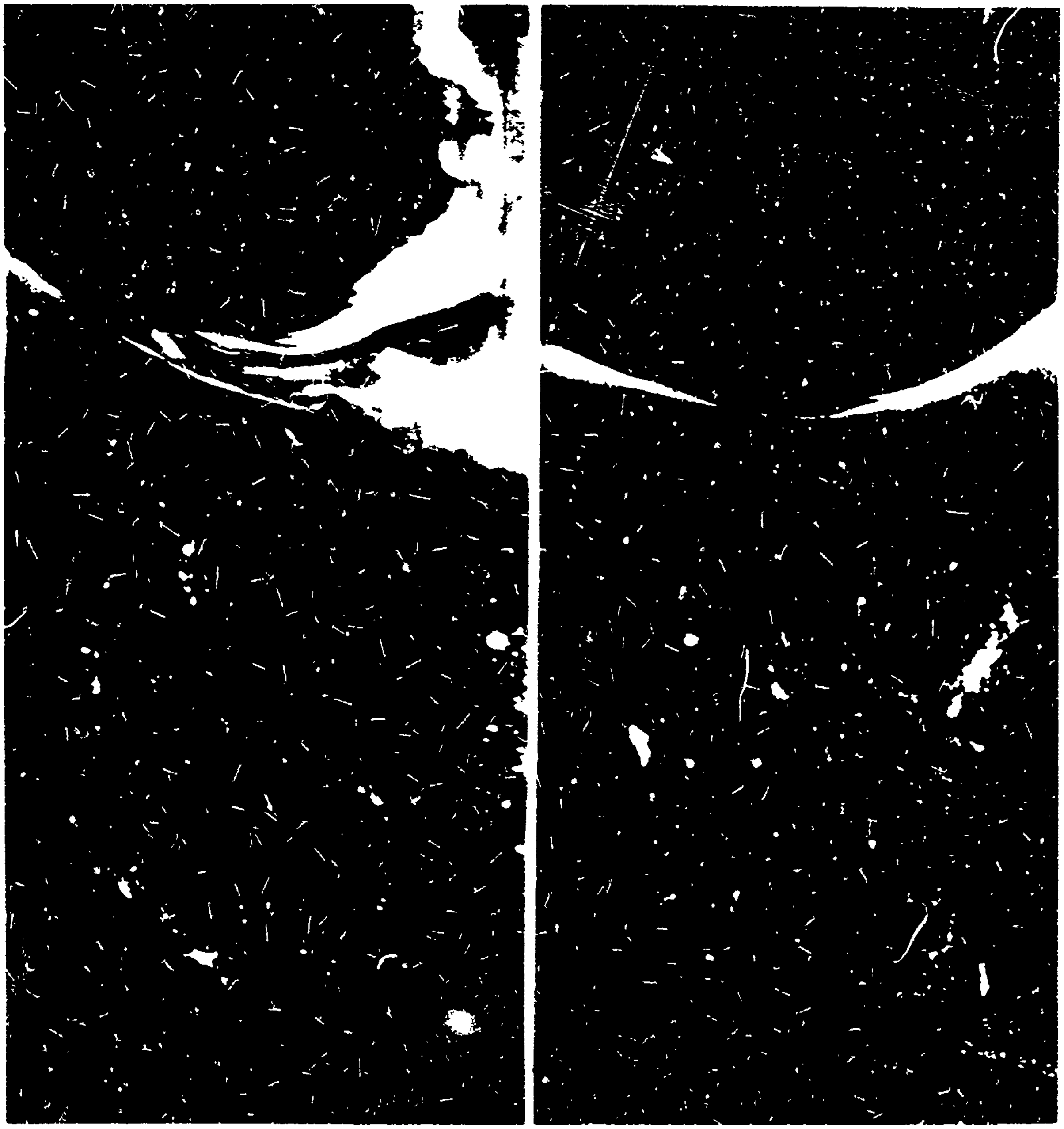


Figure 2 Two DMSP photographs showing the midnight aurora across the North American continent as outlined by city lights.

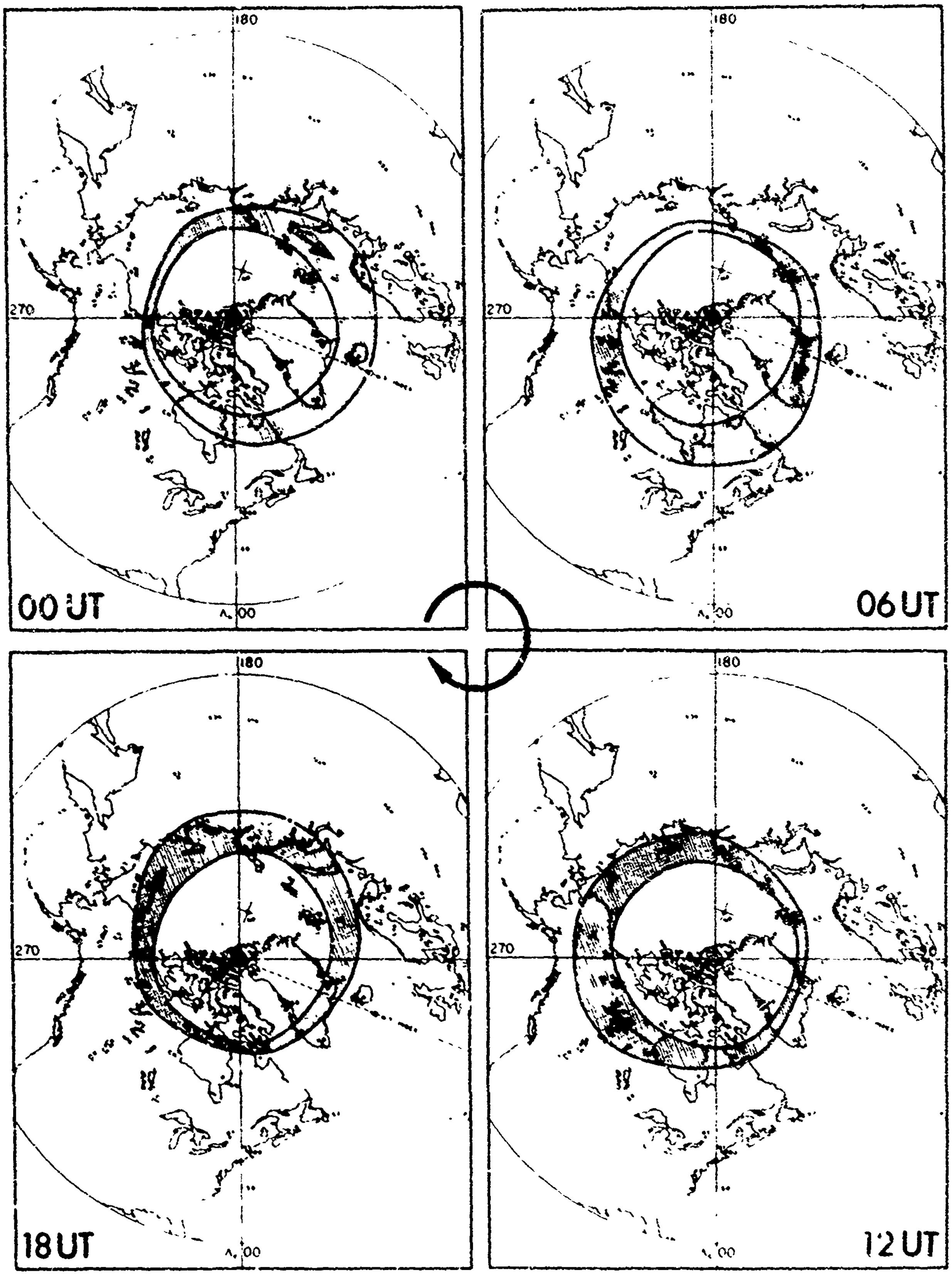

180
270
00
00 UT
180
270
00
06 UT
180
270
00
18 UT
180
270
00
12 UT

Figure 3 The auroral oval, the locus of maximum discrete (3) auroral activity, projected on the north polar region as mapped in Corrected Geomagnetic latitude and longitude. The orientation of the oval at four different values of UT are shown.

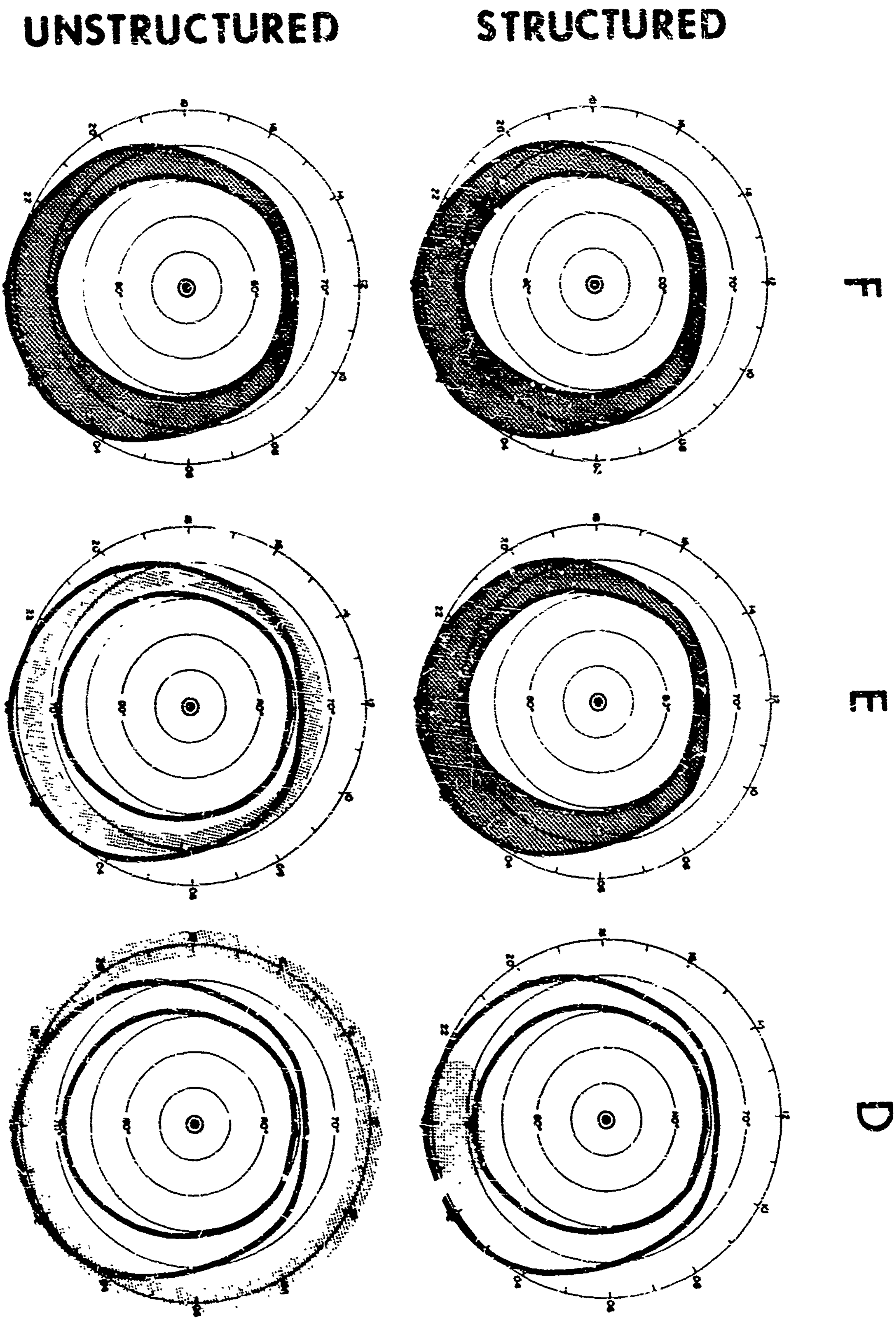
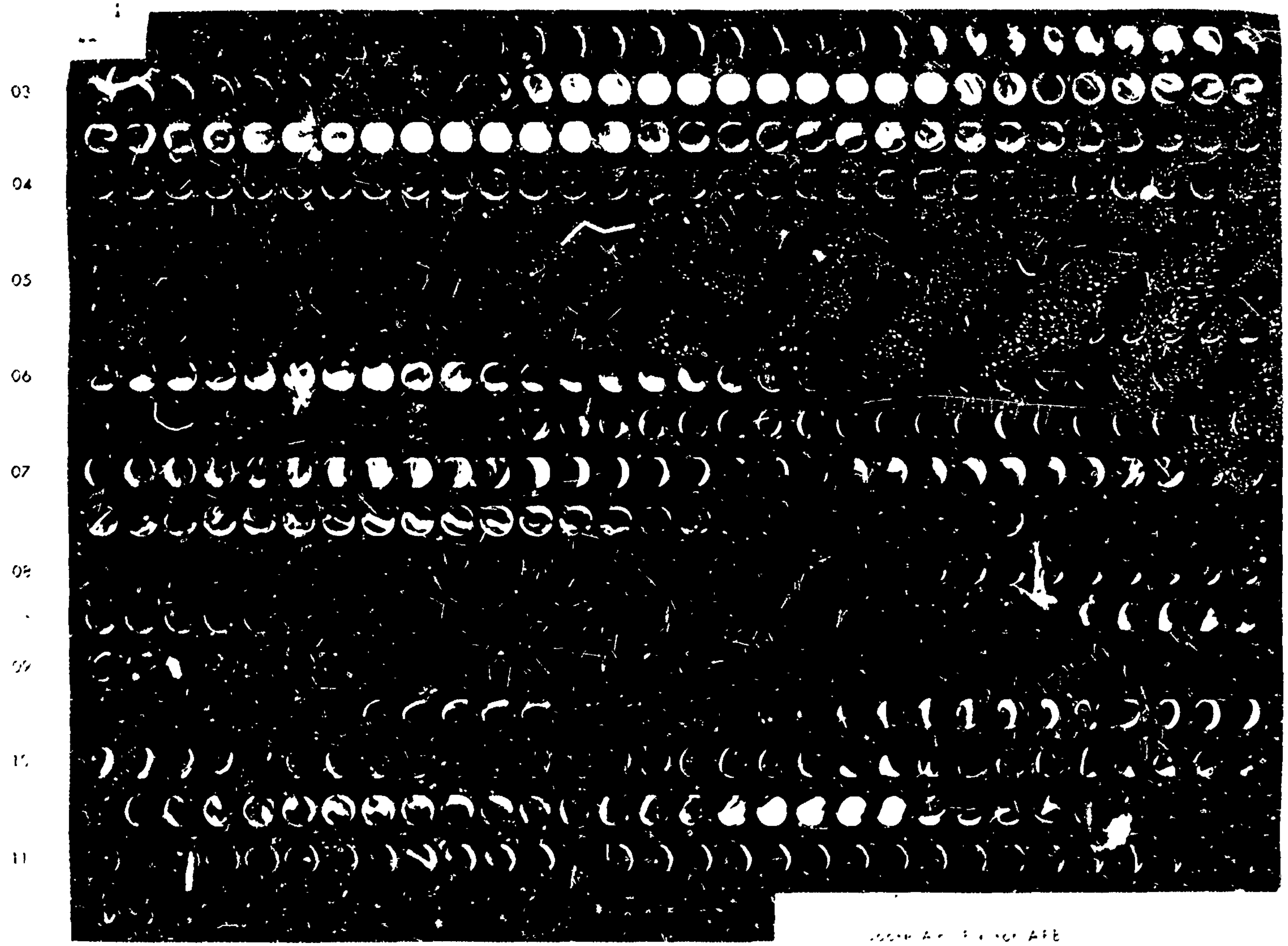


Figure 4 The regions of structured (top row) and unstructured (bottom row) auroras differentiated as to ionospheric regime (particle energy) from left to right F (~0.1 keV), E (1-10 keV) and D (≥40 keV). Coordinate systems are CG latitude and CG local time. The Q=3 auroral oval is defined by the heavy solid lines, the location of the specific auroras, by the shaded areas.



Goose Air Force AFB
[illegible] 1974

Figure 5. Montage of all sky camera photographs taken at one minute increments during AFGL flight of 2[illegible] January 1974. North is indicated in upper left-hand corner. Each strip contains 56 consecutive minutes of data. UT hours are recorded along left.

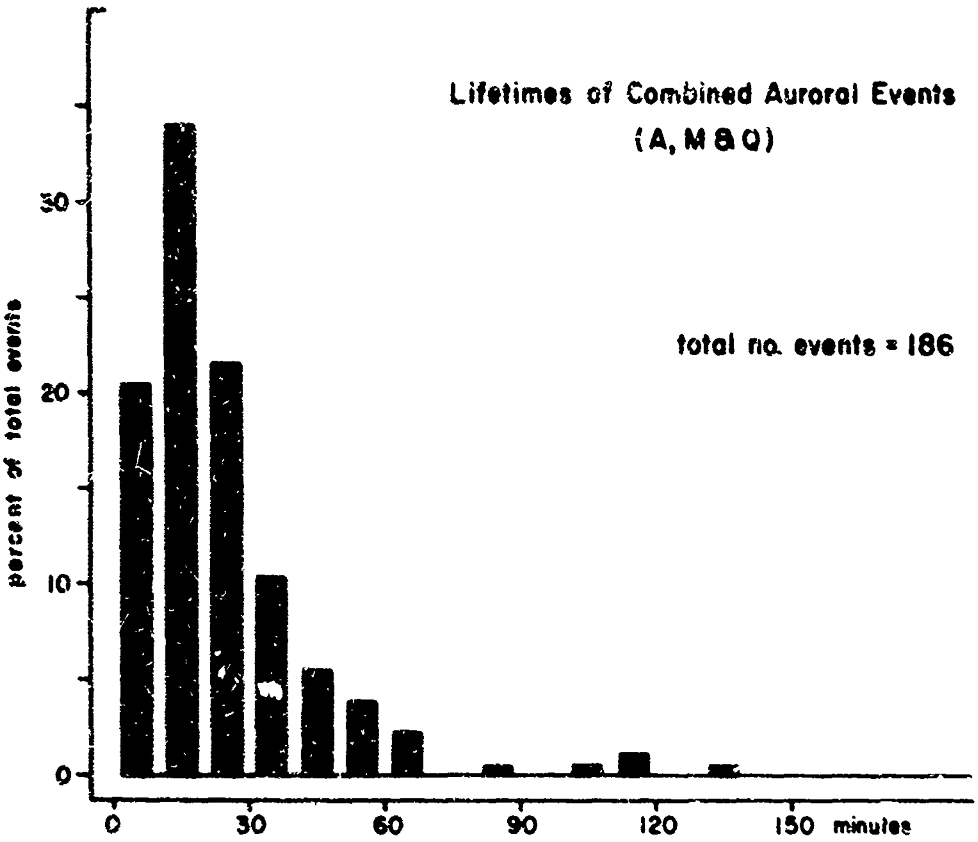


Figure 6 Histogram of the durations of the combined A, M and Q events defined by the airborne all sky camera photographs of discrete auroras near midnight.

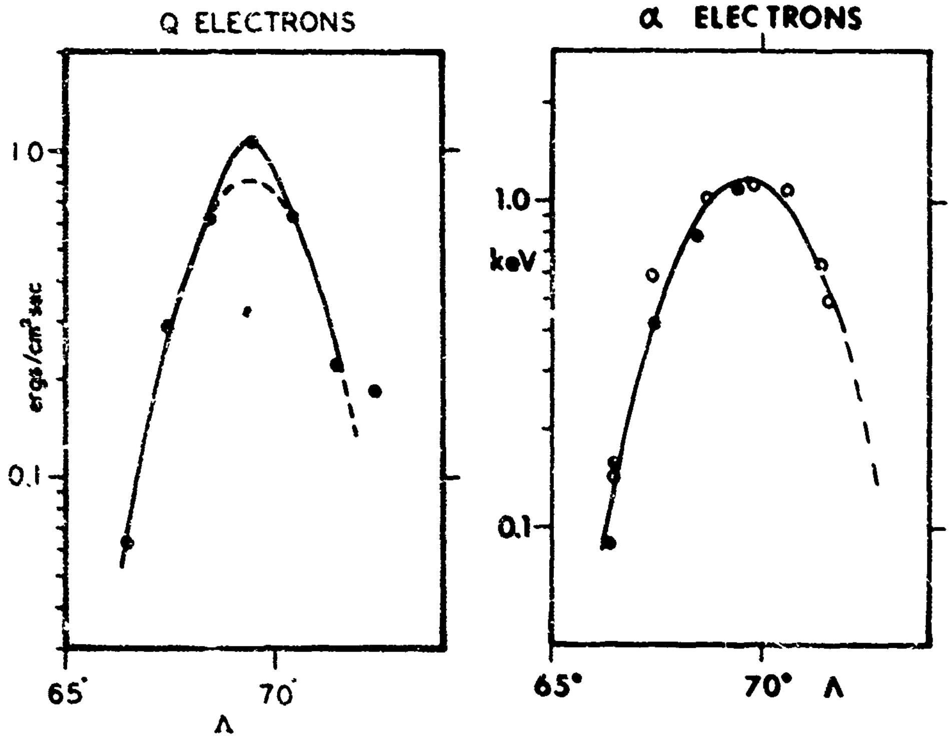


Figure 7 Distributions in invariant latitude, Λ, of the Maxwellian spectral characteristics of the precipitating electrons which produce the continuous (E) aurora. Gaussian functions are fit to the measured points: for Q, the energy flux, a Gaussian of 3.2° (FWHM); for α, the characteristic energy, a Gaussian of 3.5° (FWHM). Measurements are by ISIS-2 satellite electron spectrometer between 06:07 and 06:11 UT, 9 Dec. 1971.

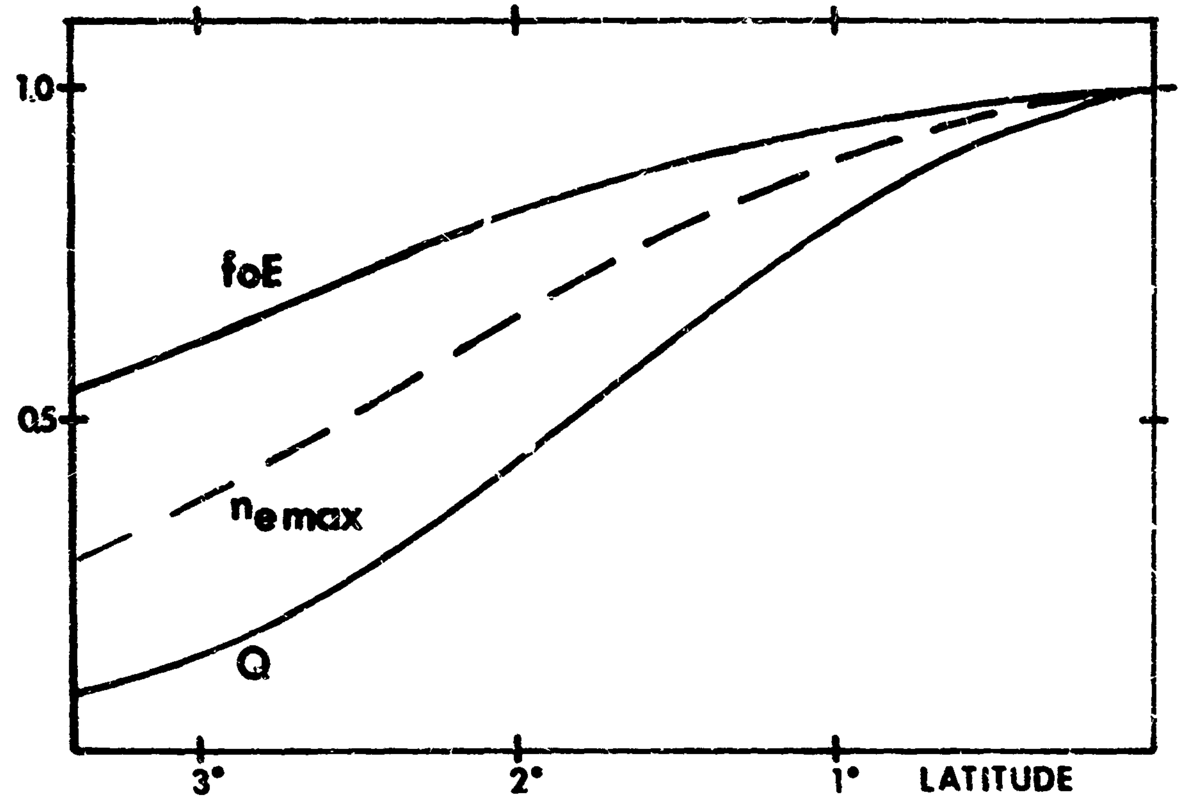


Figure 8a Gaussian latitudinal profile of Q (3.7° FWHM) shown together with the resulting maximum electron density, nemax, and E layer critical frequency, foE. All three parameters are normalised to unity at the maximum; the latitude of the maximum is taken as the origin of the latitude scale; and only the low latitude half of the distribution is shown.

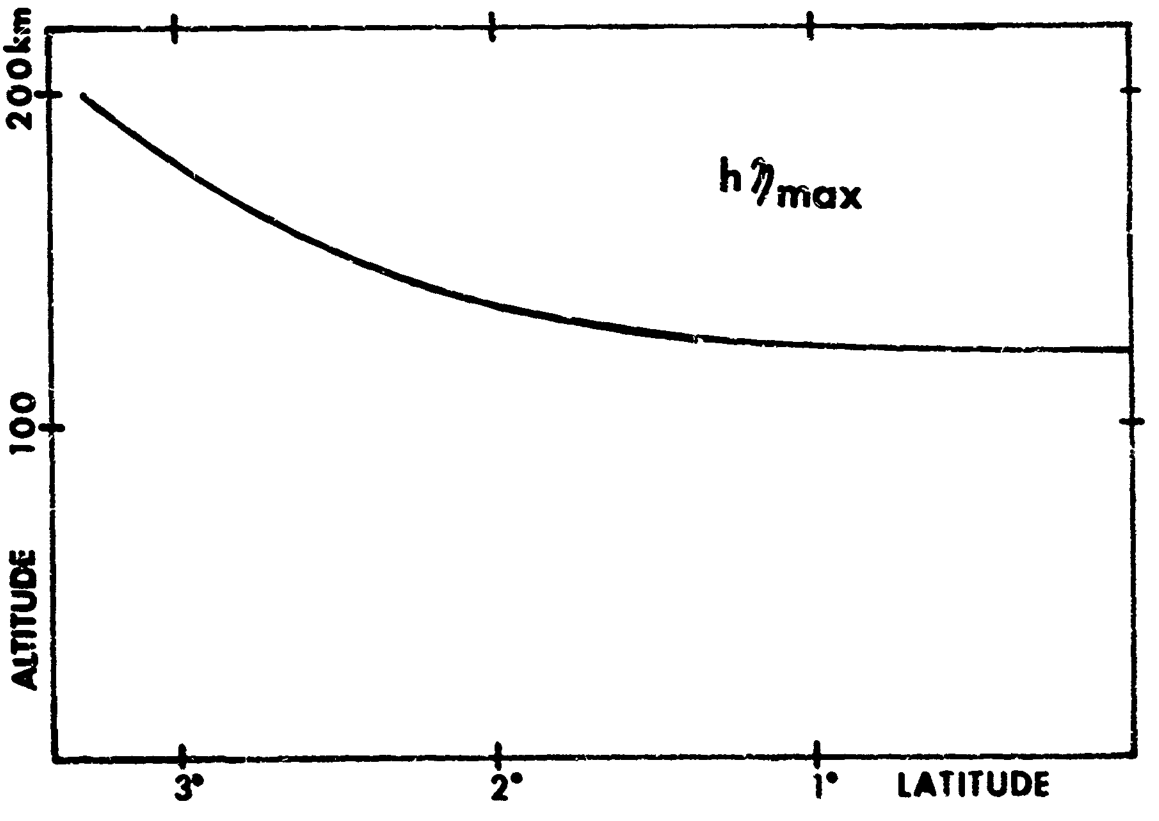


Figure 8b The latitudinal variation of height of maximum rate of ion production, $h\eta_{max}$, is plotted for a Gaussian latitudinal profile of characteristic energy, α , which is the same as that shown for Q in Figure 8a.

Perspective on the Prediction of Auroral Absorption

by
Vaughn Agy
Institute for Telecommunication Sciences
National Telecommunications and Information Administration
U. S. Department of Commerce
Boulder, Colorado 80303
United States of America

ABSTRACT

Ever since the Second Polar Year (1932-33), when a British team made the first observations of the Arctic ionosphere, special high frequency (HF) communication problems have been recognized at high latitudes. Auroral absorption is one of these special problems that has received considerable attention over the years. As many as a dozen methods have been described for the calculation of the effects of auroral absorption on HF communication circuits (or HF radar), and perhaps several dozen papers have appeared that applied more or less directly, at least qualitatively, to the operation of such circuits. A few of the methods and pertinent points from some of the papers are briefly described.

In 1956, the riometer was shown to be a powerful tool for use in the study of high latitude absorption. In particular, it allowed for a distinction to be recognized between "polar cap" absorption and what now came to be called "auroral" absorption. There are good reasons for considering the riometer our best hope for establishing an adequate data base for the prediction of auroral absorption effects on HF communication circuits. There are other reasons to explain why, after 25 years of riometer measurement, there is still no adequate data base. These reasons will be reviewed.

Study given the phenomenon--known variously as: "the no-echo condition," "polar blackout," "auroral zone blackout," and now, "auroral absorption"--has been fruitful, and there have been "breakthroughs," but much more remains to be done. Questions remain about the geographic distributions, the temporal variations, and the relationships with polar cap absorption on one hand, and with magnetospheric substorms on the other. These questions will be stated and briefly examined.

For auroral absorption, as for most high-latitude ionospheric phenomena, the beginning came during the Second International Polar Year (1932-33) when a team from the United Kingdom (UK) made vertical-incidence ionospheric observations at Tromsø, Norway. These observations were described in the classic paper by Appleton et al. (1937). The "no-echo" condition was ascribed to absorption associated with geomagnetic activity. Beginning during the nighttime hours, the absorption could continue into daytime, in some cases, beyond the end of the magnetic disturbance. Wells (1947) showed that at College, Alaska, "polar blackout" accompanied negative magnetic bays (decreases in the horizontal geomagnetic component) occurring in the early morning hours. The duration of the blackout approximated that of the bay.

These studies treated occurrences of "no-echo" or "blackout" (most of which would now be called auroral absorption) as relatively discrete events in association with geomagnetic disturbance. Newbern Smith (NBS, 1948, Chapt. 7) however, recognized the need for a different kind of treatment so that the effect of this absorption could be considered in the computation of monthly median field intensity for high-latitude high-frequency (HF) circuits. The method consists of calculating, using the map in Figure 1 (p. 150, NBS, 1948), a quantity, K', to be added to the diurnally varying absorption factor, $\bar{K}$. Smith (private communication) based the approach largely on discussions with Canadian colleagues who had noted, for example, that less absorption was observed over paths perpendicular to the auroral zone than over paths with a direction component parallel to the zone. In addition, paths wholly within (i.e., poleward of) the zone may show no auroral absorption even when radio waves propagating across the zone are seriously attenuated. Smith mentions that auroral absorption is most intense during daytime, an observation not indicated by earlier work. However, he did not try to incorporate into his method either the suggested diurnal variation or the slight equatorward displacement of the absorption zone relative to the visual auroral zone.

In the early 1950's, a number of papers appeared (e.g., Lindquist, 1951a; Meek, 1953) continuing study of the apparent relationships among observations: geomagnetic, auroral, ionospheric. But, in addition, Meek (1951) used 2 MHz LORAN data to locate the auroral absorption zone and Lindquist (1951b), divorcing blackout and geomagnetic activity, showed that the diurnal maxima for the two phenomena occurred at very different times. Agy (1954a) studied HF field strength data from a chain of stations along the 90th west meridian in an attempt to locate the auroral absorption zone. Vertical-incidence ionospheric data from eighteen northern high latitude stations were used (Agy, 1954b) to describe geographic and temporal variations in the occurrence of blackout, and to differentiate phenomenologically between "long-duration" blackouts and the "others." The frequency of occurrence and the overall total number of hours of blackout are greater in the "auroral absorption zone," but the duration of the long-lasting ones increases with latitude. The riometer, described by Little and Leinbach (1958), has been especially useful in explaining this anomaly, i.e., in differentiating between "polar cap absorption" and auroral absorption.

During the International Geophysical Year (IGY, July 1, 1957 - December 31, 1958), when a total of about 50 vertical-incidence ionosondes were in operation at high northern latitudes, there were fewer than a dozen riometers. In spite of its recognized shortcomings, fmin (the lowest frequency evident on a vertical-incidence ionogram, and, therefore, a rough measure of absorption) continued as the parameter to be used, especially for derivation of detailed geographic distributions of absorption at high latitudes during the IGY. Individual "events" involving polar cap absorption were described by Sakura (1960) and by Agy (1960). Studies aimed at determining the geographic and temporal distributions of "blackout" were undertaken by Kasuya (1960), Agy [in Davies (1965)], and Piggott and Thomas (1959). None of this work helped in any material way in calculating auroral absorption as it affects HF circuits.

The method described in NBS Circular 462 was known to be over-simplified, and the data analyses mentioned served to indicate its inaccuracies and conceptual errors. However, as recently as 1969, the Circular 462 method was revived for application to HF circuits over paths longer than 4,000 km (Vladimirov, 1969). In the mid-1940's, U. S. Signal Corps methods for calculation of transmission loss for HF circuits (Laitinen and Haydon, 1962) made use of an empirical "correction" of about 9 dB to bring calculated losses more nearly into agreement with measurement. The 9 dB was termed "excess system loss." Excess system loss was redetermined by Davis and Groome (1965), who included data from high latitude paths in their study and could conclude that excess system loss was greatest for paths at geomagnetic latitudes between 65° and 70°, and for such paths, had a diurnal maximum between 0400 and 1000 local mean time. It is clear, then, that high latitude excess system loss is largely due to auroral absorption. But since the analysis indicated no clearcut frequency dependence, other effects must also be present. In any case, based as it is on "real" data and available in tabular form, excess system loss can be included in either "hand" or computer methods by means of "table look-up" which is more convenient and quicker than computation. In the United States, specifically at the Institute for Telecommunication Sciences (ITS) or its predecessors, excess system loss replaced the method of Circular 462 when the computer replaced hand methods of computation (Lucas and Haydon, 1966) and continues in use (Barghausen et al., 1969; Lloyd and Teters, in preparation). There appears to be nothing better. However, the possible utility of riometer data for this purpose has long been considered. Analyses of riometer data in Norway (Holt et al., 1962), in Canada (Hartz et al., 1963); in the United States in Alaska (Basler, 1963), and in the Union of Soviet Socialist Republics (Driatskiy, 1966), all agreed in showing an "auroral absorption zone" (i.e., auroral absorption occurring most frequently in a limited range of latitude somewhat south of the visual auroral zone) and a local time variation in frequency of occurrence that maximizes during the pre-noon hours. Since riometer data are quantitative and since many riometers have been in operation for some years in Canada, Norway, Sweden, Finland, the U.S.S.R., and the United States (Alaska), the possibility of establishing a world-wide data base has occurred to many people interested in high-latitude, high-frequency telecommunications. In a preliminary way, a number of attempts have been made to devise schemes for the use of riometer data for the prediction of auroral absorption experienced by high-frequency radio waves: Basler (1965), Gorbushina et al. (1969), Agy (1970), Vargas-Vila (1972), and Foppiano (1975). Unfortunately, none of these authors was able to give adequate consideration to either of two sets of problems: those arising from the characteristics of the (riometer) data base, and those associated with our lack of understanding of the nature of what we call auroral absorption.

Although, in a manner of speaking, the riometer provides the only game in town, being the only data source that can be described as both quantitative and world-wide, there are problems associated with its use:

1. The riometer frequency (~30 MHz) is necessarily much higher than most of the communications frequencies in the HF spectrum to which we want to apply the measurements. The resulting insensitivity might still be overcome by using enough data. The tendency, however, has been to study relatively brief periods of time, or individual events, so most of the riometer records (estimated to be over 500 station-years worth) in the northern high latitudes remain unscaled).

2. At most installations, a finite threshold of 0.2 dB (even 0.5 dB) has been adopted for the scaling, and an absorption level of 0.2 dB measured by riometer at 30 MHz can easily correspond to 30 dB absorption at [illegible] MHz over a high latitude path.

3. The methods so far described rely on deriving the distribution of absorption values from those actually observed (i.e., above the threshold), but a threshold of 0.2 dB will be exceeded infrequently at most stations and for most months. We are in a position of deducing the entire distribution from the upper quartile--or the top decile. For most HF, the low absorption tail of the distribution is what matters, and that can only be guessed at so long as so high a threshold is insisted upon.

4. The riometer measures "opacity" so that the height at which the attenuation takes place cannot be specified without other information. One effect of the non-specific height is that scatter from the top of Es will be interpreted as absorption; but, when such scatter occurs, ground-to-ground HF propagation may be especially efficient. This possibility has been treated only briefly in a single paper (Bell and Morozumi, 1970).

So much for the data base. The problems in the other set are associated with certain aspects of the phenomenon itself which are either misunderstood, misstated, or ignored.

1. Location

a. Latitude. A latitude variation has been recognized from the beginning. Hartz and his co-workers (1963), after comparing Canadian and Norwegian data, concluded that geomagnetic latitude was not the appropriate coordinate. Since then, it has become common to assume that invariant latitude (McIlwain, 1961) or corrected geomagnetic latitude (Hakura, 1965) is appropriate, although why this should be true is not quite clear. The magnetic regimes, day and night, at the latitudes of most frequent occurrence, are very different: at night, the field lines are swept back into the magnetospheric tail, and in the day, the lines are "closed," i.e., "conjugacy" has meaning. According to Hartz and Brice (1967), the type of causative particle precipitation also changes (from night to day) from "splash" to "drizzle." Figure 2, taken from Hartz et al. (1963), exhibits a latitude variation--and a local time variation to be discussed presently.

b. Longitude. A distribution of auroral absorption non-uniform in longitude was first suggested almost twenty years ago (Agy, 1960; Davies, 1965). Even now, however, the likelihood of such a distribution is either denied out-of-hand or studied inadequately. Figure 3, based on an analysis of fmin (Davies, 1965), suggests a longitude distribution with peaks over northern Europe and over Alaska. Berkey et al. (1971) made a study of 60 "substorm events" using data from about 40 high latitude northern hemisphere riometers. Agy (1975) analyzed the results of this study to show that, in spite of editing of the data by Berkey et al. (1971) which would tend to obscure a variation with longitude, such a variation was still evident. Figure 4 shows the longitude variation and indicates rough agreement with the longitude variation derived earlier (Davies, 1965) using fmin.

c. Auroral oval. The "auroral oval," described by Feldstein (1963), is an annular area around the geomagnetic pole, eccentric toward lower latitudes at local midnight. Visual aurora occurs in the auroral oval; it encompasses the "instantaneous polar cap." The "auroral zone," on the other hand, may be thought of as the region traced out by the midnight segment of the auroral oval as it moves in universal time. It should be emphasized that auroral absorption rarely occurs in the auroral oval except in the hours around local midnight.

2. Time Variations

a. Solar cycle. It is likely that the diurnal maximum in frequency of occurrence moves to lower latitudes as the sunspot number increases, and the frequency of occurrence probably increases (Hook, 1968), but there may be a lag between long-term variations in solar activity and auroral absorption (Collins et al., 1961). Our knowledge here is incomplete--to say the least.

b. Season. For the pattern as a whole, i.e., world-wide and at all latitudes, most of us do not know (cannot describe) the seasonal variations. Those of us who do know do not agree with each other (Hartz et al., 1963; Driatskiy, 1966; Basler, 1966; Hook, 1968).

c. Local time. The strong local time dependence already indicated in Figure 2 has been known for a long time. It is likely that the local time variation is dependent on longitude and/or season and/or solar cycle, but we don't know.

d. Magnetospheric substorm. Does auroral absorption occur only during magnetospheric substorms (as implied by Berkey et al., 1971, and by Hargreaves, 1966)? In view of some of the editing by Berkey et al. (1971, p. 1), the answer seems to be, "Only if the terms 'auroral absorption' and/or 'substorm' are redefined."

The purpose of the report by Berkey et al. (1971) was to establish a data base for later study (Berkey et al., 1974) of the motion of the absorbing regions, in particular, the eastward and/or westward drift of these regions. However, despite prudent selection of the events and a careful editing of the data, the "drift" is not always evident. On the other hand, development in situ of auroral absorption in the areas of most frequent occurrence (Agy, 1975) is quite common, even usual. Figures 5 and 6 give map sequences from Berkey et al. (1971) illustrating this point.

Even more useful than a procedure for calculating monthly median auroral absorption, would be one for short-term forecast of the occurrence, development, and motion of an auroral absorption event as suggested by Elkins (1972). His approach [after having established the existence of a correlation between the magnetic disturbance index, AE, and the development of the substorm events described by Berkey et al. (1971)] would be to identify the start of a substorm by noting AE near the midnight meridian and, during the next 3/4 hour or so to project the absorbing region around to the meridian corresponding to about 0900 local time. It is here that auroral absorption (for this event) will tend to be a maximum. The editing by Berkey et al. (1971) is aimed at simplifying the contours of auroral absorption by: deleting substorm events still in progress when "this" event begins along with those that start before this one ends; and by ignoring the long-lasting (co-rotating with the earth) absorption which has "evidently nothing to do with the substorm effects." Possible consequences of the editing and of the "bootstrap" approach which does not relate the "substorm events" to substorms (perhaps better defined by reference to magnetograms) will be to render Elkins' approach untenable. For example, Figure 6 represents an event that overlaps a magnetospheric substorm whose ionospheric effects have been described by Park and Meng (1973). But the substorm, according to Park and Meng, started an hour before the riometer event. According to Elkins' proposed model, during that hour the absorbing region should have drifted eastward some 135 degrees in longitude and maximized in intensity. Of course, conclusions based on a statistical analysis cannot be invalidated by a single contrary instance; but the restrictions placed on the events described by Berkey et al. (1971) by their definition and by subsequent editing of the data must be kept in mind. The limitations set are such that we may no longer be dealing with the "real world" of auroral absorption at all.

It is not necessarily true that "anything is better than nothing." Until the time comes that we must no longer guess at the time distribution of riometer measurements of auroral absorption (and, in particular, the low absorption tail of that distribution), we are well-advised to do nothing. For most of the currently available riometer data, the important part (from the standpoint of the user of HF) of the distribution is lost below the effective threshold of the instrument and/or the scaling procedure. Until we are prepared to scale (or rescale) immense piles of riometer recordings, any other effort will be pointless: we have reached a dead end.

REFERENCES

Agy, V. (1954a), The location of the auroral absorption zone, J. Geophys. Res., 59, 267.

Agy, V. (1954b), The geographic and temporal distribution of polar blackouts, J. Geophys. Res., 59, 499.

Agy, V. (1960), Polar blackouts during the international geophysical year--an animated film, part of the oral presentation by V. Agy and K. Davies (1961), Worldwide patterns of ionospheric blackout occurrence, J. Atmos. Terr. Phys., 23, 202.

Agy, V. (1970), HF radar and auroral absorption, Radio Sci., 5, 1317.

Agy, V. (1975), On the geographic distribution of auroral absorption, J. Atmos. Terr. Phys., 37, 681.

Appleton, E. V., R. Naismith, and L. J. Ingram (1937), British radio observations during the second international polar year 1932-33, Phil. Trans. Roy. Soc. A., 236, 191.

Barghausen, A. F., J. W. Finney, L. L. Proctor, and L. D. Schultz (1969), Predicting long-term operational parameters of high-frequency sky-wave telecommunication systems, ESSA Tech. Rept. ERL 110-ITS78.

Basler, R. P. (1963), Radio wave absorption in the auroral ionosphere, J. Geophys. Res., 68, 4665.

Basler, R. P. (1965), A method for predicting auroral path losses over HF propagation circuits, presented at U. S. Spring USNC-URSI meeting, April 20-24, Washington, D. C.

Basler, R. P. (1966), Annual variation of auroral absorption, J. Geophys. Res., 71, 982.

Bell, C. D. and H. M. Morozumi (1970), A comparison of absorption measured by riometer and by ionosonde at high latitudes, J. Atmos. Terr. Phys., 32, 257.

Berkey, F. T., V. M. Driatskiy, K. Henriksen, D. H. Jelly, T. I. Shchuka, A. Theander, and J. Yliniemi (1971), Temporal development of the geographical distribution of auroral absorption for 30 substorm events in each of IQSY (1964-1965) and IASY (1969), Rep. UAG-17, World Data Center A, National Atmospheric and Oceanic Administration, U. S. Department of Commerce, Boulder, Colorado.

Berkey, F. T., V. M. Driatskiy, K. Henriksen, D. H. Jelly, T. I. Shchuka, A. Theander, and J. Yliniemi (1974), A synoptic investigation of particle precipitation dynamics for 60 substorms in IQSY (1964-1965) and IASY (1969), Planet. Space Sci., 22, 255.

Collins, C., D. H. Jelly, and A. G. Matthews (1961), High-frequency radio-wave blackouts at medium and high latitudes during a solar cycle, Canadian J. Phys., 39, 35.

Davies, K. (1965), Ionospheric Radio Propagation, NBS Monograph 80, U. S. Government Printing Office.

Davis, R. M. and N. L. Groome (1965), The effect of auroral zone absorption on high frequency system loss, U. S. Department of Commerce, NBS Report 8810.

Driatskiy, V. M. (1966), A study of the space and time distribution of auroral absorption according to observations of the riometer network in the Arctic, Geomag. Aeron., 6, 828.

Elkins, T. J. (1972), A model of auroral substorm absorption, AFCRL-72-0413, Environmental Research Report No. 404.

Feldstein, Y. I. (1963), Some problems concerning the morphology of auroral and magnetic disturbances at high latitudes, Geomag. Aeron., 3, 183.

Foppiano, A. J. (1975), A new method for predicting the auroral absorption of HF skywaves, United Kingdom contribution to section D3 of revised CCIR Report 252-2.

Gorbushina, G. N., V. M. Driatskiy, and E. M. Zhulina (1969), Instruction for the calculation of short wave radio links at high latitudes, Arctic Antarctica Institute, Acad. Sci., U. S. S. R.

Hakura, Y. (1960), Polar blackout and auroral zone blackout, J. Radio Res. Lab., Japan, 7, 583.

Hakura, Y. (1965), Tables and maps of geomagnetic coordinates corrected by the higher order spherical harmonic terms, Rep. Ionos. Space Res. Japan, 19, 121.

Hargreaves, J. K. (1966), Auroral absorption of HF radio waves in the ionosphere--a review of results from the first decade of riometry, Proc. IEEE, 57, 1348.

Hartz, T. R., L. E. Montbriand, and E. L. Vogan (1963), A study of auroral absorption at 30 Mc/s, Canadian J. Phys., 41, 581.

Hartz, T. R. and N. M. Brice (1967), The general pattern of auroral particle precipitation, Plan. Space Sci., 15, 301.

Holt, D., B. Landmark, and F. Lied (1962), Analysis of riometer observations obtained during polar radio blackouts, J. Atmos. Terr. Phys., 23, 229.

Hook, J. L. (1968), Morphology of auroral zone radiowave absorption in the Alaska sector, J. Atmos. Terr. Phys., 30, 1341.

National Bureau of Standards (1948), Ionospheric Radio Propagation, U. S. Department of Commerce, NBS Circ. 462.

Kasuya, I. (1960), Statistical study in the occurrence of polar blackouts, J. Radio Res. Lab. Japan, 6, 451.

Laitenen, P. O. and G. W. Haydon (1962), Analysis and prediction of sky-wave field intensities in the high frequency band. U. S. Army Signal Corps, Radio Prop. Agency, Fort Monmouth, N. J., Tec. Rept. No. 9 (Revised Oct., RPU203).

Little, C. G. and H. Leinbach (1958), The riometer--a device for the continuous measurement of ionospheric absorption, Proc. IRE, 47, 315.

Lindquist, R. (1951a), A survey of recent ionospheric measurements at the ionospheric and radio wave propagation observatory at Kiruna, Arkiv Geophys., 1, 247.

Lindquist, R. (1951b), Polar blackouts recorded at the Kiruna Observatory, Trans. Chalmers Univ., Gothenburg Res. Lab. of Elect., Rept. No. 16.

Lucas, D. L. and G. W. Haydon (1966), Predicting statistical performance indexes for high frequency ionospheric telecommunications systems, ESSA Tech. Rept. IER 1-ITSA-1.

McIlwain, C. E. (1961), Coordinates for mapping the distribution of magnetically trapped particles, J. Geophys. Res., 66, 3681.

Meek, J. H. (1951), reception of 2 Mc/s loran in central Canada, Defence Research Board, Canada, Rept. No. 9.

Meek, J. H. (1953), Correlation of magnetic, auroral, and ionospheric variations at Saskatoon, J. Geophys. Res., 58, 445.

Park, C. G. and C. I. Meng (1973), Distortions of the nightside ionosphere during magnetospheric substorms, J. Geophys. Res., 78, 3828.

Piggott, W. R. and L. Thomas (1959), Studies in polar blackout morphology, URSI-AGI, Committee Report.

Vargas-Vila, R. (1972), Auroral absorption predictions for high-latitude HF propagation, presented at U. S. Spring USNC-URSI meeting, April 20-24, Washington, D. C.

Vladimirov, V. (1969), Simplified method for calculating auroral absorption for distances of over 4000 km, Telecom. J., 36, 175.

Wells, H. W. (1947), Polar radio disturbances during magnetic bays, Terr. Mag. Atmos. Elec., 52, 315.

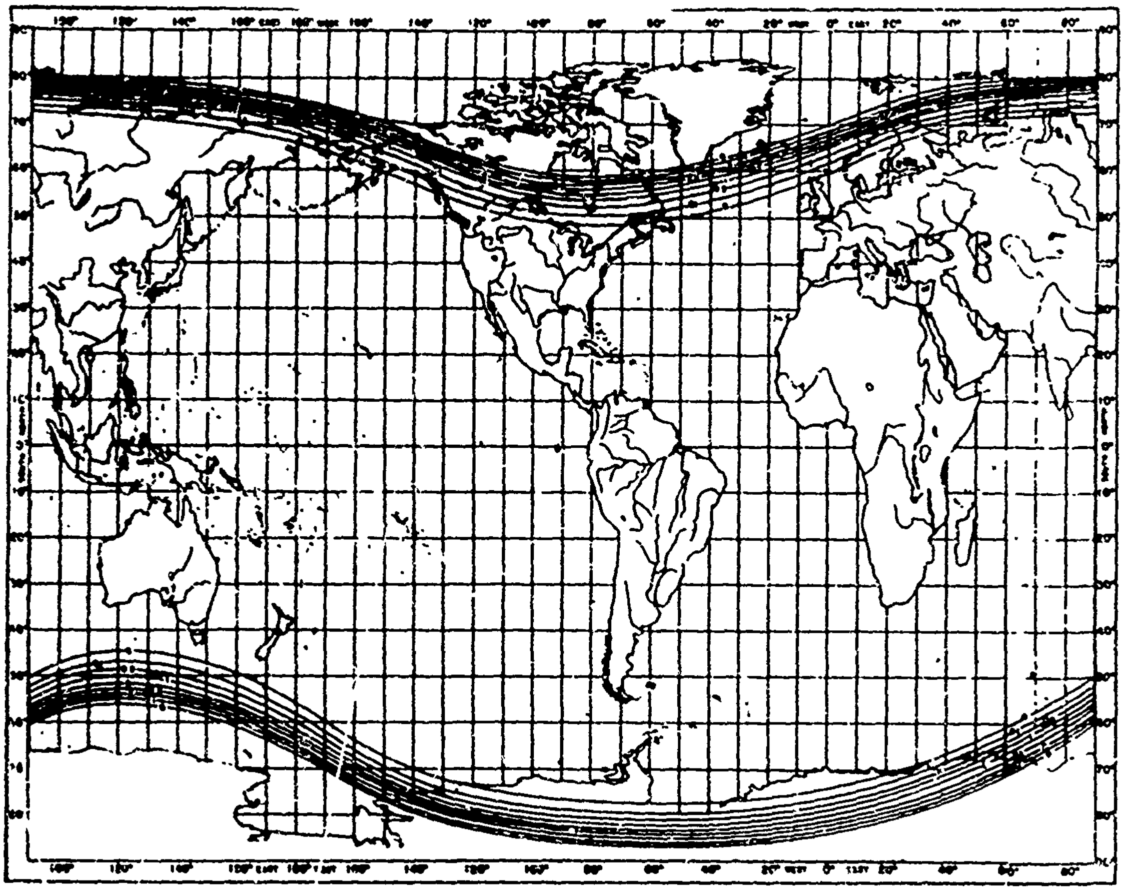


Figure 1. Auroral zone absorption map, numbers on curves are absorption factor, K'
(from Ionospheric Radio Propagation, NBS Circular 462, 1948).

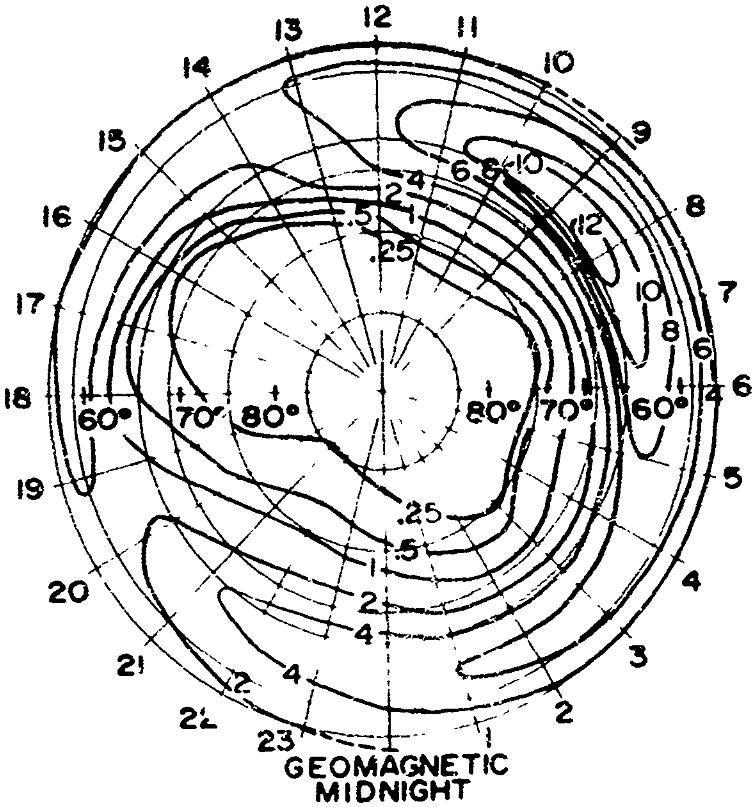


Figure 2. Percent time auroral absorption of 30 MHz cosmic noise, $\geq$ 1 dB (from Hartz et al., 1963).

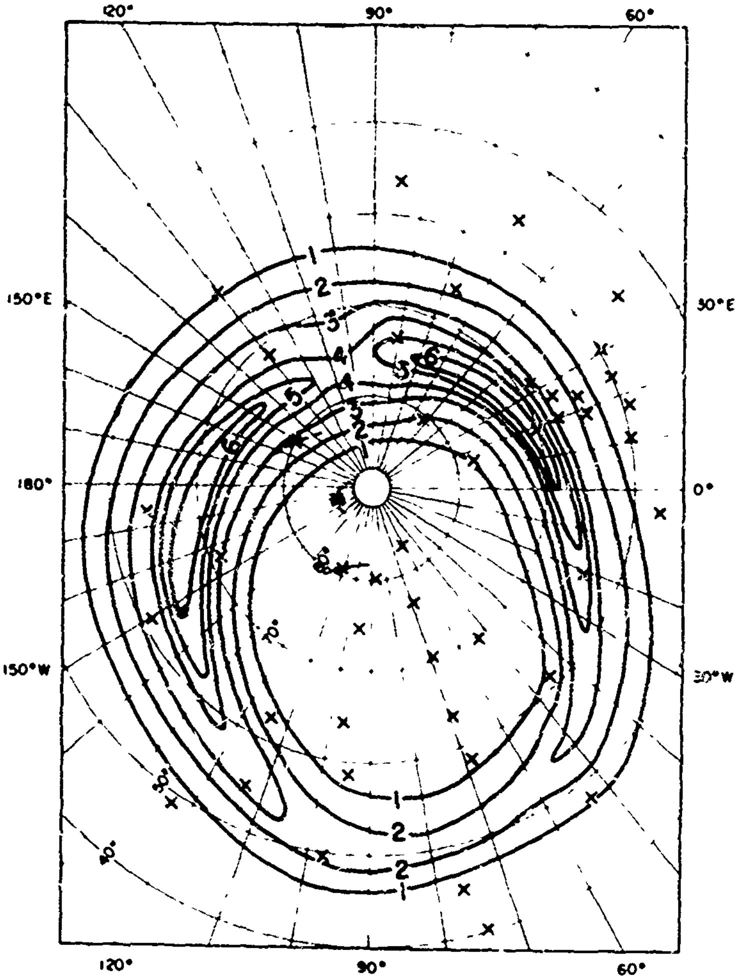

120°
90°
60°
150°E
30°E
180°
0°
150°W
30°W
1
2
3
4
5
6
30°
40°
70°

Figure 3. Mean percent time of IGY "auroral zone" blackout occurrence - Northern Hemisphere.

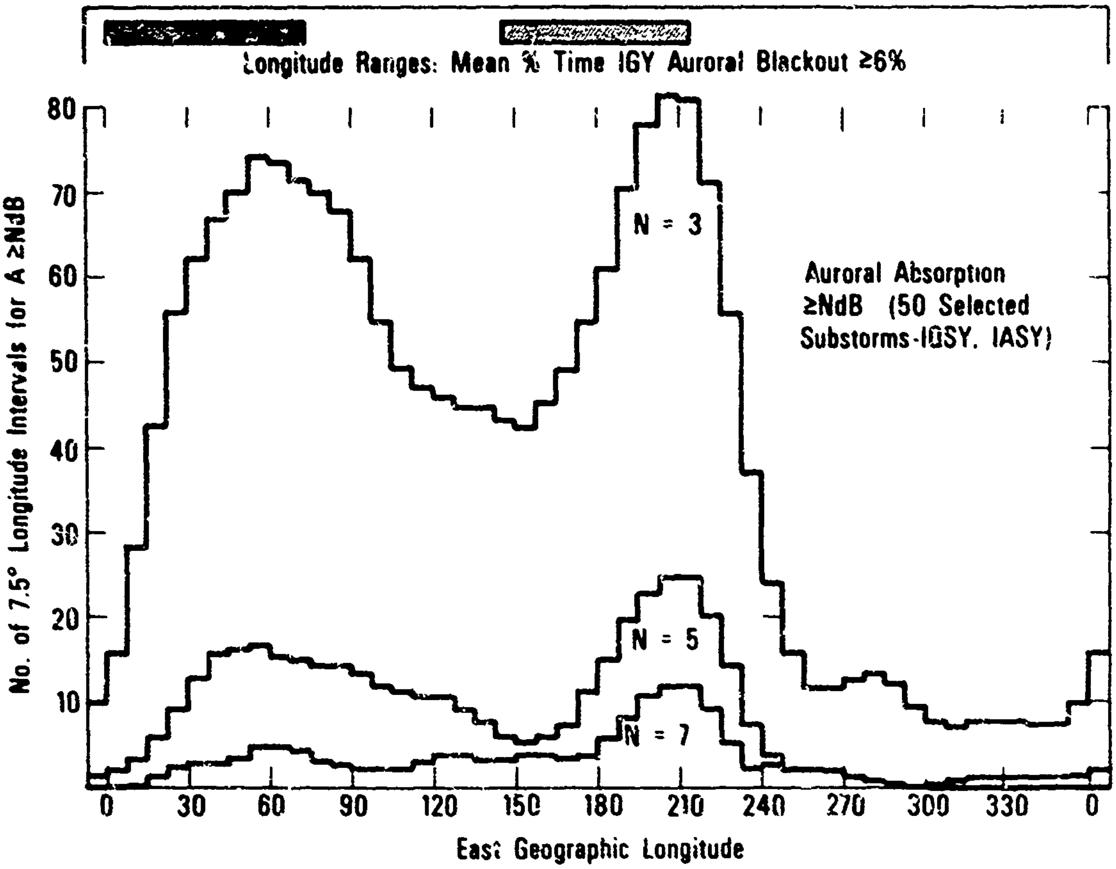


Figure 4. Auroral absorption $\geq$ N dB, (50 selected substorms - IQSY, IASY).

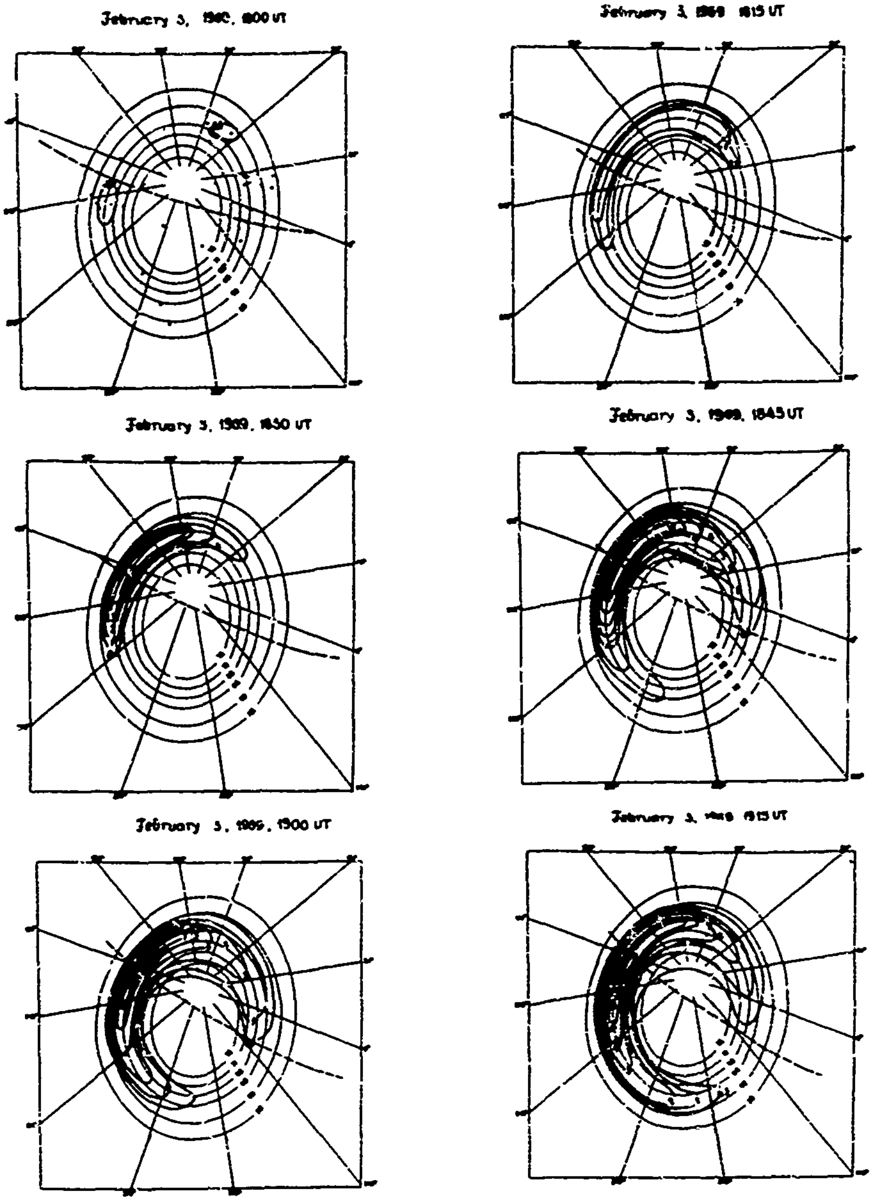
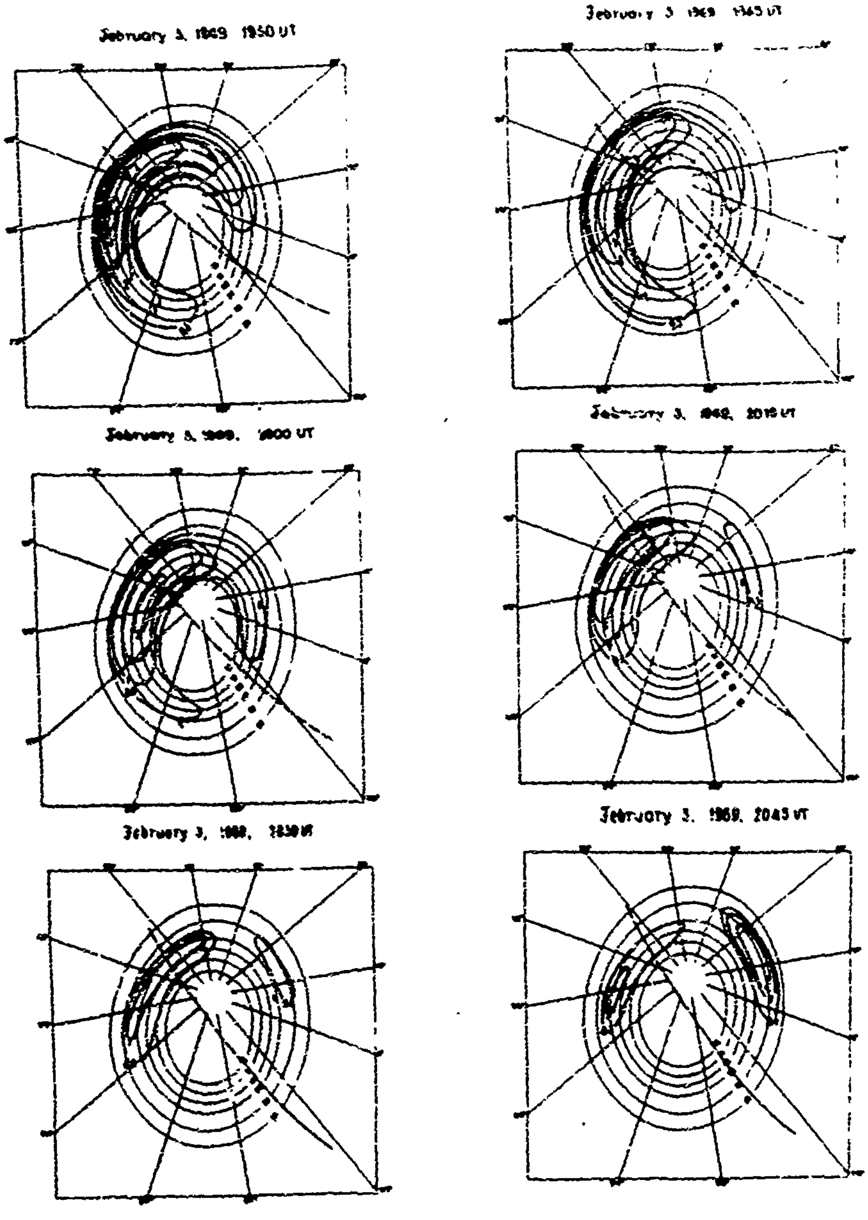

February 3, 1969, 1800 UT
February 3, 1969 1815 UT
February 3, 1969, 1830 UT
February 3, 1969, 1845 UT
February 3, 1969, 1900 UT
February 3, 1969 1915 UT

Figure 5(a). Auroral absorption patterns, February 3, 1969, (from Berkey et al., 1971).



February 3, 1969 1750 UT

February 3, 1969 1945 UT

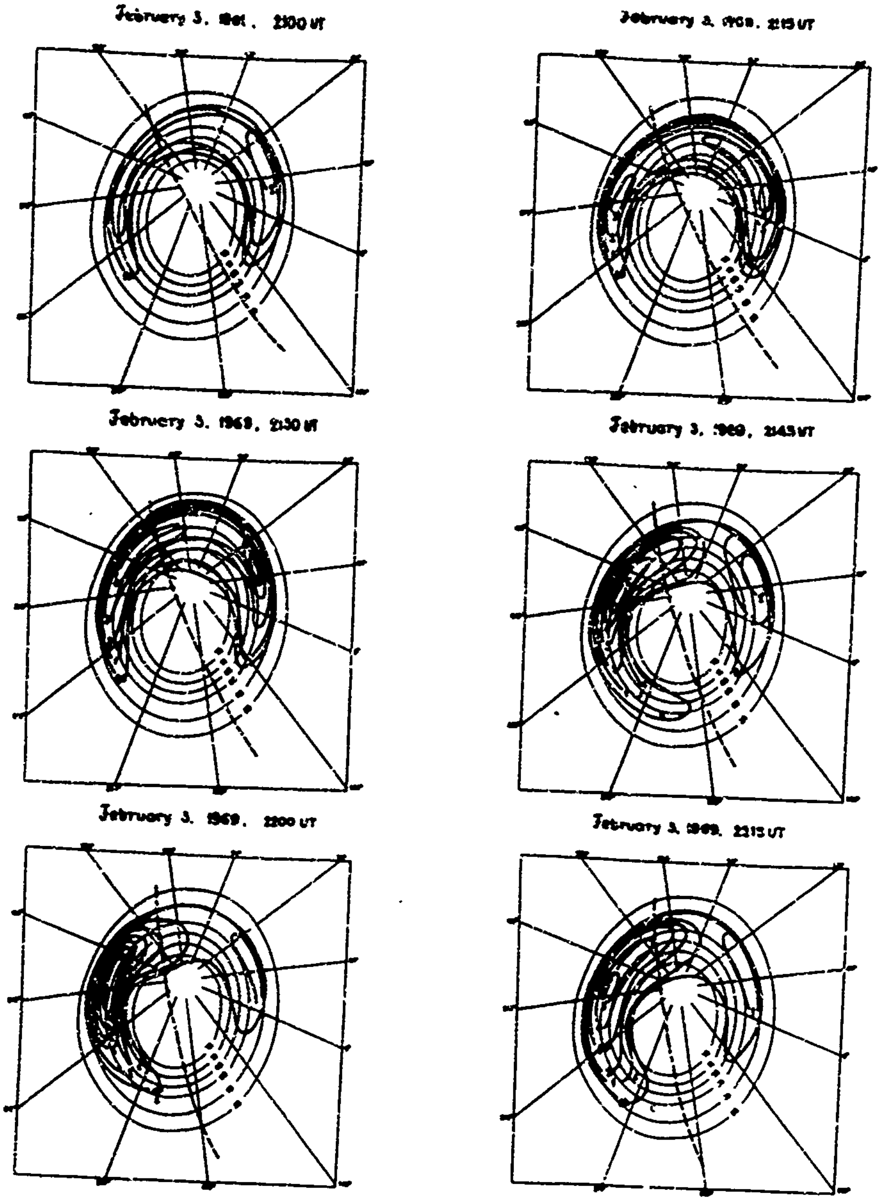
February 3, 1969, 1800 UT

February 3, 1969, 2015 UT

February 3, 1969, 1830 UT

February 3, 1969, 2045 UT

Figure 5(b). Auroral absorption patterns, February 3, 1969, (from Berkey et al., 1971).



February 3, 1969, 2100 UT

February 3, 1969, 2115 UT

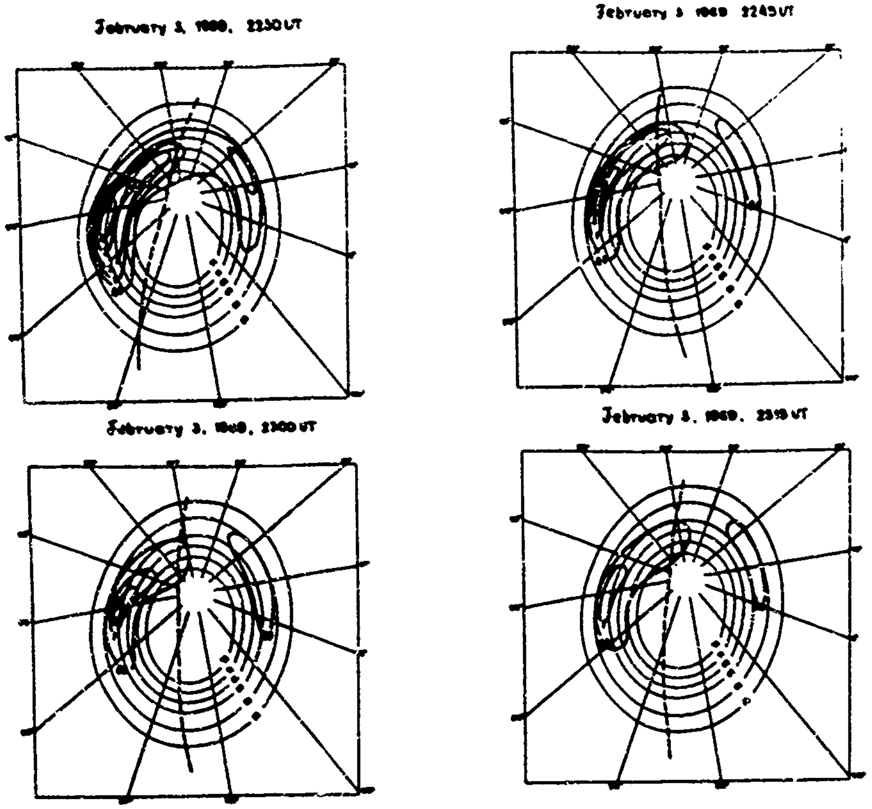
February 3, 1969, 2130 UT

February 3, 1969, 2145 UT

February 3, 1969, 2200 UT

February 3, 1969, 2215 UT

Figure 5(c). Auroral absorption patterns, February 3, 1969, (from Berkey et al., 1971).



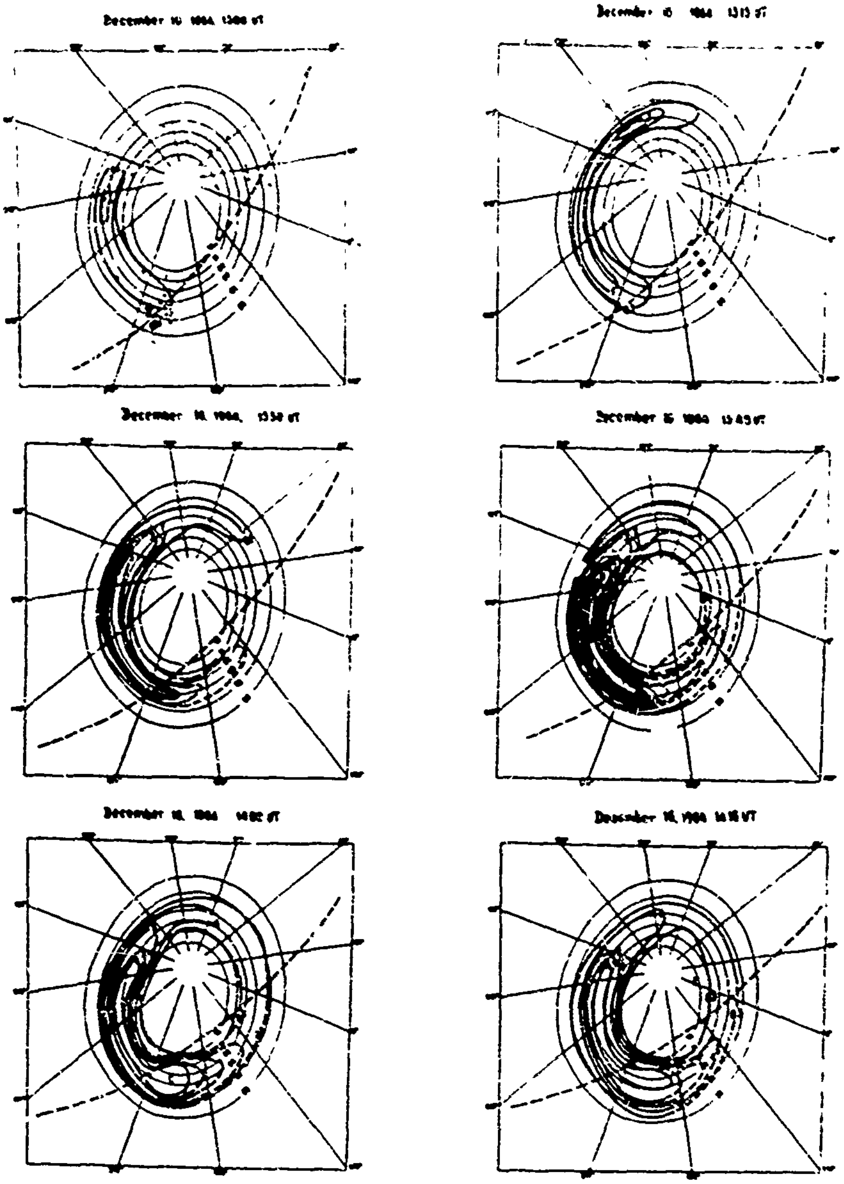
February 3, 1969, 2230 UT

February 3 1969 2245 UT

February 3, 1969, 2300 UT

February 3, 1969, 2315 UT

Figure 5(d). Auroral absorption patterns, February 3, 1969, (from Berkey et al., 1971).



December 16 1964 1300 UT

December 16 1964 1315 UT

December 16, 1964, 1330 UT

December 16 1964 1345 UT

December 16, 1964 1400 UT

December 16, 1964 1415 UT

Figure 6(a). Auroral absorption patterns, December 16, 1964.

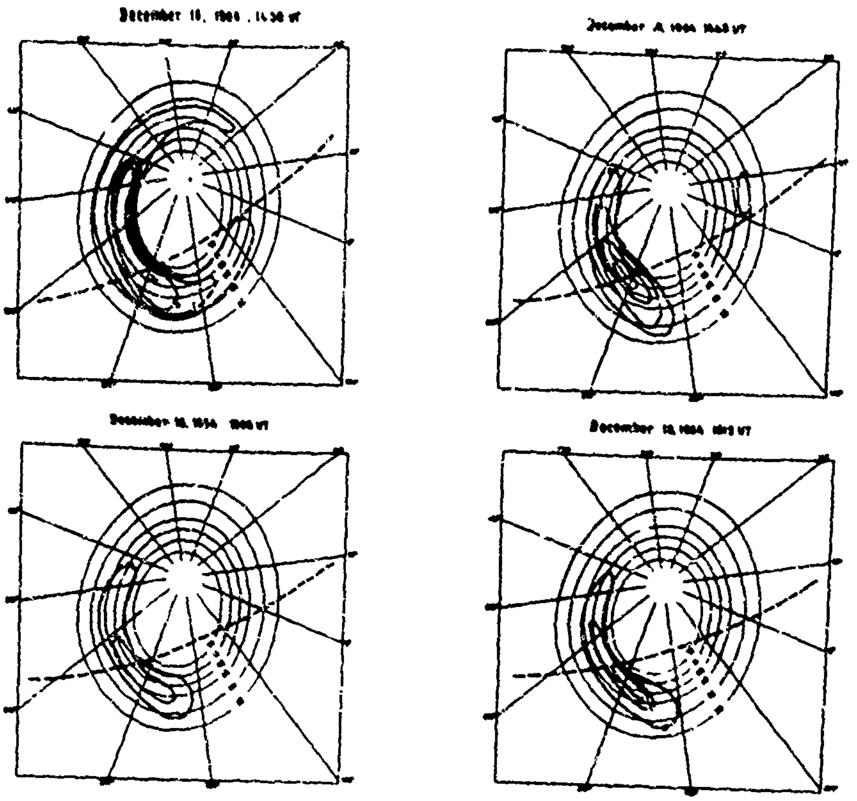


Figure 6(b). Auroral absorption patterns, December 16, 1964.

DISCUSSION

P.A. Bradley, UK

I support Mr Agy's comments that there is a need to analyse the large amounts of riometer data currently within the world data centers and also to carry out further observations. It is only by detailed studies of such results that we can hope to achieve a major improvement in our understanding of the morphology of auroral absorption. Unfortunately there are major difficulties with existing data in that there has been no standardization of measurement and record-scaling procedures. In particular times of measurement and types of antenna used vary considerably. What really is needed is to produce international measurement and scaling standards in the same way as has been done for vertical incidence ionospheric sounding. I fear though, within the current economic climate, that we are too late for this, and we must in the main make do with the data we already have.

Many riometer installations are operated principally in support of rocket and other special investigations where the interest as in auroral absorption events. For these applications a fixed threshold of detection of the absorption of a few tenths of a decibel at 30MHz is not serious, but I agree this is troublesome when these data are used for morphological mapping. Useful results can be obtained in such cases none the less by extrapolation of values measured when there is significant absorption. For example, Foppiano has shown from data collected in the middle of the northern auroral zone at a time of day when there is much absorption that the day to day variations of absorption with a month follow a log-normal law. This law can then be assumed for other times and locations in order to deduce for example the monthly median absorption, even when this is below the threshold of detection.

The critical question is how riometer data can be related to vertical-incidence absorption. Mr Agy has mentioned difficulties due to sporadic-E ionisation. Other factors influencing riometer data are (1) residual absorption inherent in the reference calibration curve of maximum received cosmic noise power (2) the effects of a finite core of noise reception, within which the noise is incident thereby experiencing correspondingly greater absorption than for vertical incidence and (3) absorbing ionisation patches not always filling the whole of the antenna beam, thereby reducing the measured absorption. Quantitative tests have been made between scaled values of riometer absorption and absorption deduced from calculations based on simultaneously measured vertical electron concentration profiles deduced from rocket flights and incoherent scatter observations. To a first order it is concluded that riometer absorption and vertical incidence absorption on a single ionospheric traverse are the same.

I would like to comment on the absorption prediction procedure of Foppiano mentioned. This is based on riometer data from available stations in the northern hemisphere grouped into four separate longitude zones. His mapping is matched to the Hartz and Brice precipitating particles patterns and assumes the absorption to be the resultant of two components due to the splash and drizzle particles, each varying in a gaussian manner with latitude and time of day. Longitude, seasonal and solar-cycle changes are also incorporated.

Some recent work in conjunction with Dr E Thrane of NDRE has compared circuit reliability on oblique N-S HF sky-wave paths in Norway with predictions. Measured data at night indicate poor performance at the lower frequencies giving only a narrow window of useful frequencies. This is consistent with predictions which include the Foppiano absorption model. By comparison, predictions based on the excess-system loss concept which ignores the frequency dependence of the auroral absorption predict increasing circuit reliability the lower the frequency. While I concede that the Foppiano absorption model is tentative in many aspects, I submit that nonetheless it provides a useful first-order way of assessing quantitatively auroral absorption effects on system performance for high latitude paths.

Author's Comment

Of course, I can't argue with success, but a single satisfactory comparison with measurement is not enough to "prove" the validity of the approach. Many more comparisons must be made. I repeat the distribution of auroral absorption values used by Foppiano is guessed at — it is not known. Also, incorporated or not, the longitude, seasonal and solar cycle variations of auroral absorption are not known. The effect of sporadic E is quite different from the others mentioned. increased riometer "absorption" in the case of sporadic E may well be accompanied by increased signal over the circuit rather than in signal attenuation.

Comment by Klaus Bibl, US

There is a large amount of auroral absorption data available in the compressed digital soundings of the Goose Bay, Labrador, Station of the US Air Force Geophysics Laboratory. The optimum presentation of digital amplitudes in dB as function of frequency and time of day, separated for E- and F- region echoes, would permit the relatively easy analysis of auroral absorption over many years and a calibration of the riometer measurements.

DIRECTION AND DOPPLER CHARACTERISTICS OF MEDIUM AND LONG PATH HF SIGNALS WITHIN THE NIGHT-TIME SUB-AURORAL REGION

R.W. Jenkins, E.L. Hagg and L.E. Montbriand
Department of Communications
Communications Research Centre
P.O. Box 11490, Station H
Ottawa, Canada
K2H 8S2

SUMMARY

During January 27 - February 4, 1976, the large HFDF receiving array near Ottawa was used to record HF transmissions from a USAF Geophysics Laboratory aircraft 2000 km away over the northwest Atlantic on four evening flights. Geomagnetic conditions ranged from quiet to active. Signals from individual array elements were separately recorded, and later analyzed to provide the Doppler shift and direction of the signal. Also used were oblique ionograms recorded on the aircraft for a similar path.

The F-mode MUF's were below 5 MHz for the major portions of all evening flights. However, several weaker modes were observed to be present at much higher frequencies. They are examined for their potential utility in extending HF communications and surveillance:

(i) Sporadic-E was a persistent feature on the oblique records; a vertical ionosonde on board the aircraft did not see the same phenomenon. This mode had a path centered close to great-circle with typical spreads in bearing of 8 degrees.

(ii) Skip-distance-focussed ground sidescatter to the south was persistently present. Position estimates of the scatterers made from Doppler and direction measurements imply severe multipaths for this mode which will sometimes limit data rates to less than 25 baud.

(iii) A mode involving scatter from auroral irregularities was observed which was extremely variable in time and generally weaker than the other modes.

(iv) Occasional reflections from the northern edge of the mid-latitude trough in ionospheric electron density were seen, but were limited to frequencies less than 1 MHz above the F-mode MUF.

It is concluded that the Sporadic-E and skip-distance-focussed ground-sidescatter modes present an opportunity for limited HF operations in the subauroral region during evening hours, when very low F-mode MUF's would otherwise make such operations impossible.

1. INTRODUCTION

Nighttime HF operations in the sub-auroral regions are hampered by the existence of very low ionospheric electron densities associated with the mid-latitude trough (Muldrew, 1965; Stanley, 1966). The resulting low MUF's for F-mode communications force many users into a small, congested range of frequencies at the lower end of the HF spectrum where equipment problems (e.g. poor antenna efficiencies and directivities) and ionospheric absorption tend to be worst.

The present paper is concerned with the results of an HF experiment involving highly-sensitive detection of signals propagated over a relatively long distance (~2000km) in the vicinity of the nighttime trough in electron density. In addition to the normal F-modes whose MUF's were, for much of the experiment, below the frequencies used, several other modes were found to be present most of the time, at frequencies well above the F-mode MUF. These modes included:

(a) Sporadic E,

(b) Skip-distance-focussed ground sidescatter (from the south where F-MUF's were higher),

(c) Scattering from auroral irregularities.

Another mode, sometimes seen at frequencies slightly above the F-mode great-circle MUF, was that of oblique reflection from the northern edge of the ionospheric trough.

These modes are either weaker or less predictable in their occurrence than the usual F-mode and are not normally considered in HF system planning. However, they present a possible means of extending HF coverage at subauroral latitudes for both communications and surveillance and should be considered.

The experiment was performed using HF transmissions from a series of five aircraft flights, four in the evening and one in the afternoon. Thus the results should not be regarded as a very statistically significant sample of the variety of propagation conditions that can occur during evening hours at subauroral latitudes. However, various geomagnetic conditions ranging from quiet ($K_p=0$) to active ($K_p=5+$) were experienced for these flights, and the results can be considered as representative of at least some of the conditions that occur with reasonable frequency. With the intent of providing information for system planners, the present paper examines the modes of propagation that were present during evening hours when the normal F mode was unavailable as a result of its MUF having dropped below the available frequencies. Parameters considered were those indicative of the potential of these modes for communications and surveillance, i.e., their relative amplitudes, stability, direction of arrival and Doppler shift characteristics, and inferred multipath properties.

2. EXPERIMENT DESCRIPTION

During the period Jan 25-Feb 4, 1976, the Canadian Department of Communications HFDF receiving system (Rice and Winnacott, 1977) near Ottawa was used to record unmodulated HF transmissions from a specially-equipped USAF Geophysics Laboratory aircraft flying out of Goose Bay, Labrador, over the northwest Atlantic on a series of one afternoon and four evening flights. The geometry for each flight was approximately as illustrated in Figure 1. The great-circle propagation path to Ottawa was expected to lie to the south of the auroral oval (Feldstein and Starkov, 1967), within the mid-latitude trough in ionospheric electron density, during the evening flights. Oblique ionograms, made on board the aircraft using FM/CW transmissions from Rome, N.Y. (230 km south of Ottawa), yielded additional information about the modes of propagation.

Figure 2 illustrates the antenna configuration for the HFDF system for this experiment. It consisted of a 1219 m by 244 m crossed array of 42 and 16 vertical monopoles respectively. The signals at the antennas were fed through matched cables to separate fixed-gain receivers controlled by a common synthesized frequency. The resultant base-band in-phase and quadrature components were digitally sampled and recorded on magnetic tape. The data sampling rate for each antenna was 128 Hz. Aliasing problems were avoided by the use of 30 Hz low-pass filters on the output stages of the receivers.

A two-dimensional (i.e., Doppler shift and direction of arrival) analysis was performed on the tape-recorded data. The Doppler analysis consisted of a cosine-squared preweighted Fourier analysis in time, over a 4-second integration time. Only the innermost -16 to +16 Hz of the resultant 128 Hz-wide Doppler window proved to be of interest and was retained for display. The direction analysis consisted of a cosine-squared preweighted Fourier analysis in position performed separately for each arm of the crossed array. The long paths and resultant low elevation angles allowed a one-to-one correspondence to be made between a particular direction cosine and azimuth, at least for the signals that came from directions other than near-parallel to the array arm used. A computer line printer was used to encode and print out the resultant signal strength as a function of Doppler shift and direction cosine (interpreted as azimuth). The display encoding was carried out using a set of characters arranged in 4 dB steps, representing the signal level for each particular azimuth and Doppler shift. Resolution was 0.3 Hz in Doppler, and 3 degrees in azimuth for most of the directions of interest, at the lowest transmitted frequency. A contour plot of signal level versus azimuth and Doppler shift, taken from a line printer display, is shown in Figure 3. No attempt was made at this point to remove the shift due to the transmitter/receiver frequency offset from the data.

The narrow-band filtering of the signal combined with the processing gain achieved by the direction and Doppler analysis facilitated the examination of very weak signals. The direction analysis permitted a useful dynamic range of 32 dB at any one time, for signals with the same Doppler shift but different directions. When the Doppler shifts differed as well, the dynamic range was much greater.

Data were recorded on the HFDF system on each of three transmitted frequencies (5.621, 8.989, and 11.136 MHz), for eight seconds, once every 5 or 10 minutes throughout the flights. Corresponding oblique ionograms were obtained at 15 minute intervals. For the purposes of this paper, the evening data were restricted to that obtained for times between 0050 and 0515 UT (approximately 2050 to 0115 LT at the mid-point of the HF propagation path) and aircraft locations within the area illustrated in Figure 1. The afternoon flight data, which was used only to obtain an estimate of the F-mode signal strengths for comparison purposes, was taken on January 25 between 2040 and 2230 UT (1640-1830LT) when the aircraft was in the same portion of its flight path.

3. RESULTS

3.1 Available Modes of Propagation

Based upon inspection of the Doppler vs direction displays and oblique ionograms, it was found that, when the F-mode MUF's were too low for propagation between the aircraft and Ottawa, nearly all the reception at Ottawa could be classified into 3 modes, all of which were present for most of the evening periods of analysis. Figure 3 is a contour plot taken from a Doppler vs. direction display for a time and frequency where all three modes appeared with nearly equal intensity. The three modes are as follows:

(a) Sporadic E (E_s) Propagation

Sporadic E extending to frequencies well above those transmitted by the aircraft was observed on most of the oblique ionograms taken during the evening flights. The corresponding HFDF displays showed strong signals, with only a narrow Doppler spread in any direction (<0.5 Hz), spread over a set of directions which usually included that of the great-circle. The maximum signal strengths appeared to be almost randomly distributed in direction, and to move significantly in the 5 to 10 minute gap between measurements. The Doppler shift showed only a weak variation with direction. Such signals for the most part were easily identified on the displays, even at times when they were too weak to be seen on the oblique ionograms, and were categorized as "E_s".

(b) Auroral Signals (A)

Auroral signals, denoted by "A", appeared to be unrelated to the E_s signals in time or location. These signals were seen to the north of the great-circle direction, mostly late in the evening, and were thought to be the result of sidescattering from ionospheric irregularities located in a region of auroral activity. The Doppler shifts often showed a strong variation with direction. Sometimes this variation was well-ordered, i.e., the signals lay along well-defined Doppler shift versus azimuth "traces" such as in Figure 3; at other times signals appeared whose Doppler varied randomly with bearing. The observed shifts could not be explained solely in terms of the aircraft motion. At least several Hertz of Doppler shift were normally attributed to the motions of the auroral scatterers. Doppler spreads were typically more than 1 Hz.

(c) Skip-distance-focussed Ground Sidescatter from the South (SS)

These signals, categorized as "SS", were relatively weak but persistent, appearing whenever the display sensitivity was sufficient. They occurred well to the south of great-circle, and lay along well-defined traces on the Doppler shift vs. azimuth displays. They often arrived from an azimuth more than 90 degrees south of the great-circle direction and required the displays from both arms of the crossed array for their interpretation. The Doppler shift vs. azimuth behaviour could be explained in terms of the aircraft motion and scattering from a relatively immobile set of scatterers. These signals were attributed to skip-distance-focussing of ground sidescatter (to the south where F-MUF's were expected to be higher than that of the great-circle MUF's), a mode which has been observed previously on trans-Atlantic circuits (Hagg and Rolfe, 1963).

A fourth observed mode which appeared on two of the evening flights was that of oblique propagation from the northern edge of the ionospheric trough in electron density. This mode had an MUF 0.5 to 1.0 MHz in excess of the great-circle F-mode MUF, it was seen at Ottawa on the lowest transmitted frequency only, and it persisted for about an hour after the great-circle F-mode had disappeared. Its properties differed on the two flights for which it was seen. One set of observations from a geomagnetically quiet period indicated little spreading in Doppler shift, direction of arrival or path-length (determined from the oblique ionograms), and relatively little Doppler shift (<1 Hz) due to ionospheric motion. The other observations from a geomagnetically active period indicated significant spreading of the signal direction and path length, and larger Doppler shifts due to ionospheric motion, presumably as a result of a more disordered trough boundary.

It should be noted that in addition to the above, short-lived modes, which were likely of meteoric origin, were occasionally observed. These are not considered further in this paper.

3.2 Relative Amplitudes

The strengths in dB of the three types of frequently occurring modes, Sporadic-E (E_s), auroral (A), and skip-distance focussing of ground sidescatter (SS) were examined for the four evening flights. In order to obtain values which would be of meaning to operators and system planners, the amplitudes were measured relative to typical F-mode signals. The F-mode signals used for this purpose were taken from data obtained during the one-hour period in the late afternoon immediately prior to the disappearance of the F-mode when the MUF dropped below the transmitted frequency. The January 25 afternoon flight was used for this purpose, for the 8.989 and 11.136 MHz frequencies, and all 5 flights were used for the 5.621 MHz frequency. The aircraft location at these times was similar to that for the portion of interest of the evening flights.

The resulting distributions of signal strength for the E_s, A, and SS modes are shown in Figures 4A, 4B, 4C and 4D for the evening flights. Mean values and standard deviations for these distributions are given in table 1.

TABLE 1: Mean signal strengths and standard deviations for the three modes: Sporadic E (E_s), Auroral (A), and Southern Skip-distance-focussed ground sidescatter (SS) Side-Scatter (SS), for the four evening flights.

Frequency (MHz)	Mode	Mean signal strengths and standard deviations in dB relative to typical F-mode signal strengths.			
		Jan. 27	Jan. 30	Feb. 1	Feb. 4 (U.T.)
5.621 MHz	E_s	-10±8	-25±6	-29±3	-36±6
	A	-26±4	-41±6	-38±7	-35±6
	SS	-34±3	-29±4	-30±5	-32±5
8.989MHz	E_s	-40±9	-53±6	-54±8	-56±6
	A	-----*	-----*	-61±11	-68±9
	SS	-59±4	-47±4	-48±6	-49±4
11.136MHz	E_s	-53±10	-66±6	-64±5	-68±6
	A	-----*	-----*	-69±12	-----*
	SS	-74±4	-56±4	-64±8	-63±4

* For more than half the measurements, the signal level was below the background level and could not be obtained.

Not suprisingly, the signal strengths seen for these modes relative to the corresponding F-mode were strongly frequency dependent. Typical values for the 3 modes were -30 dB at 5.621 MHz, -50 dB at 8.989 MHz, and -65 dB at 11.136 MHz. However, there was considerable scatter about these values for the various modes and flights.

Significant differences exist between the various days. January 27, a geomagnetically quiet day, experienced significantly stronger sporadic-E than the other days and, at the same time, weaker skip-distance focussing of sidescatter. This was true at all three frequencies. Longer skip-distances and a poorer antenna response at the resulting lower elevation angles, rather than increased absorption, was likely the cause of the reduced sidescatter, since the effect was more noticeable at the higher frequencies. Estimates of the location of the ground-sidescatters presented later in this paper corroborates this conclusion. The sporadic-E on this day was remarkably high in strength, especially at the 5.621 MHz frequency, when one considers that for some of the path lengths involved (~2500 km), at least two hops were required for E_s propagation.

The three modes differed significantly in their amplitude variations. The auroral mode was noted to be the most variable with time; it exhibited measurement-to-measurement changes as high as 26 dB and varied as much as 40 dB over an evening. Large changes in signal strength, as much as 26 dB in mean value, were noted for the E_s mode between the four evening flights; during the course of an evening moderate changes (typically 6 dB rms) were experienced. The mode experiencing the least variation during the course of an evening was the SS mode. Its standard deviation (typically 4 dB) was generally the lowest and its measurement-to-measurement change (2.8 dB average) was about half that of the other two modes.

The four evening flights are shown in Figure 5. Similar patterns were found for the other two frequencies. Several things are apparent in this figure:

(i) the E_s mode dominated propagation on the evening of January 27;

(ii) the SS mode tended to dominate propagation for the other three evenings;

(iii) the E_s mode varied considerably over the course of each evening;

(iv) the A mode tended to be weakest of the three, but had large rapid increases and a general increase after local midnight (~0400 UT) which allowed it to dominate at times.

The fraction of measurements for which each of the three modes dominated the propagation are plotted for the three frequencies and four evening flights in Figure 6. The dominance of the E_s mode for the January 27 measurements and the SS mode for the other evenings is readily apparent. The observed relative strength of the E_s mode for appreciable periods of time makes it potentially interesting for extending communications and surveillance to frequencies that would not normally be available. The relative persistence and frequent dominance of the SS mode are reasons for examining this mode further. The auroral mode is of lesser interest owing to its relative weakness, much greater time variation and undesirable Doppler characteristics. The remainder of this paper will concentrate on the E_s and SS modes and their Doppler and direction properties. The other modes, i.e. the auroral mode and trough-edge mode, will be considered in the light of the problems their occasional appearance may introduce.

3.3 Sporadic E Mode (E_s)

The E_s mode observed during the flights did not show a clear association with auroral activity. The very strong E_s mode on January 27 was especially surprising as geomagnetic activity was low for that evening and visual auroral observations made on board the aircraft (J. Buchau, personal communication) indicated extremely quiet auroral conditions with only a weak continuous aurora observed low on the horizon.

The HFDF measurements suggest that the E_s mode is a persistent feature of the subauroral region during evening hours. General studies of sporadic E occurrence at high latitudes relying on vertical ionosonde data (Gassman, 1971; Wagner and Pike, 1971) restrict persistent E_s to the Feldstein auroral oval (Feldstein and Starkov, 1967). The southern boundary of the instantaneous Feldstein oval (Reilly, 1977; Whalen, 1970) was determined to be several degrees to the north of the great-circle for the present flights. However, a vertical ionosonde on board the aircraft indicated little or no sporadic E for flights when strong E_s was seen to persist on the oblique records (J. Buchau, personal communication). Therefore, the oblique measurements, which are a more direct indication of modes available for HF propagation, imply a persistent and sometimes relatively strong E_s mode which is not necessarily observed on vertical incidence sounding records, and extends further south than previously suggested.

A typical E_s-propagated signal, as seen on the HFDF displays, is illustrated in Figure 7. It consists of an enhanced region relatively narrow in Doppler and elongated in bearing, having a slight change in Doppler with bearing, with one or more maxima along its length.

The bearings of maximum signal strength were tabulated; the resulting distributions are plotted in Figure 8. When an equally strong signal was experienced in several directions at once, the count for that measurement was divided among those directions. For operational purposes, the strength of the E_s mode relative to other modes will determine its usefulness. Thus the distributions were divided on the basis of the relative strength of the E_s mode. No significant differences between the three frequencies of transmission were found, and their results were combined in order to improve the statistics.

The results of Figure 8 are an indication of the limits on bearing accuracy that may be obtained by surveillance systems extending their times and frequencies of operation through use of the E_s mode. When the E_s mode was by far the strongest mode (> 20 dB), its direction of maximum nearly always lay within 2 degrees of the great circle direction. Slightly weaker E_s modes (8 to 16 dB above the next strongest mode) showed much more spread in direction, with 25% having a bearing more than 5 degrees away from great-circle. This spreading of arrival directions away from great-circle became even more pronounced when the E_s mode dropped to a level comparable to or less than the next strongest signal. The weaker E_s modes also revealed a tendency for the propagation to occur slightly northward of great-circle, where more E_s would be expected.

The spreads in bearing (taken between the 4 dB points) of the E_s-propagated signal were scaled from the line-printer displays and tabulated; the resulting distributions are shown in Figure 9. The data for all three frequencies were combined. A separation was made on the basis of signal strength relative to the other modes. The results of Figure 9 follow a pattern similar to that of the previous figure. For times when the E_s mode was very strong (≥ 20 dB) relative to the other modes, the median azimuthal spread was 4 degrees when it was moderately stronger (8 to 16 dB), the median spread was 8 degrees; and for times when the E_s mode was approximately the same strength as the next strongest mode, the median spread was 16 degrees.

These spreads in azimuth and the previously-noted differences from great-circle in the directions of maximum strength are larger than that normally experienced for conventional F and E-mode propagation, and will limit the effectiveness of certain signal enhancement techniques (e.g., adaptive beam-forming) which are designed for signals arriving from a single well-defined direction.

The Doppler behaviour noted for the E_s mode is well illustrated in Figure 7. In any one direction the spread was quite narrow, often less than the 0.3 Hz resolution of the experiment. However, the E_s signal was spread in azimuth and experienced a slight change in Doppler shift with azimuth. This change was due to the motions of the aircraft and E_s scattering centers, each of which contributed differently in the different directions. An analysis of the motions involved in the Figure 7 example (8.989MHz, 0251 UT, January 30) is illustrated in Figure 10. For the purpose of the analysis, the angle between the propagation path and great-circle was assumed to be the same at the aircraft as at Ottawa, i.e. the E_s was assumed to be located at the midpoint of the path; however, the basic conclusions do not change when the relative sizes of the two angles are changed. The aircraft contribution was computed from the velocity information obtained from the aircraft log, and subtracted from the observed frequency shift for the various directions. In the great-circle direction, E_s motion cannot cause any significant Doppler shift because of the geometry involved; the shift in that direction was therefore the transmitter-receiver offset. This was then subtracted from all directions. The final residual Doppler shift was that due to the component of E_s motion perpendicular to the great circle direction. The observed negative component south of great-circle, becoming positive north of great-circle, implied a component of motion in the southward sense of the order of 100 m/sec. Over the 20 degrees spread in azimuth experienced for the example of Figure 7, the E_s motions gave rise to a change in Doppler of approximately 2 Hz. Over the more typical 8 degrees spread in arrival directions, E_s motions of the order of 100 m/sec would cause a spreading in Doppler of 0.5 to 1.0 Hz.

3.4 Skip-distance-focussed Ground Sidescatter (SS mode)

A typical example of skip-distance-focussed ground sidescatter, at 8.989 MHz, is illustrated in Figure 11. The results from the line printer displays for both arms of the array were combined to make this plot. Contour levels of signal strength were plotted as a function of bearing and Doppler shift. The SS mode signals were generally seen to lie along one or several preferred "traces", in this case, two lines running from bottom left towards upper right.

The Doppler shift of the SS mode signals could be assumed to be entirely due to the component of aircraft velocity in the direction of the ground scatterers, since the scattering was from ground (or sea) irregularities whose motion was negligible relative to that of the aircraft. Thus, the bearing of the ground scatterers at the aircraft could be determined from the measured Doppler shift and known aircraft velocity vector. This, together with the measured bearing at Ottawa, enabled a rough triangulation estimate of the ground scatter locations.

The location estimates thus derived for the 8.989 MHz measurements illustrated in Figure 11 are plotted in Figure 12. The distances involved in the SS mode can be seen to be very large, with resultant propagation times varying from 13 to 40 ms. The positions of the local maxima, numbered 1 to 8 in Figure 11, were used in the estimates. The resulting locations can be seen in Figure 12 to fall into one of two groups, according to the line along which they lay in Figure 11, numbers 1 to 5 being in the closer group, and numbers 6 to 8 about twice as far from the great-circle joining Ottawa and the aircraft. The two groups appear to correspond to one-hop and two-hop F-mode skip distances. Also shown in Figure 12 are the concurrent results for 5.621 MHz. The derived locations for this frequency fall into two similarly-spaced groupings which lie somewhat closer than the corresponding groupings for the higher frequency. This is consistent with the skip-distance focussing explanation; shorter skip distances are expected at lower frequencies.

Location estimates of ground scatterers were made for a number of examples from the four evening flights. Results were in general similar to those of Figure 12. The SS-mode propagation times found from these estimates are plotted in Figure 13 for the three frequencies and four evening flights. Several trends are apparent in this figure. The general increase in propagation time with frequency can be seen by comparing the distributions for the three frequencies for any single flight. Differences exist between the flights. The propagation times (and distances involved) for January 27 tended to be higher than those for the other three flights, presumably as a result of longer skip distances. The correspondingly lower elevation angles at Ottawa combined with the reduced antenna gain at low angles could explain the lower SS-mode signal amplitudes noted earlier for that flight.

The very long delay times and the spread in delay times noted for the SS mode will limit the rate of communications. The SS mode was often observed (see Figure 5) to be comparable in amplitude to the much shorter-path E_s mode. The resulting multipath interference can be expected to limit the rates of communication via these modes to less than 25 baud at times.

3.5 Auroral (A) Mode and Trough-Edge Mode

These modes, which were noted to occur to the north of great-circle were either normally weak, with occasional large increases in strength (A-mode), or extremely limited in frequency (trough-edge). They are not as potentially useful for communications and surveillance as the E_s and SS modes. However they may have a disruptive effect on communications or surveillance systems by virtue of their occasional strong appearance and should be considered in this light.

The delay times for these modes would not be expected to be as long as those of the SS mode. Thus, for a system operating at data rates suitable to the multipath conditions of the SS mode, the auroral and trough-edge modes should pose no problem. However, for direction-finding systems these modes will cause the sporadic appearance of strong signals in directions far removed from the transmitter bearing. The auroral propagation tended to be highly variable in its direction of arrival, Doppler shift, and amplitude, as well as normally possessing several Hz spread in Doppler, and so might be discriminated against on this basis. The same is not true for the trough-edge mode: it was observed on January 27 to possess a near constancy in direction, a narrow Doppler spread, and a significant relatively constant strength for at least one hour at the 5.621 MHz frequency.

4. CONCLUSIONS

The present results on medium and long-path HF propagation suggest the persistent existence of two relatively weak propagation modes in the evening subauroral region which may be used to extend the range of operating frequencies normally used for communications and surveillance well above the very low F-mode MUF's normally present for that time and region:

(i) A sporadic-E mode, the existence of which is not necessarily indicated on vertical ionograms. This mode is characterized by a narrow Doppler spread, it is centered close to great-circle, and it possesses an amplitude-dependent spread in direction of the order of 8 degrees for moderately strong propagation. This mode shows appreciable variation in strength from evening to evening, as well as moderate variations within the course of an evening (see Table 1). Slight Doppler shifts in off-great-circle directions (~1 Hz) occur as a result of ionospheric motions.

(ii) Skip-distance-focussed ground sidescatter. This mode occurs in directions well to the south of great-circle, and involves propagation times much in excess of typical great-circle modes. It is more constant in strength than the E_s mode. It was visible throughout the four evening flights, whenever experiment sensitivity warranted its being seen. Doppler shifts observed for this mode were explained in terms of the aircraft motion.

The relatively weak nature of these modes (Table 1) would have to be considered in planning for their use. The severe multipath effects inherent in the ground-sidescatter mode place additional constraints on the data rate achievable when using that mode. Thus these modes can be regarded as a means of allowing limited HF communications at times when extremely low F-layer MUF's make normal communications impossible.

Acknowledgments

We would like to thank Drs. T. Elkins, J. Buchau, and G. Sales of the USAF Geophysics Laboratory, who arranged for the aircraft transmission and oblique-ionogram recordings. They also provided many helpful discussions.

References

Feldstein, Y.I. and G.V. Starkov, 1967, "Dynamics of Auroral Belt and Polar Geomagnetic Disturbances", Vol. 15, p. 209, Planet. Space Sci.

Gassman, G.J., 1971, "On Modelling the Arctic Ionosphere", AGARD Conference Proceedings 97, Radar Propagation in the Arctic.

Hagg, E.L., and W. Rolfe, 1963, "A Study of Trans-Atlantic Radio Propagation Modes at 41.5 Mc/s", Vol. 41, p220, Can. J. Phys.

Muldrew, D.B., 1965, "F Layer Ionization Troughs Deduced from Alouette Data", Volume 70, p2635, J. Geophys. Res.

Reilly, A.E., 1977, "Analysis of Sweep Frequency Oblique Polar Region High Frequency Radio Propagation Measurements", Report No. AFGL-TR-77-0102, USAF Geophysics Laboratory.

Rice, D.W., and E.L. Winnacott, 1977, "A Sampling Array for HF Direction-Finding Research", CRC Report 1310, Department of Communications, Ottawa.

Stanley, G.M., 1966, "Ground-Band Studies of the F-region in the Vicinity of the Midlatitude Trough", Volume 71, p. 5067, J. Geophys. Res..

Wagner, R.A. and C.P. Pike, 1971, "A Discussion of Artic Ionograms", AGARD Conference Proceedings 97, Radar Propagation in the Arctic.

Whalen, J.A., 1970, "Auroral Oval Plotter and Nomograph for Determining Corrected Geomagnetic Local Time, Latitude, and Longitude for High Latitudes in the Northern Hemisphere", Report No. AFCRL-70-0422, USAF Geophysics Laboratory.

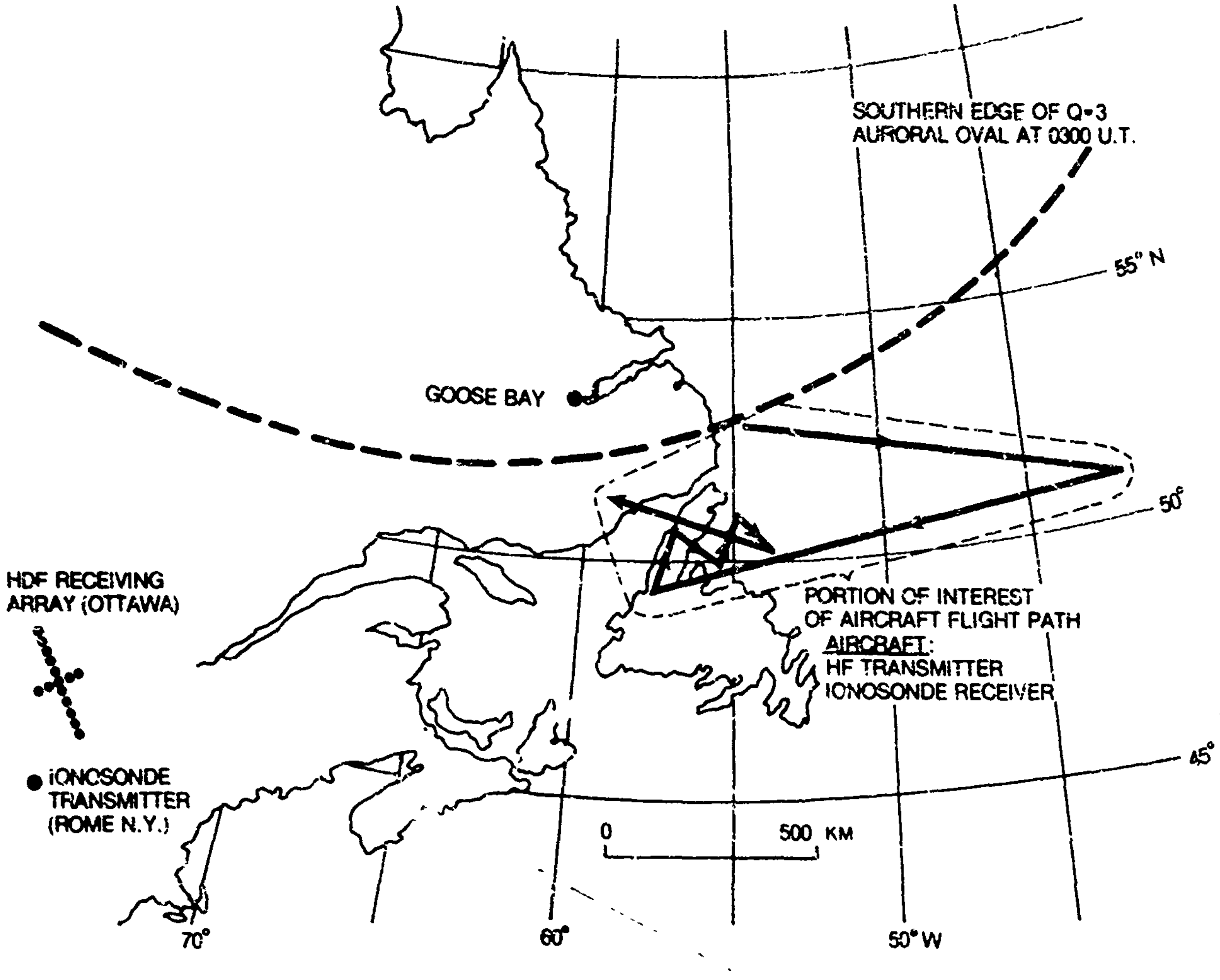

SOUTHERN EDGE OF Q=3
AURORAL OVAL AT 0300 U.T.
55° N
GOOSE BAY
50°
HDF RECEIVING
ARRAY (OTTAWA)
PORTION OF INTEREST
OF AIRCRAFT FLIGHT PATH
AIRCRAFT:
HF TRANSMITTER
IONOSONDE RECEIVER
45°
IONOSONDE
TRANSMITTER
(ROME N.Y.)
0
500 KM
70°
60°
50° W

Figure 1 Geometry of experiment

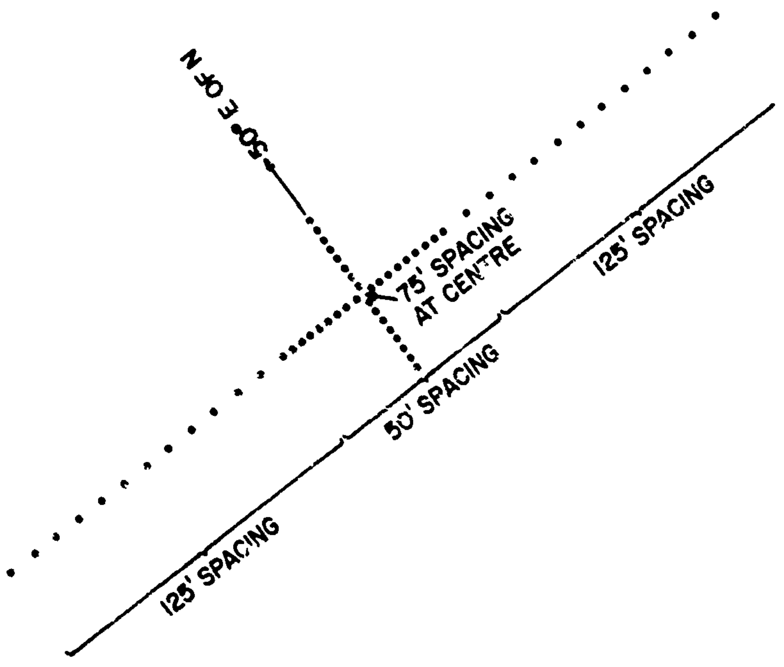

50° E OF N
75' SPACING
AT CENTRE
125' SPACING
50' SPACING
125' SPACING

Figure 2 HFDF receiving array near Ottawa. 8.2 m vertical elements used.

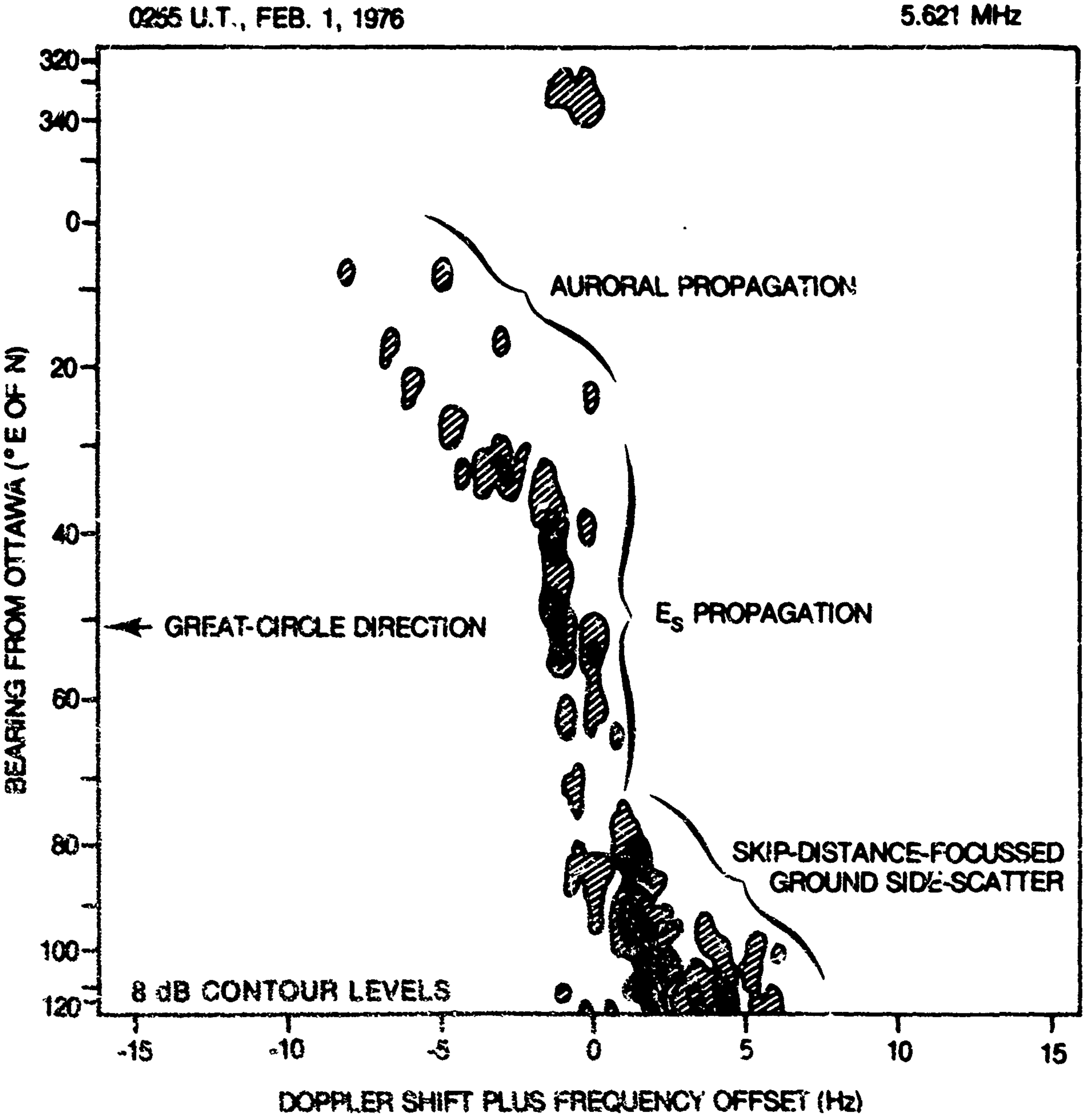


Figure 3 Contour plot taken from a line printer display of received signal strength vs. Doppler shift and bearing. 8 dB contour levels shown.

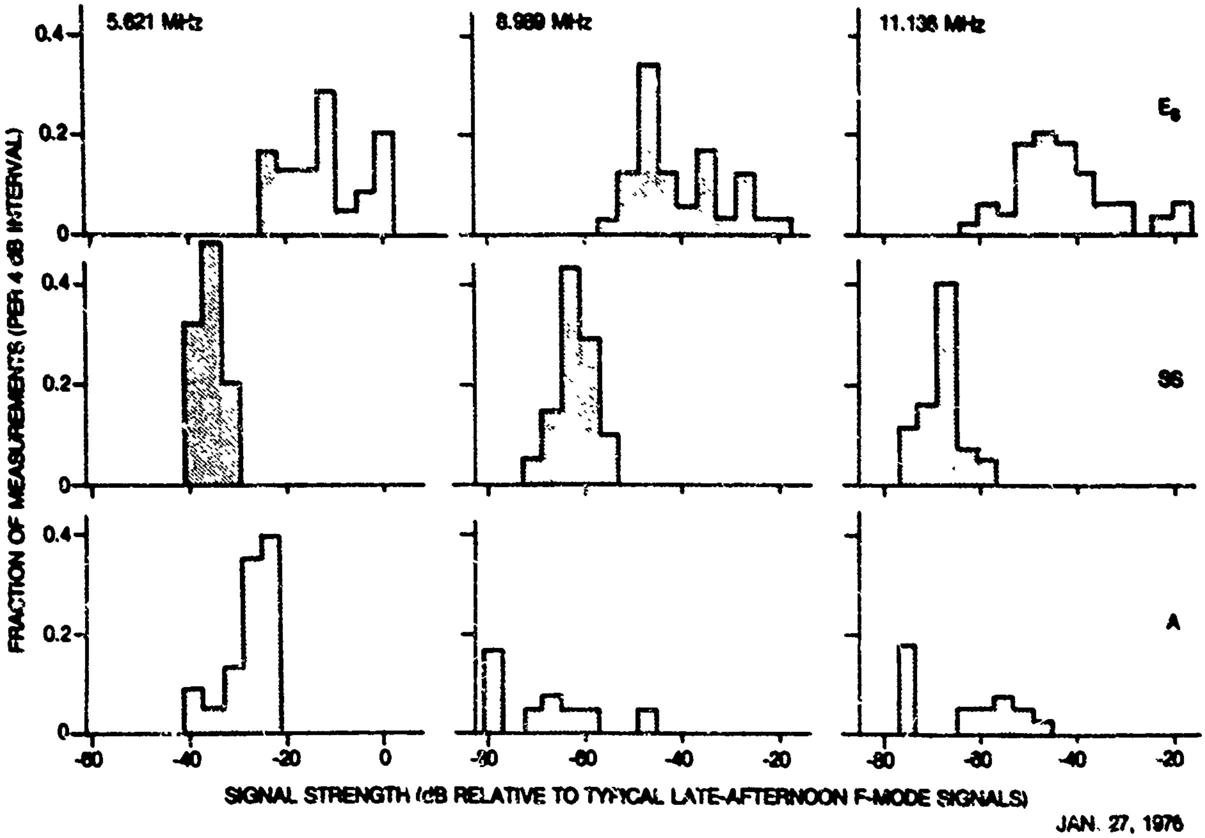


Figure 4A Distributions of received signal strengths for the E_s, SS and A modes at the three transmitted frequencies for 0050-0512UT, Jan. 27 (evening flight 1).

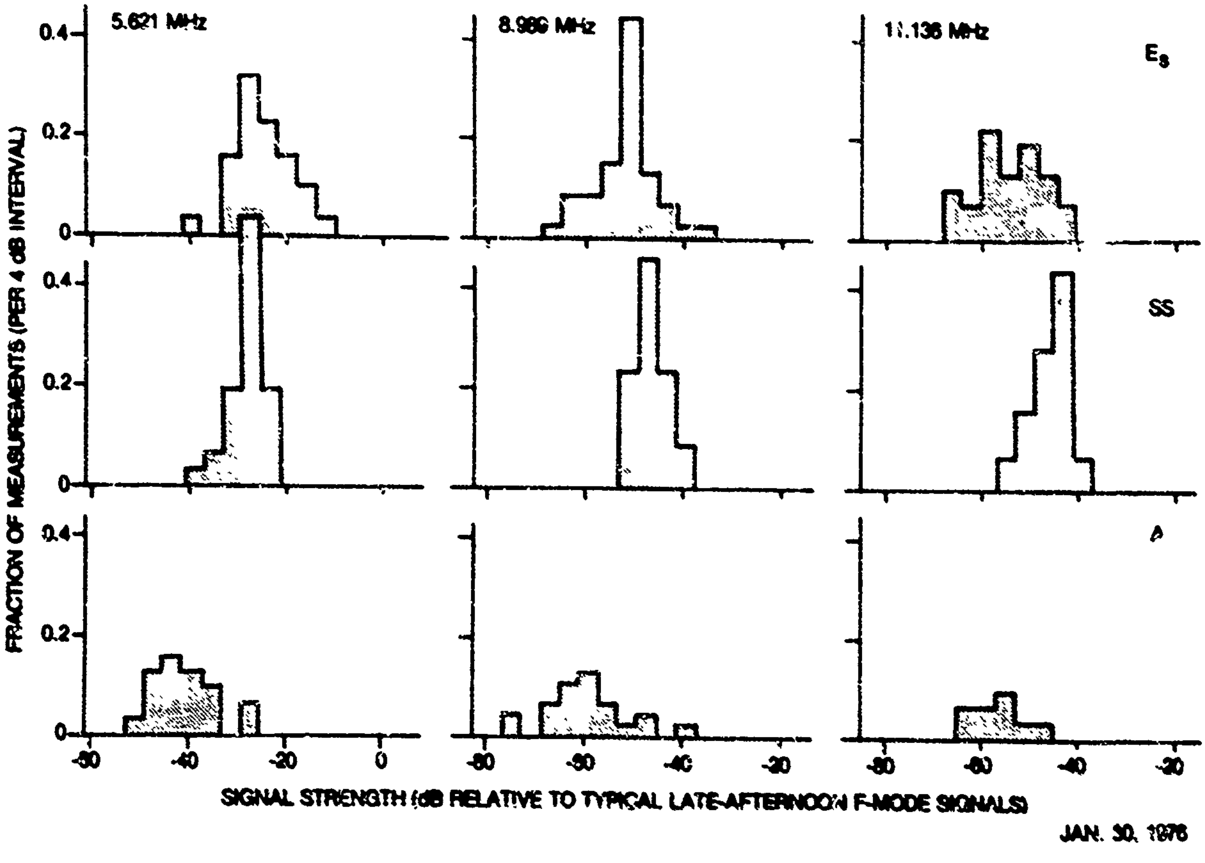


Figure 4B Distributions of received signal strengths for the E_s, SS and A modes at the three transmitted frequencies for 0135-0512UT, Jan. 30 (evening flight 2).

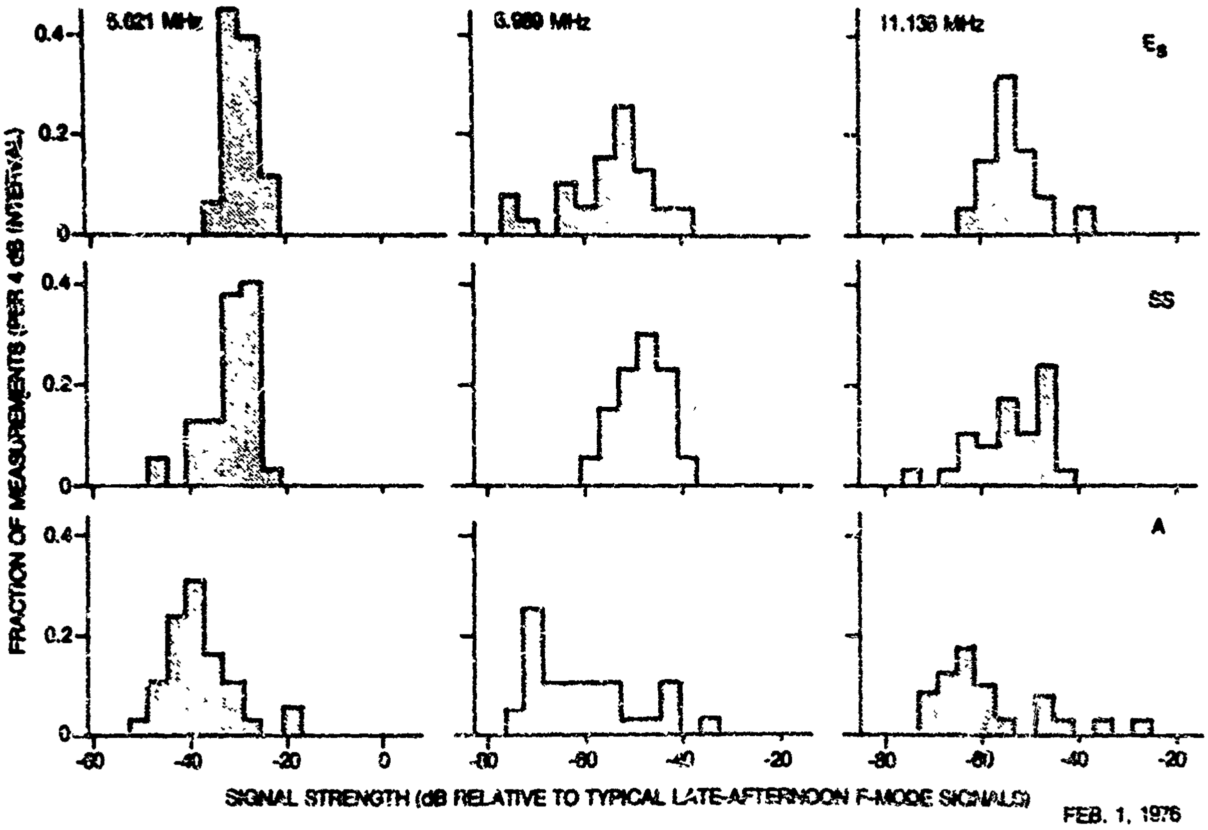


Figure 4C Distributions of received signal strengths for the E_s, SS and A modes at the three transmitted frequencies for 0050-0412UT, Feb. 1 (evening flight 3).

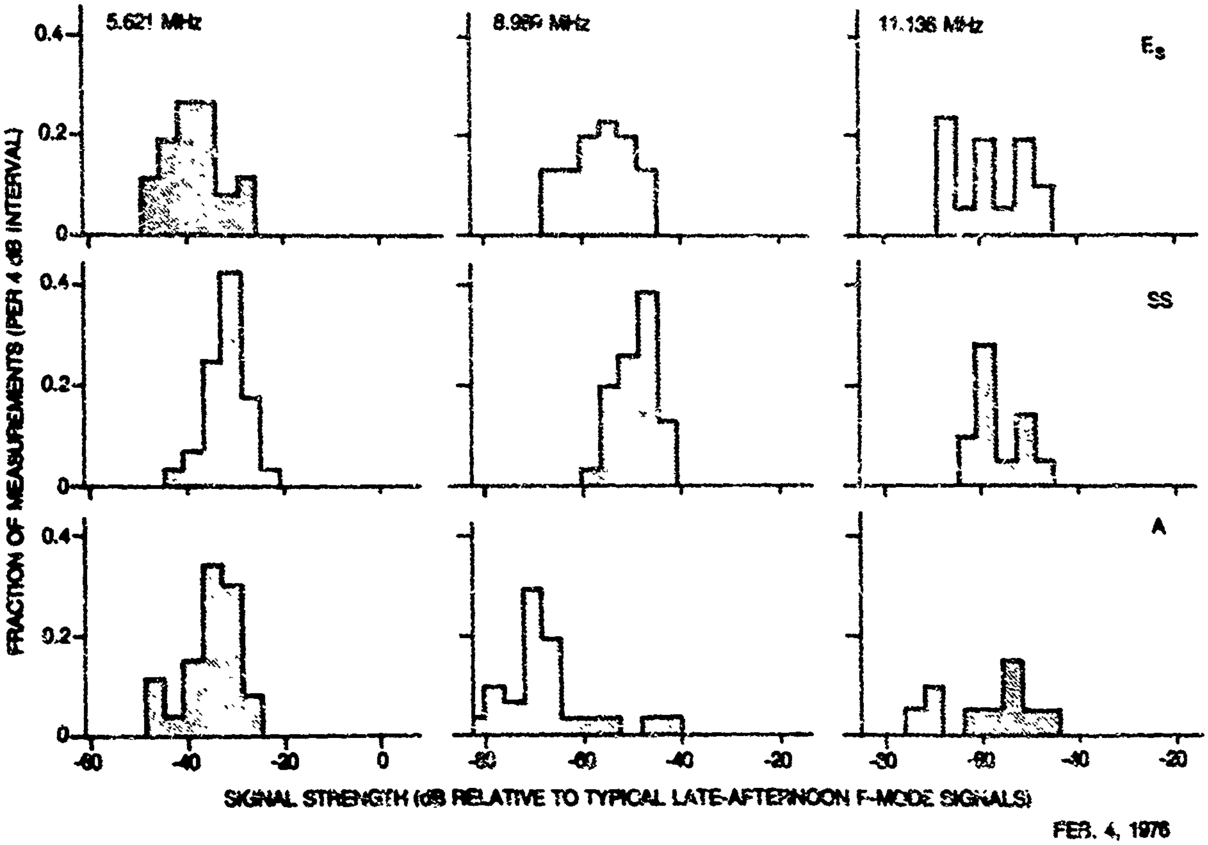


Figure 4D Distributions of received signal strengths for the E_s, SS and A modes at the three transmitted frequencies for 0050-0312UT, Feb. 4 (evening flight 4).

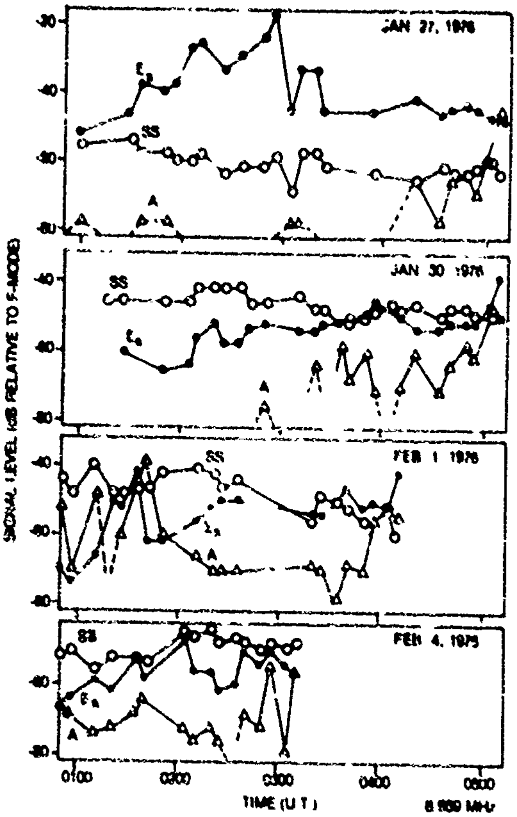


Figure 5 Signal strength as a function of time, for the E_s, SS, and A modes, for the four evening flights.

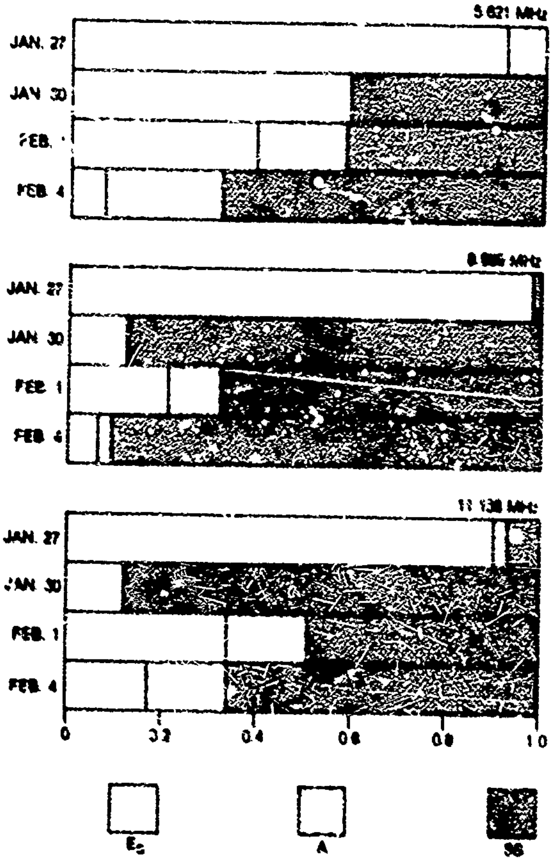


Figure 6 Fraction of measurements for which each mode dominated, for the three transmitted frequencies and four evening flights.

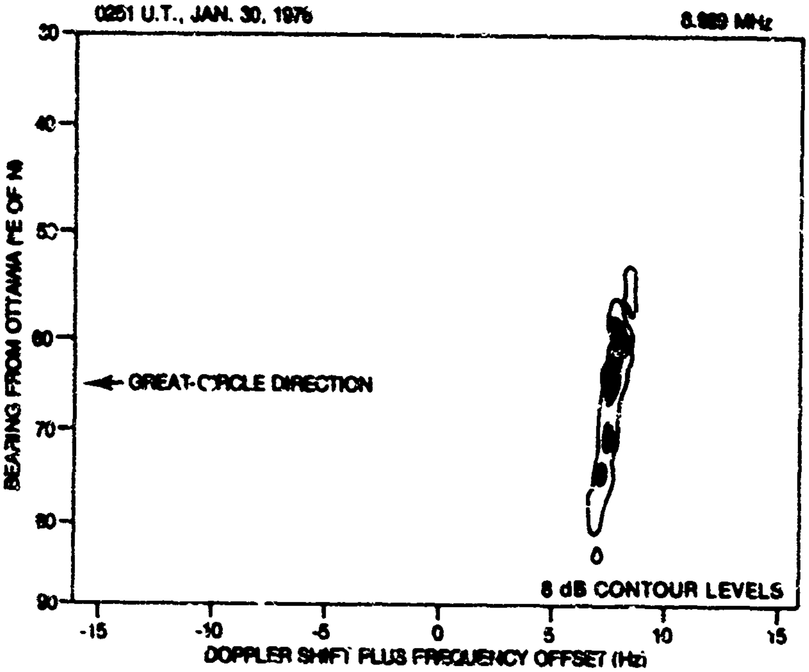


Figure 7 Contour plot taken from a line printer display of received signal strength as Doppler shift and bearing, for the E_s-propagated signals. 8 dB contour levels shown.

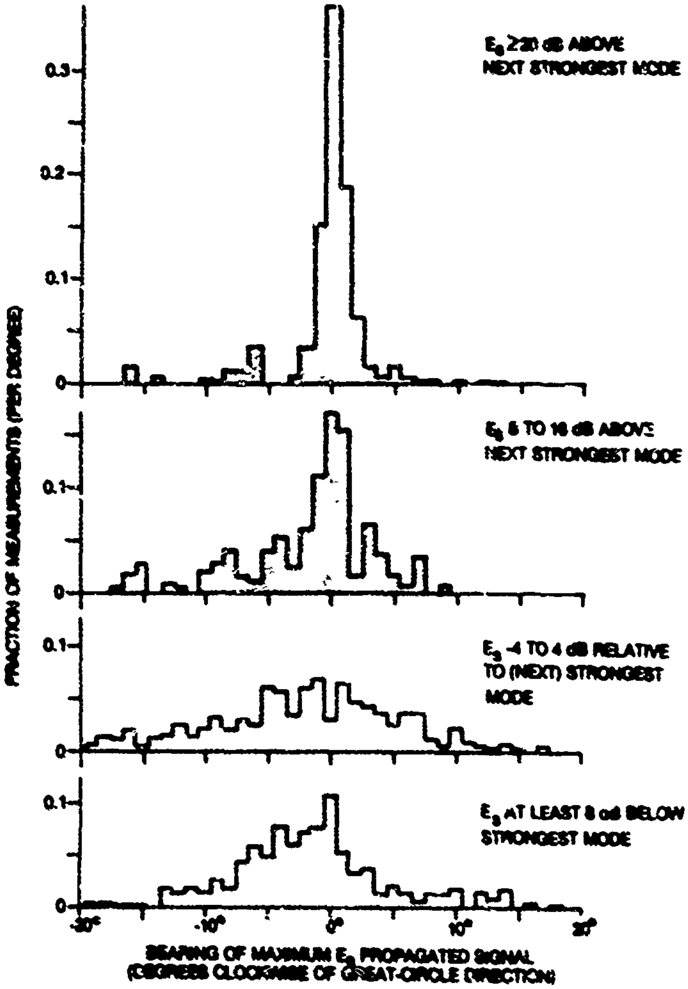


Figure 8 Distributions of maximum signal direction for the E_s mode arranged according to strength of the E_s mode relative to other modes present.

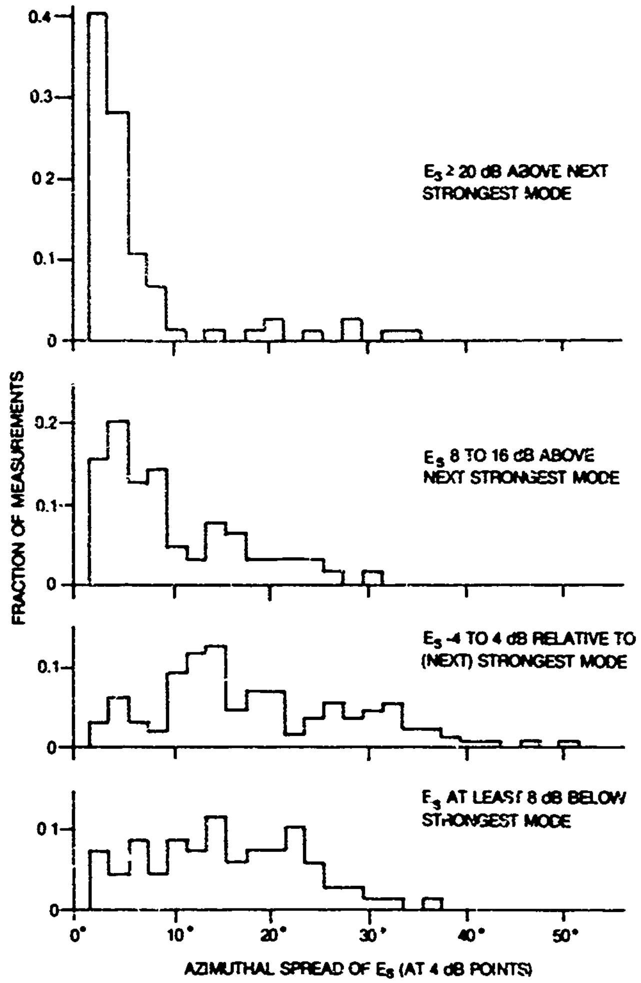


Figure 9 Distributions of spreads in bearing for the E_s mode, arranged according to strength of the E_s mode relative to the other modes present.

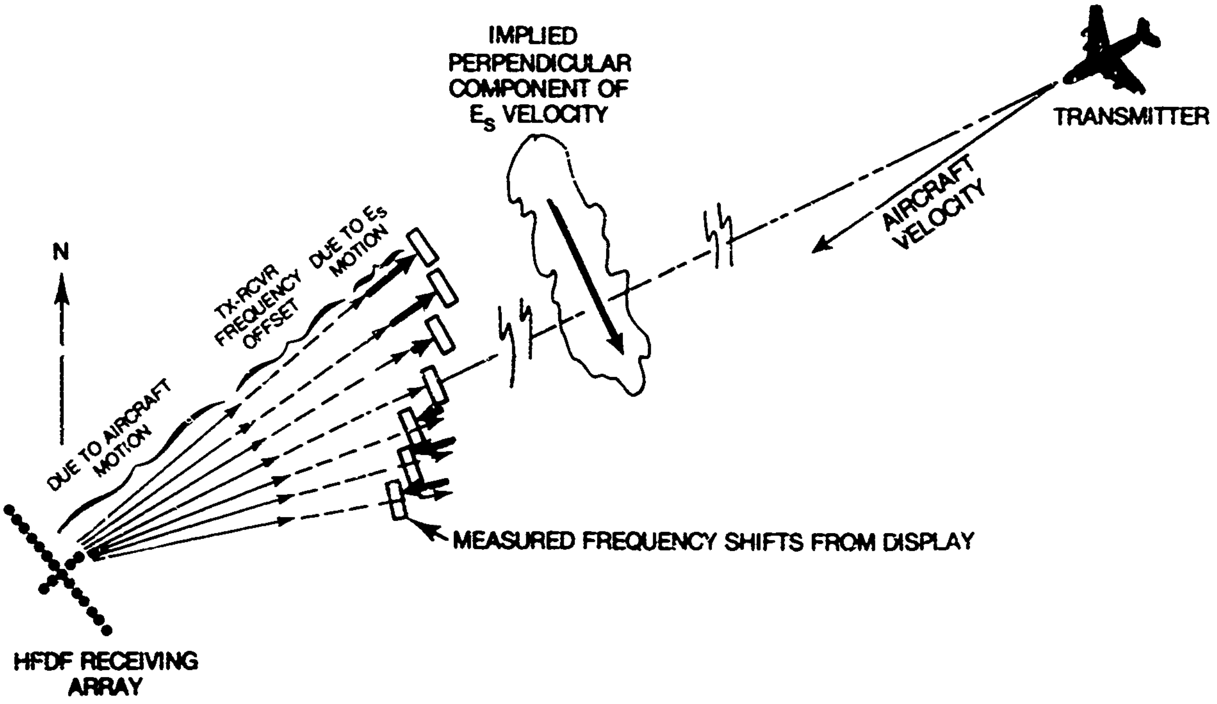

IMPLIED PERPENDICULAR COMPONENT OF Es VELOCITY
TRANSMITTER
AIRCRAFT VELOCITY
N
DUE TO Es MOTION
TX-RCVR FREQUENCY OFFSET
DUE TO AIRCRAFT MOTION
MEASURED FREQUENCY SHIFTS FROM DISPLAY
HFDF RECEIVING ARRAY

Figure 10 Determination of the Doppler shifts due to E_s motion, for the display of Figure 7.

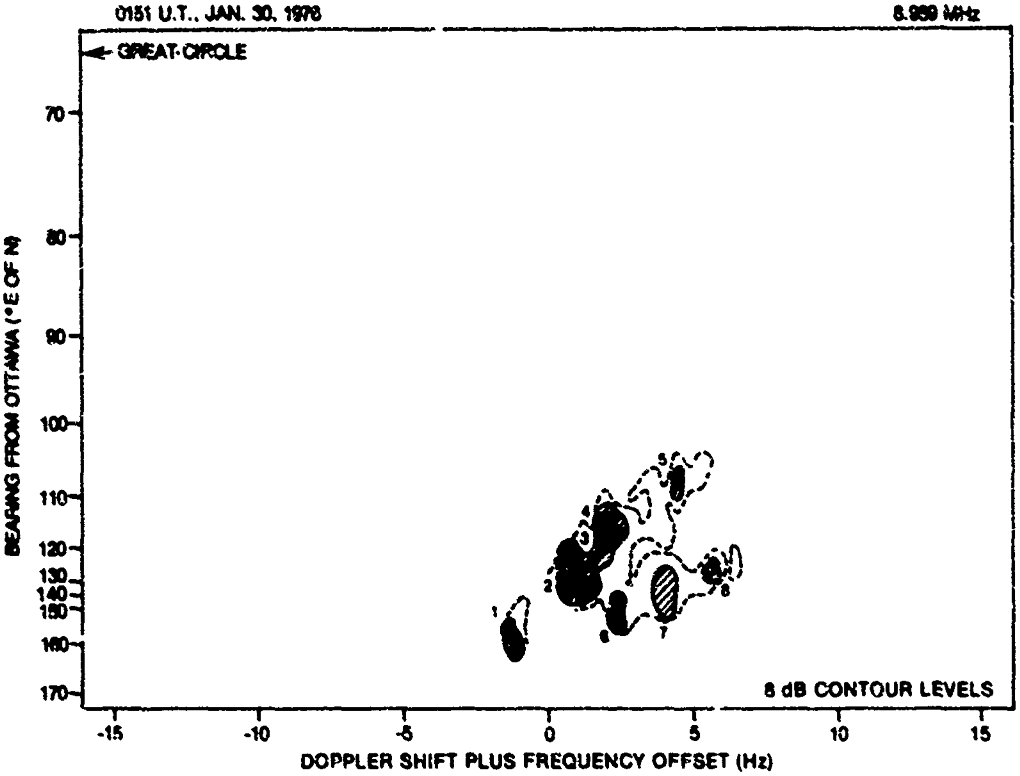


Figure 11 Contour plot taken from line printer displays of received signal strength vs. Doppler shift and bearing, for skip-distance focussed ground side-scatter. 8 dB contour levels shown.

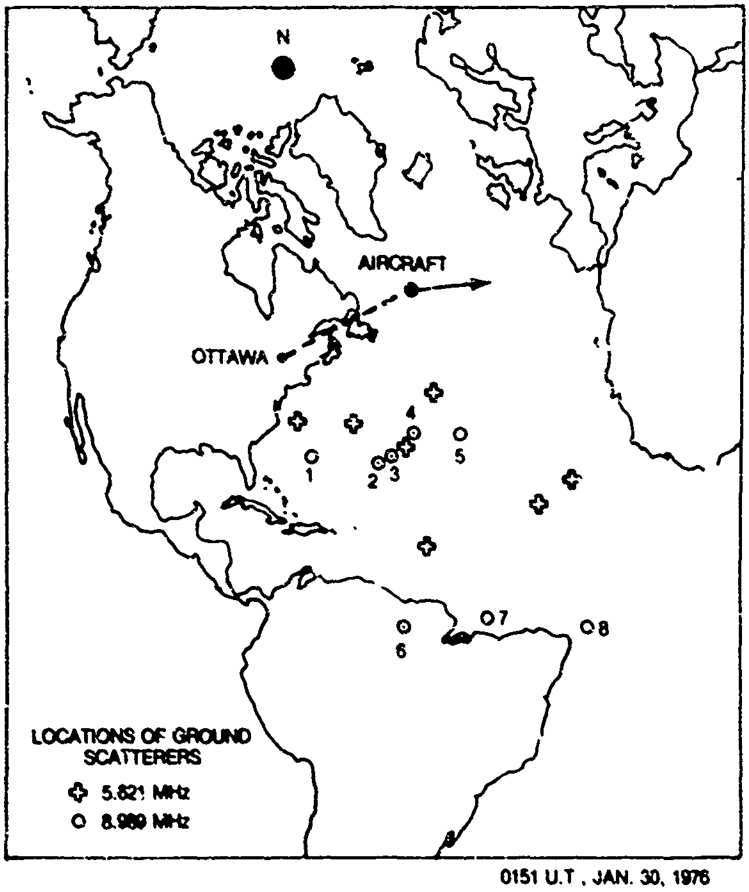

N
AIRCRAFT
OTTAWA
1
2
3
4
5
6
7
8
LOCATIONS OF GROUND SCATTERERS
✚ 5.821 MHz
○ 8.989 MHz
0151 U.T., JAN. 30, 1976

Figure 12 Approximate positions of ground scatters for the SS mode, as determined from Doppler shift and bearing measured at Ottawa.

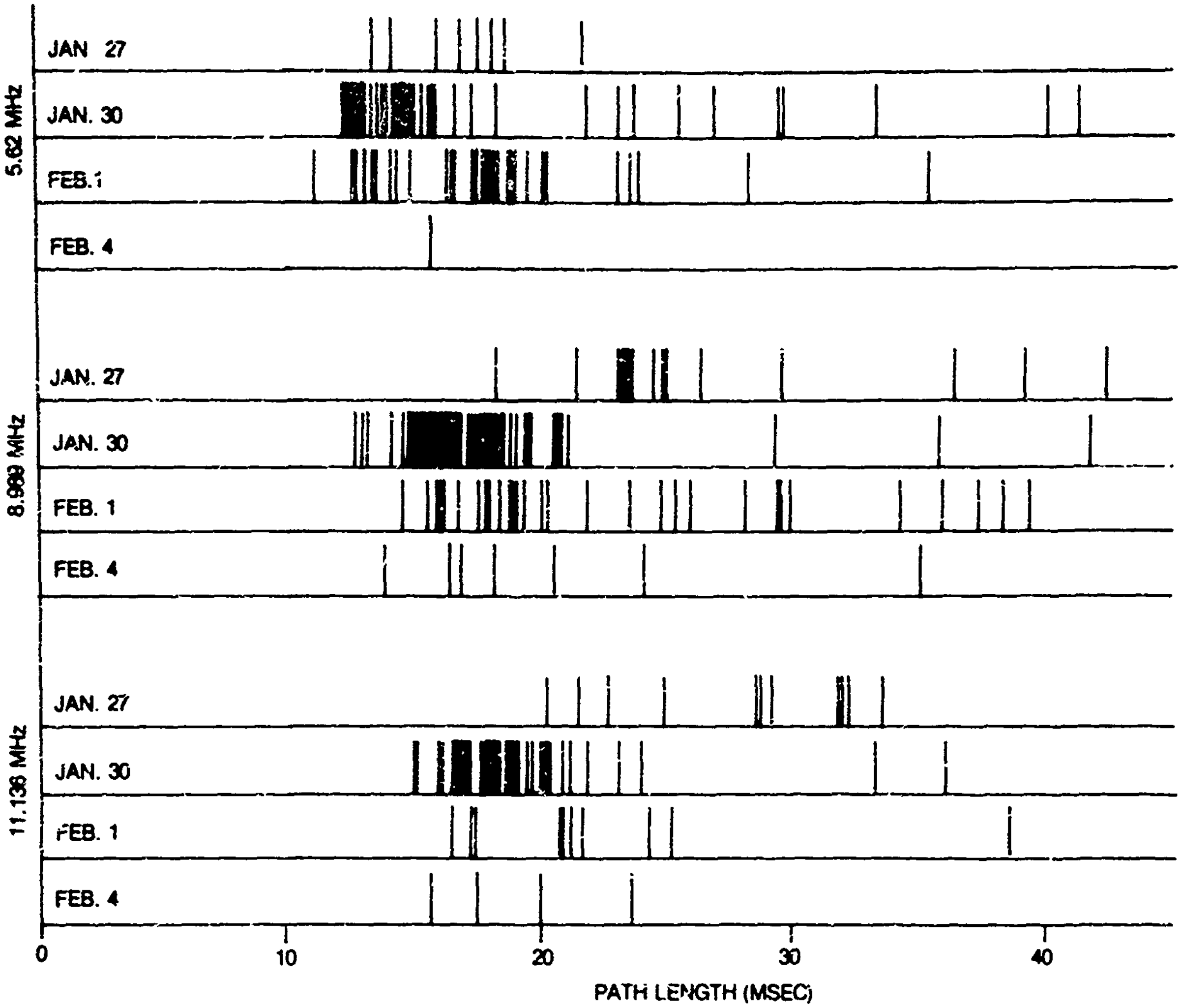


Figure 13 Distributions of path length for the SS mode.

DISCUSSION

James A. Whalen, US

In Figure 3: Do the differences in Doppler shifts – negative for auroral propagation, zero for Es, and positive for the southern scatter mode – occur when the aircraft was moving toward the north? If so, is it possible to detect and measure north-south (or in general perpendicular to line of sight) components of Doppler echoes from aircraft by OTH-B radar with such a technique?

Author's Reply

The position of the Doppler shift attributable to aircraft motion would be preferentially positive toward the north, and negative toward the south when the aircraft is moving northward. However, added to that must be a shift due to the motions of the ionospheric reflection or scattering point. Under certain conditions, e.g. reflection from the northern edge of a well-ordered trough, it may be possible to obtain the measurement you suggest.

RECENT ADVANCES IN HF PROPAGATION SIMULATION

Terence J. Elkins
Rome Air Development Center
Deputy for Electronic Technology
Hanscom AFB, MA 01731

1. Introduction

The mathematical modelling of High Frequency (HF) radio propagation in the ionosphere is now a well developed branch of radio science, having been in existence for over a half century. Early workers in this field were concerned with rigorous analytical formulations of the modelling problem of which the well-known Appleton-Hartree formula represents a major achievement in the theory of wave propagation. This formula, together with mathematical models for the ionospheric electron density distribution developed by Chapman (1931) enabled reasonably accurate calculations to be made of the distribution of HF radio energy in the earth-ionosphere cavity. The complexity of radio propagation in the ionosphere is such, however, that solutions to practical problems usually could be obtained only by making several simplifying assumptions or after extremely laborious calculation. As more information became available, in the 1940s and subsequently, concerning detailed ionospheric structure, requirements developed for improved computational techniques. These requirements were substantially answered by the advent of digital computers and by the development of a rigorous mathematical solution (Haselgrove, 1955) for the trajectories of radio waves in the ionosphere.

The major emphasis of HF radio propagation simulation for many years was in connection with long range, point-to-point communication circuits. The development of the "ITS-78" computer algorithm (Barghausen et al, 1969) represented a major advance in this area. While HF communications is still a major application of propagation simulation techniques, recent technological advances have generated requirements which are in some respects more demanding. Over-the-horizon (OTH) radars, operating in the HF band, are a prime example of a new technology which has placed very stringent demands on radio propagation simulation techniques. Satellite-to-ground propagation in the HF and low VHF bands is another area in which new simulation procedures have become necessary. The continual quest for greater accuracy in HF direction finding and surveillance systems has led to requirements for modelling relatively small scale ionospheric features, such as travelling ionospheric disturbances (TIDs). Modes of propagation which have hitherto been regarded as scientific curiosities may soon be exploited for practical purposes; thus techniques have been developed for simulating HF propagation in chordal and ducted modes, with appropriate consideration of the mode coupling problem (injection/ejection) using both natural and artificially generated ionospheric irregularities.

A major thrust of HF propagation simulation, in recent years has been the development of real time operational techniques for propagation management. This has entailed the development of high speed computational procedures, suitable for implementation in relatively small computers in some cases, and appropriate techniques for incorporating some form of "real-time" data into the ionospheric model in order to render it more closely representative of the actual ionosphere through which the desired signal is propagating. This paper will attempt to summarize these recent developments in HF propagation simulation and relate them to practical applications.

2. Ionospheric Models

The first prerequisite in the solution of a propagation simulation problem is the generation of an appropriate ionospheric model. Naturally, the complexity of the model is determined to some extent by the accuracy required in the final result. On the other hand, there is a reasonably well-established limit to this accuracy which is set primarily by the degree of realism with which the model represents the ionospheric structure--in other words, by the extent to which small departures from the modelled ionosphere, whether random or deterministic, influence the final desired result. Another factor which is important in selection of a mathematical model is the kind of mathematical ray tracing which is to be performed in the modelled ionosphere. In general, rapid, efficient ray tracing algorithms require analytic models, while the increased accuracy inherent in full three dimensional ray tracing can best be exploited with a numerical specification of the ionosphere.

2.1 Numerical Specification

The numerical model of the F-region developed by the Institute for Telecommunications Sciences (ITS) and implemented in the computer algorithm HFMUFES-4 has remained for two decades the dominant component of most ionospheric models. This model specifies the global distribution of F-region critical frequency (f_oF2), height of peak ionization (h_mF) and parabolic semi-thickness (YmF) - the last parameter is specified indirectly as the difference between h_mF and the height of the bottom of the layer. For accurate three dimensional ray tracing, it is usually necessary to substitute a more realistic vertical ionospheric profile for that contained in HFMUFES-4. An example of a suitable vertical profile is that of Rush et at (1973), in which an E-region model similar to that of HFMUFES-4 is merged smoothly and analytically to the ITS model for the F-layer peak.

Subsequent to the development of the ITS model, advances in theoretical and empirical knowledge of the polar regions, together with requirements for improved accuracy in polar ionospheric HF simulation, led to efforts to modify the ITS model in the polar regions. It was understood by this time that the polar ionosphere exhibits a substantial component which is ordered in geomagnetic rather than geographic coordinates. A system of coordinates known as the "Corrected Geomagnetic System (CGS)" (Hakura, 1965) was found to be useful in specifying a modifying term to the ITS model at high latitudes. Furthermore, it was also recognized that the level of geomagnetic activity plays a more complex role in controlling high latitude ionospheric structure than at temperate latitudes. A numerical specification has been developed at RADC (Miller and Gibbs, 1975) which combines the above factors to produce a global ionospheric model reflecting current understanding of the polar ionosphere. Among the major features of this model is a numerical specification of the ionospheric trough and the auroral oval, in the CGS. Figure 1 illustrates the model for f_oF_2 for the winter season at Sunspot Number 100, for a level of the Planetary Magnetic Index (K_p) of [illegible]

CGS. Figure 1 (a) is the result of transforming the ITS model for the appropriate conditions into the CGS and applying a "correction" factor, illustrated in Figure 1 (b). This factor was derived empirically by means of regression analysis of a large amount of vertical incidence ionosonde data against the ITS model, ordering the residuals in the CGS and extracting a K_p dependence.

Figure 2 illustrates the geometry of a particular auroral propagation path, showing the locations of the auroral oval and the solar terminator. Figure 3 shows the contours of plasma frequency in the vertical plane through the propagation path, predicted by the combined RADC and SRI models. Also shown are sample traced rays at various elevation angles from the transmitter, illustrating an example of a ducted ray at 10° elevation.

2.2 Quasi-Analytic Specification

The structure of the ionosphere on a global scale is too complex to permit a fully analytic specification. The term "analytic" here has a special meaning, related to the subsequent solution of the ray tracing problem. In practice it is not possible to represent the vertical profile by a suitable mathematical function while still maintaining acceptable accuracy. This is not possible in modelling horizontal variations on a global basis, although over a limited geographic extent, even the horizontal ionospheric gradients may be represented by simple analytic functions, such as linear or quadratic functions of range. An example of a model of this kind is one in which the vertical ionospheric profile is represented by a series of three quasi-parabolic layers (Croft and Hoogasian, 1968) connected by matching segments which preserve mathematical continuity of the electron density and its first derivative with altitude. The horizontal structure of the model is similar to that contained in the ITS specification. Such a model was recently developed (Miller and Gibbs, 1979), with the name FIMP (Fast Ionospheric Modelling Program), for quasi-real time system application. Figure 4 shows an example of the vertical profile generated by FIMP. The virtue of this particular model is that, when coupled to a suitably designed ray tracing procedure, the particular analytic form for the vertical profile permits extremely rapid radio path computation

3. Computation of Absorption

The computation of absorption has traditionally been treated in quasi-empirical manner in which the nondeviative D-region absorption is calculated as a function of solar zenith angle, together with empirically determined constants. Recent work by Bradley and George (1974) has refined this procedure by providing a global map of these empirical constants, incorporating their seasonal dependence and also containing a correction for deviative absorption in the E-region. This model is suitable for rapid and reasonably accurate computation.

Another approach to the calculation of absorption is the calculation of the attenuation term in the Appleton-Hartree equation for refractive index during numerical ray tracing. This is a straightforward approach involving a representation (either analytic or numerical) of the spatial distribution of electron collision frequency in the ionosphere, which amounts in practice mostly to the vertical (altitude) distribution. A recent development in this area was the incorporation of electron-ion collisions, in addition to electron-neutral collisions (Rush and Elkins, 1975). This enables an accurate estimate to be made of F-region deviative absorption which hitherto had normally been ignored in propagation simulation. This component of absorption is important, for example, in calculations of HF ionospheric ducting and in HF satellite to ground propagation. In general, the numerical ray-tracing approach to the calculation of absorption is useful and often essential in cases where atypical collision frequency profiles exist (auroral region, nuclear burst regions) or when the propagation path is not of the classical "hop" type.

3.1 Auroral Absorption

In the auroral regions there exists a component of D-region ionization, produced by energetic charged particle precipitation, which is best ordered in corrected geomagnetic coordinates. The absorption which this ionization induces on HF propagation paths often greatly exceeds the normal solar produced component. Auroral absorption is, however, sporadic in nature and is necessarily represented in statistical terms in propagation calculations. Recent published work on this subject, relevant to numerical computation of HF absorption, includes that of Foppiano (1975), Vondrak et al (1978) and Elkins (1972).

4 Numerical Ray Tracing

Recent developments in numerical ray tracing have been largely confined to the realm of computer software technology. The most rigorous approach to ray tracing is the solution of the so-called "Haselgrove equations" (Haselgrove, 1955), for which purpose a computer program was developed by Jones(1968). This program is in wide use in the HF radio propagation community, although it has, in some instances, been extensively modified to improve its flexibility and to tailor it to specific problems. At RADC, for example, the following modifications have been made to the original program of Jones:

1) Input of fully numerical ionosphere specified on a three dimensional lattice of arbitrary (including variable) spacing.
2) Location of computational pole (i.e., the symmetry axis of the spherical coordinate system) at the transmitter.
3) Incorporation of corrected geomagnetic coordinate system for polar ionosphere representation.
4) Optional sporadic-E reflection incorporated.
5) Incorporation of optional multiple order spherical harmonic expansions of the geomagnetic field.
6) Computation along the ray path of the magnetic aspect angle, i.e., the angle between the propagation vector and the local geomagnetic field.
7) Inclusion of generalized electron collision frequency profiles, containing electron-ion, as well as electron-neutral components.
8) Provision of a variety of output options, permitting plotting in various formats, calculation of minimum group path, Doppler shift and other derived parameters.
9) A homing feature is incorporated, in which a logical search procedure is implemented to enable the propagation path from a given transmitter to terminate at a specified point, either on the earth or in space.

5. HF Auroral Radar Clutter

A major factor in the performance of radars operating near the auroral regions is auroral scatter giving rise to unwanted ("clutter") returns. From another point of view, it has been proposed (Bates, 1971) that a single HF radar located near the geomagnetic pole could be used to monitor continuously the location and level of activity of the auroral oval, by means of auroral backscatter. In both of these applications, an appropriate propagation simulation procedure is essential--in the former case to permit optimization of radar performance in an auroral environment, and in the latter to permit interpretation of the data.

Auroral scatter is caused by small scale ionospheric irregularities generated by a variety of plasma instabilities. Because of the strongly anisotropic nature of ionospheric conductivity, these plasma irregularities tend to be elongated in the direction of the local geomagnetic field, with the result that their radar cross section is usually highly directional, with a maximum value in a direction perpendicular to the magnetic field. A procedure for HF auroral clutter computation involves four principal elements:

1) A model for the distribution in the ionosphere of small scale irregularities.
2) A model for the geomagnetic field structure.
3) A model for the ambient ionosphere.
4) A ray tracing procedure in which the geomagnetic aspect angle (see section 4) is computed during refraction of the ray.

Combination of these four elements permits calculation of the regions from which auroral clutter will be observed, namely those regions containing small scale irregularities in which the magnetic aspect angle, after partial refraction, achieves the value 90°. In addition, if the backscatter cross section of the auroral irregularities is known, as a function of geomagnetic aspect angle, the magnitude of the clutter return can be computed by a simple extension of this procedure.

5.1 Clutter Amplitude

Figures 5 and 6 show models for the volumetric backscatter cross section of E- and F-region ionospheric irregularities respectively (Elkins, 1979). (The volumetric cross section is the equivalent point target cross section per unit illuminated volume.) The cross section is depicted in the CGS, for the winter season at a moderate level of geomagnetic activity ($K_p = 2$). Due to the strong anisotropy of auroral scatterers, in which they appear greatly elongated along the local geomagnetic field direction, auroral clutter returns are experienced by a radar only from those regions in which the propagation vector, after refraction, achieves near-orthongonality with the geomagnetic field. The computation of range-azimuth dependence of received auroral clutter power, for a particular radar, thus involves two steps:

a) Determination of the regions of "orthogonality" by means of ionospheric modelling and ray tracing;
b) Solution of the standard radar equation, using the radar parameters together with the auroral clutter cross sections as modelled and illustrated in Figures 5 and 6.

The computation of equivalent point target cross section from volumetric cross section involves estimation of the volume of auroral irregularities illuminated by the radar. In general, this is a far more difficult task for a HF radar than for a line-of-sight (UHF) radar, due to the large ionospheric refraction, and involves an exhaustive ray tracing analysis. Figure 7, however, illustrates an approximate method of computing illuminated volume, in terms of an effective vertical illumination angle. This angle ($\Delta\varepsilon$) is determined, at a given range and for a given radar pulse length, by the requirement for geomagnetic orthogonality, as illustrated.

The calculation of F-region auroral clutter power P_c experienced by a hypothetical experimental HF radar at 60° geomagnetic latitude proceeds as follows:

$$P_c = \frac{P_T\, G_T\, D_R\, \lambda^2\, \sigma_V\, V_S}{(4\pi)^3\, R^4\, L^2}$$

where P_T = transmitter average power (1 kw)
G_T = transmit antenna gain (15 dB)
D_R = receive antenna directivity (15 dB)
λ = wavelength (50 meters)
σ_V = volume backscatter cross section
V_S = illuminated volume
= $\frac{1}{2}R^2\ \Delta\varepsilon \cdot \phi \cdot c\tau$
R = range
ϕ = azimuthal beamwidth (10 deg)
τ = radar pulse length (100 μsec)
c = velocity of light
L = one way absorption loss

The antenna radiation patterns are assumed to be broad in the vertical plane (short monopoles). Figure 8 shows the regions of geomagnetic orthogonality in a 120° sector to the North of the radar, for 2 hop propagation, determined by modelling and ray tracing at 6 MHz frequency. Figure 9 shows contours of constant P_c calculated from the radar equation. The value of $\Delta\varepsilon$ was determined by ray tracing, in this case, to be ~0.5°.

5.2 Doppler Shift

The motion of auroral scatterers imparts a Doppler Shift to the backscattered radar signal. In the F-region this Doppler shift is principally due to the plasma motion driven by magnetospheric electric fields. Figure 10 shows, in qualitative fashion, the F-region plasma convection pattern over the polar region. Shown dashed is the equatorial boundary of the statistical auroral oval. Elkins (1979) has developed a quantitative model for the plasma velocities, based on theoretical consideration together with data from satellite and auroral radar experiments. Using this model, the Doppler shift experienced by the hypothetical radar at local midnight, at a frequency of 10 MHz, has been computed and is illustrated in Figure 11. A positive Doppler shift (by convention) indicates a receding velocity. The equatorial edge of the auroral oval is again shown as a dashed line.

6. Travelling Ionospheric Disturbance

A particular small-to-medium scale ionospheric feature which is very common, and which can be modelled for simulation purposes is the "travelling ionospheric disturbance" (TID) Francis (1974) has developed a technique for accurately modelling the effects of TIDs on radio systems. Figure 12 shows a vertical cross section through a horizontally stratified Chapman layer with a superimposed TID. The dashed lines are loci of constant phase of the internal gravity wave in the neutral atmosphere, which provides the driving force for the TID. Francis' technique has been modified at RADC to permit the TID perturbation to be applied to arbitrary ionospheric models, in particular, to numerically specified models.

In order to illustrate the practical application of TID modelling an experiment has been simulated in which an OTH radar is required to operate in such a way that a distant transponder is maintained in the skip focussing region while a TID crosses the propagation path orthogonally. Figure 13 shows the geometry of the experiment, together with contours of f_oF_2 modelled for the ambient ionosphere by the ITS technique. A three dimensional ionospheric model was constructed in the region of interest using the RADC technique (Rush et al, 1974). This 3-dimensional ambient model was then perturbed according to the formulation of Francis, by a TID having a period of 40 minutes, horizontal phase velocity of 200 meter/sec, and vertical velocity of 50 meter/sec. The horizontal wavelength of this TID was thus 480 km.

The leading edges of backscatter ionograms were synthesized in the direction of the transponder by means of 3-dimensional ray tracing in the perturbed model. Figure 14 shows examples of backscatter leading edges for the ambient ionosphere and for two values of TID phase, separated by a half period (20 minutes). The points labelled along these curves represent take-off elevation angles at the radar. The frequency required to maintain the transponder at the first hop skip range was determined by trial and error using 3-dimensional ray tracing with a homing feature, using the backscatter leading edges as a first guess. Figure 15 shows this frequency as a function of time (TID phase). Figure 16 and 17 show the corresponding variations of take-off elevation and azimuth at the radar.

7. Ducted Ionospheric Propagation

HF energy from ground-based sources can be trapped in ionospheric ducts and propagate to large ranges, provide suitable conditions exist in the ionosphere in the vicinity of the transmitter. Figure 3 shows an example of a ray injected into a duct as a result of natural ionospheric gradients near the transmitter. Energy induced plasma irregularities generated by a high-power ground-based "heater". Radio waves are scattered from the magnetic field aligned irregularities with the result that the heated region acts like a "radio mirror" or virtual source, which can be placed in the ionospheric duct by suitable choice of heater frequency. Figure 18 shows a vertical ionospheric profile containing a "valley" between E and F layers. The critical frequency of the F-layer varies with range from an elevated virtual source as indicated in Figure 19, while the E-layer and valley plasma frequencies remain constant. Figure 20 shows the results of ray tracing in this ionospheric model with the source at 160 km altitude (i.e., in the valley). Rays are directed horizontally at 2 MHz intervals in the frequency range 12 MHz - 34 MHz. All rays remain trapped in the duct out to ranges in excess of 9000 km, except the 34 MHz ray, which escapes near the f_oF minimum.

8. Summary

Advances in ionospheric modelling and numerical ray tracing, over the past few years have been mainly directed towards the solution of specific applied problems in HF propagation. Ionospheric models have achieved greater versatility in computational terms and earlier models have been modified to better represent polar ionospheric features. The calculation of radio absorption has been improved and deviative absorption can now be estimated accurately. Numerical ray tracing techniques developed in the 1960s have been given a similarly expanded versatility, permitting their application to a wide range of problems. Examples have been presented illustrating some recent applications of current HF propagation simulation techniques. These included the estimation of HF auroral clutter impact on an experimental OTH radar, the effect of TIDs on backscatter ionograms and the computation of ionospheric ducting, using a "virtual source" in the ionospheric duct.

References

Bates, H.F. (1971), The Aspect Sensitivity of Spread-F Irregularities, J. Atmos. Terr. Phys., 33, 111.

Barghausen, A.F., J.W. Finney, L.L. Proctor, L.D. Schultz (1969), Predicting Long-Term Operational Parameters of High-Frequency Sky-Wave Telecommunication Systems, ESSA Tech. Rept., ERL-110-ITS-78.

Bradley, P.L. and P. George (1974), A New Method of Predicting the Ionospheric Absorption of High Frequency Waves at Oblique Incidence, Telecomm. J., 41, 307.

Chapman, S. (1931), The Absorption and Dissociative or Ionizing Effect of Monochromatic Radiation in an Atmosphere on a Rotating Earth, Proc. Phys. Soc., 43, 26.

Croft, T.A. and H. Hoogasian (1968), Exact Ray Calculations in a Quasi Parabolic Ionosphere with No Magnetic Field, Radio Science, 3, 69.

Elkins, T.J. (1972), A Model of Auroral Substorm Absorption, AFCRL-TR-72-0413.

Elkins, T.J. (1979), In Preparation.

Francis, S.H. (1973), Acoustic-Gravity Modes and Large-Scale Traveling Ionospheric Disturbances of a Realistic, Dissipative Atmosphere, J. Geophys. Res., 78, 2278.

Foppiano, A. (1975), CCIR, IWP 6/1, Docs. 3 and 10.

Jones, R.M. (1968), A Three Dimensional Ray Tracing Computer Program, Radio Science, 3, 93.

Hakura, Y. (1965), Tables and Maps of Geomagnetic Co-ordinates Corrected by the Higher Order Spherical Harmonic Terms, Repts. Iono. Space Res. (Japan), 19, 121.

Haselgrove, J. (1955), Ray Theory and a New Method for Ray Tracing, Proc. Cambridge Conf. on Physics of the Ionosphere, Phys. Soc. (London), 355.

Miller, D. and J. Gibbs (1975), Ionospheric Analysis and Ionospheric Modelling, AFCRL-TR-75, 0549.

Miller, D. and J. Gibbs (1979), RADC Tech. Rept. (In Press).

Rush, C.M (1976), An Observation Network for use in Short-Term Propagation Predictions. Telecomm. J., 43, VIII, 544.

Rush, C.M. and T.J. Elkins (1975), An Assessment of the Magnitude of the F-region Absorption on HF Radio Waves using Realistic Electron Density and Collision Frequency Models, Telecomm. J., 42, 476.

Rush, C.M., D. Miller and J. Gibbs (1974), The Relative Daily Variability of f_oF_2 and h_mF_2 and Their Implications for HF Radio Propagation, Radio Science, 9, 749.

Vondrak, R.R., G. Smith, V.E. Hatfield, R.T. Tsunoda, V.R. Frank and P.D. Perreault (1978), Chatanika Model of the High Latitude Ionosphere for Application to HF Propagation Prediction, RADC-TR-78-7.

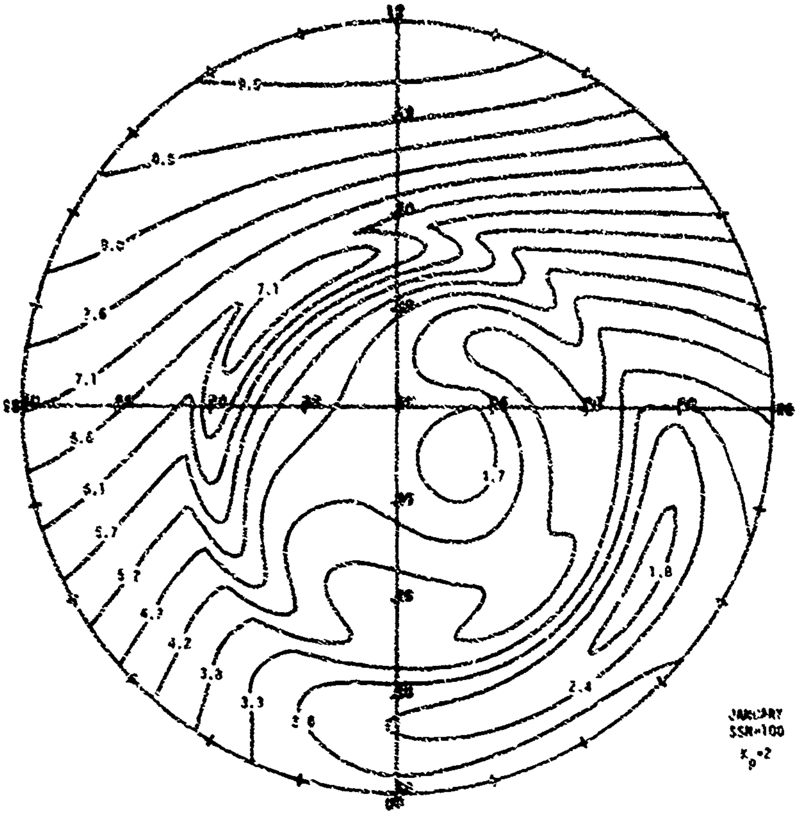


FIGURE 1(a): Model for F_oF_2 over the polar region for January at Sunspot Number 100, $K_p = 2$, in corrected Geomagnetic Co-ordinates.

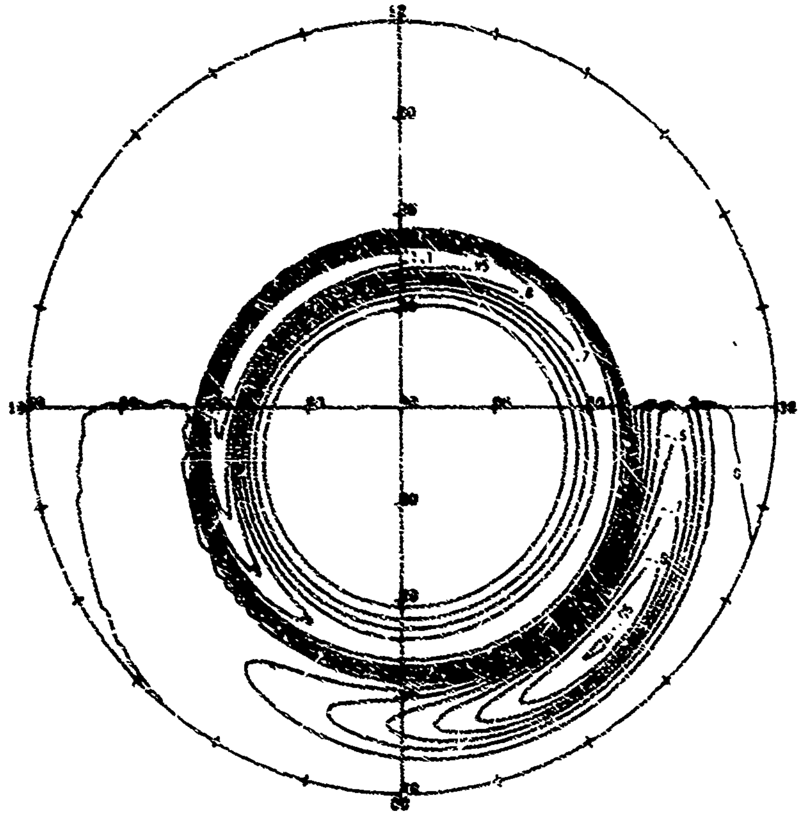


FIGURE 1(b): Empirical correction term applied to ITS-78 ionospheric model to produce the polar model of f_oF_2 in Figure 1(a).

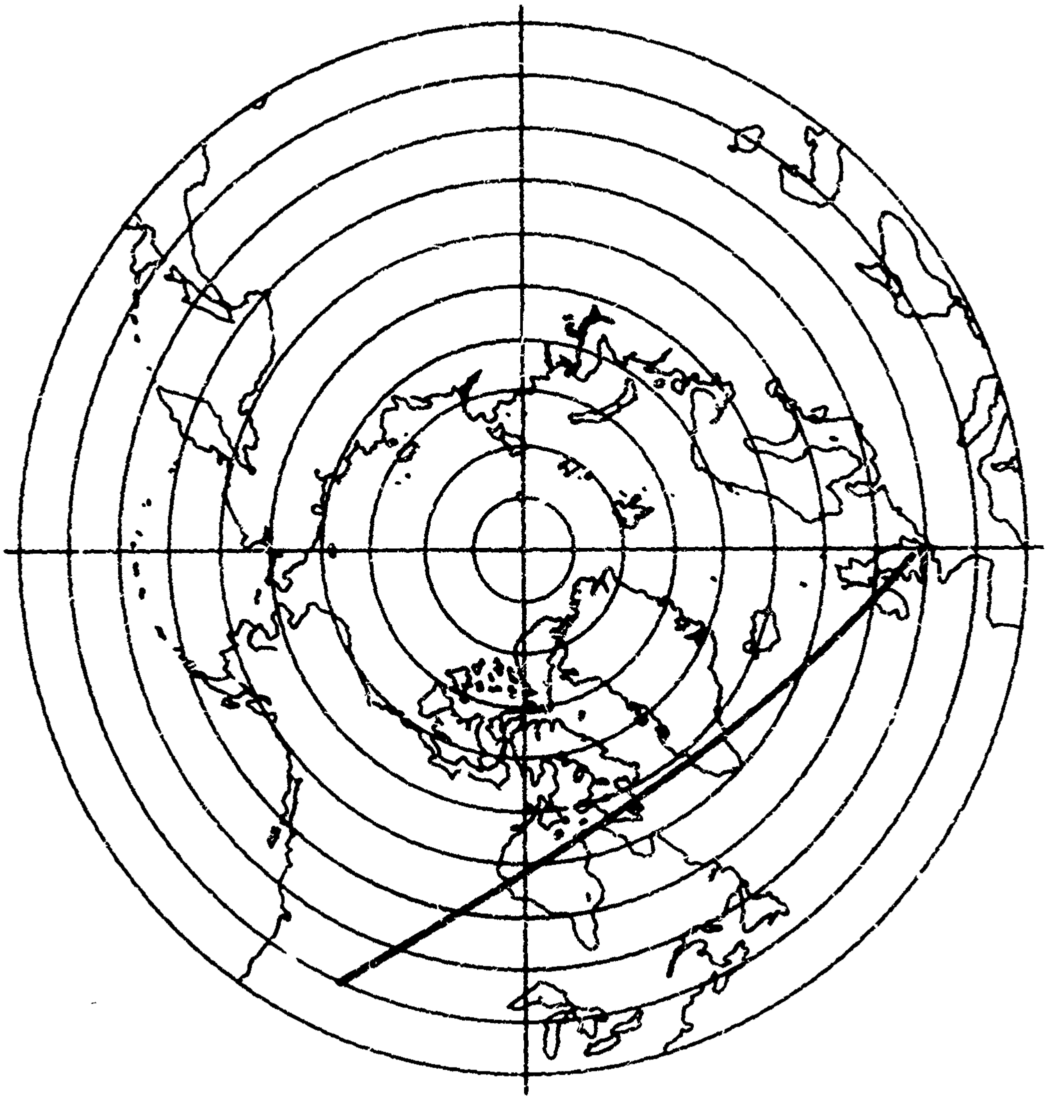


FIGURE 2: Polar view of Northern hemisphere showing selected propagation path.

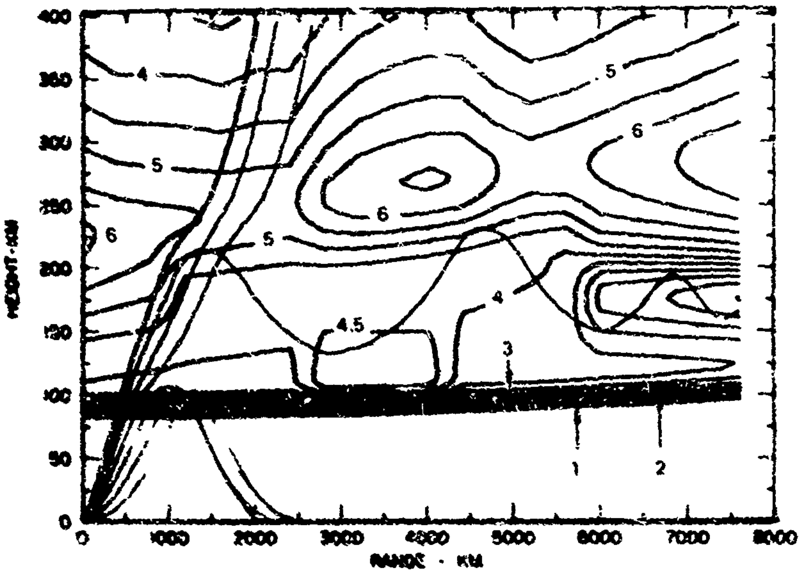


FIGURE 3: Ionospheric contours of local plasma frequency (MHz) generated by the RADC model for propagation path in Figure 2 (UT = 18; SSN = 50; K_p = 2). Ray trajectories are indicated (after Vondrak et al, 1978).

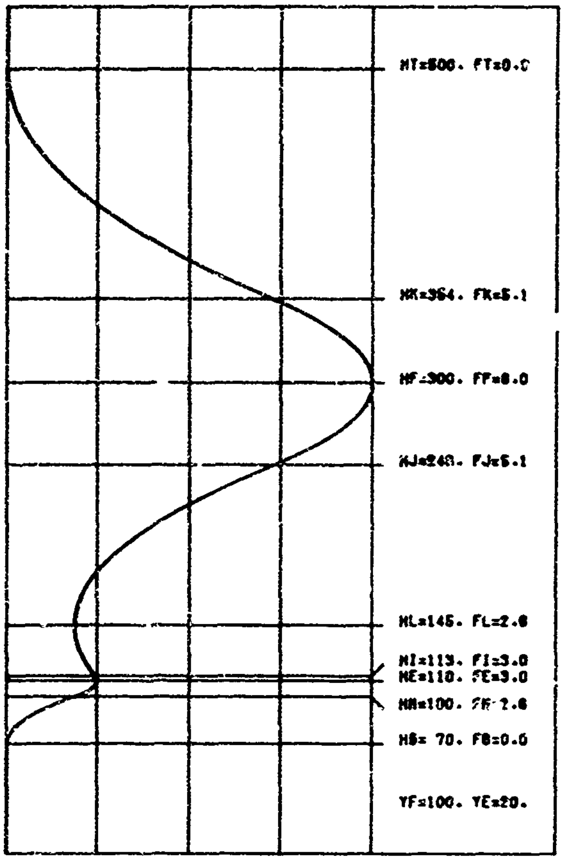


FIGURE 4: Example of a vertical profile generated by program FIMP, consisting of quasi-parabolic layer segments linked by analytic matching segments. The quantities indicated are the heights and plasma frequencies at key points on the profile (H_i and F_i) and layer Semi-thickness (Y_F, Y_E).

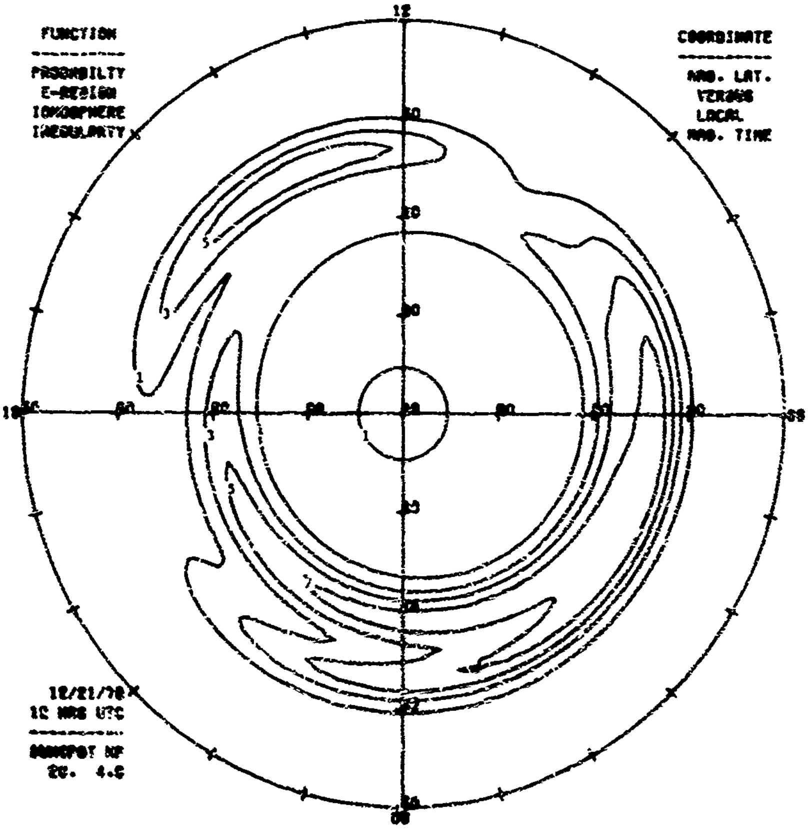


FIGURE 5: Median volumetric backscatter cross section of E-region auroral irregularities in CGS. Key to contours (at 10 MHz frequency): 1 = -102; 3 = -94; 5 = -89; 7 = -84. Units are dB (M^2/M^3).

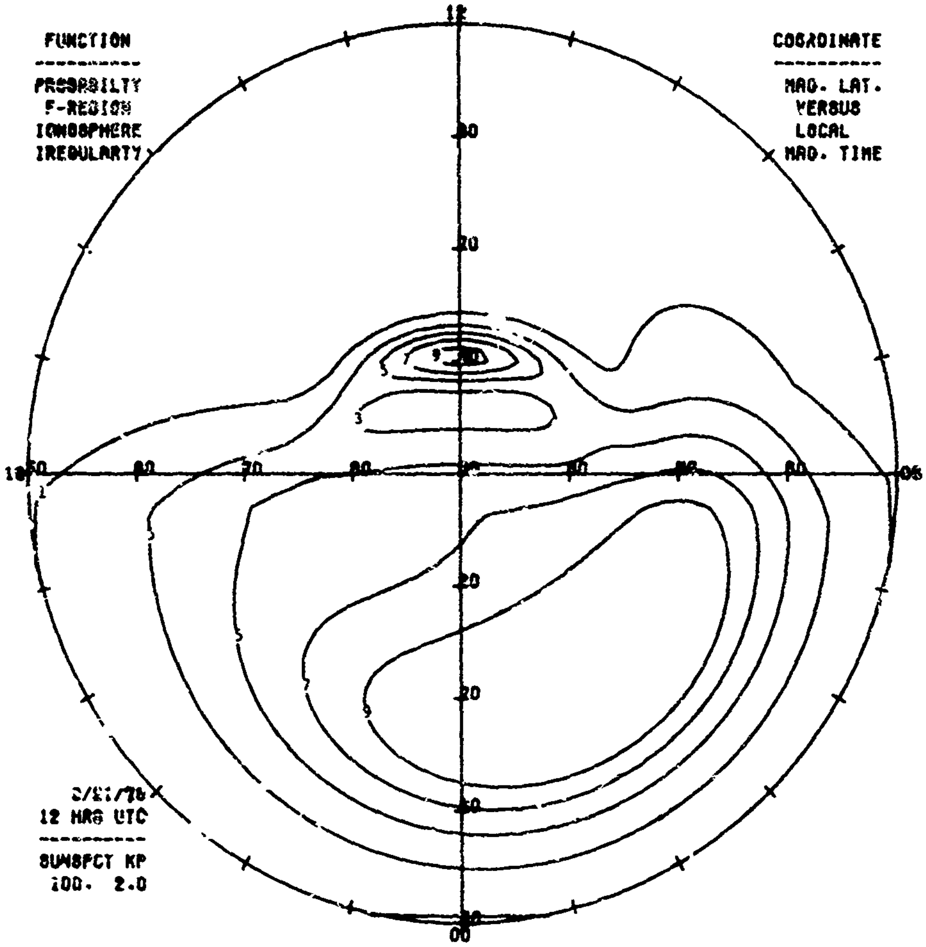


FIGURE 6: Median volumetric backscatter cross section of F-region auroral irregularities in CGS. Key to contours (at 10 MHz frequency): 1 = -95; 3 = -85; 5 = -77; 7 = -71; 9 = -61. Units are dB (m^2/M^3).

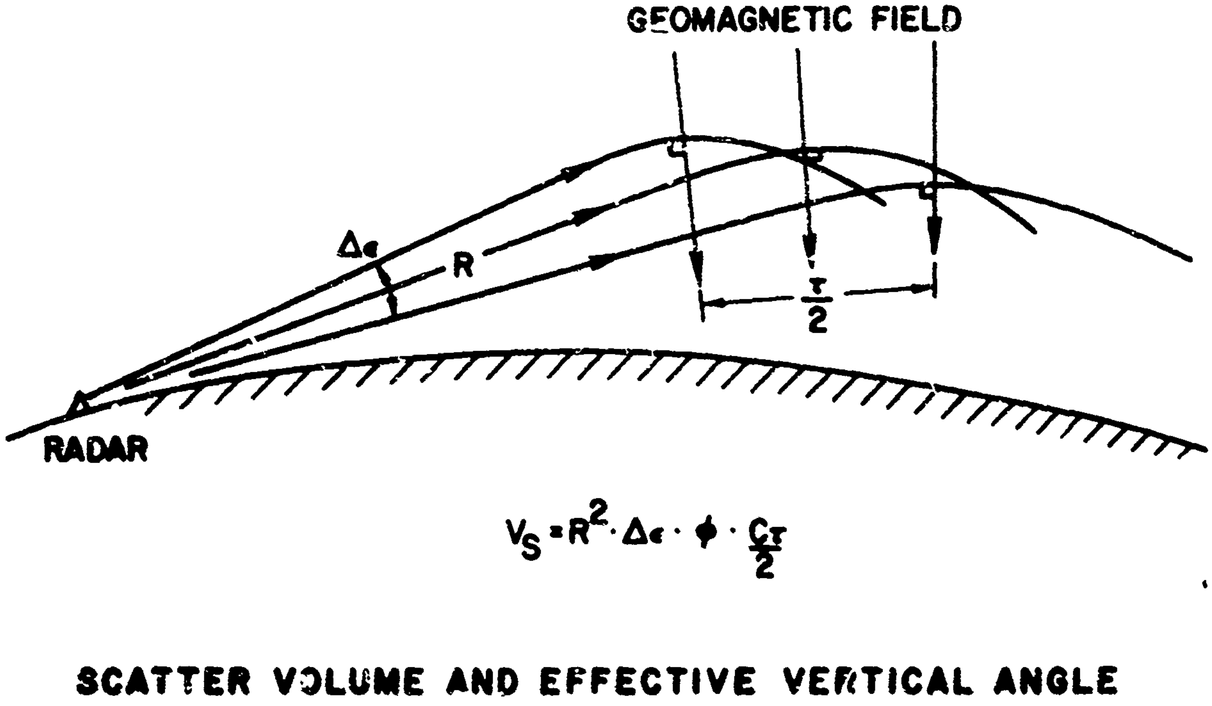
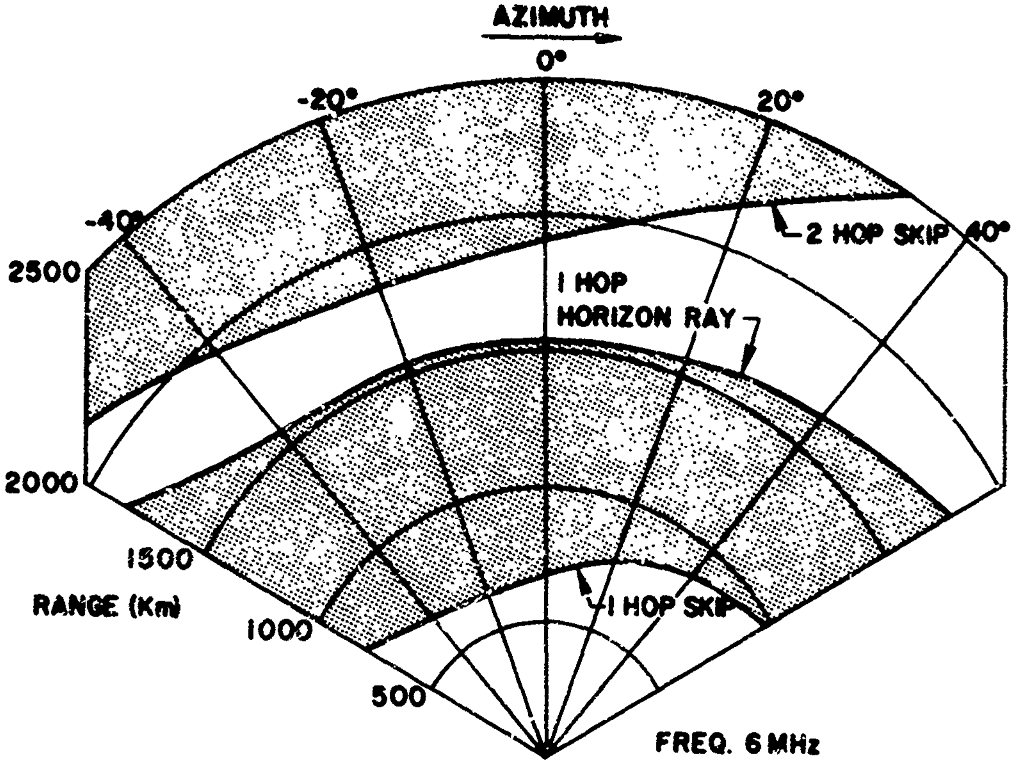
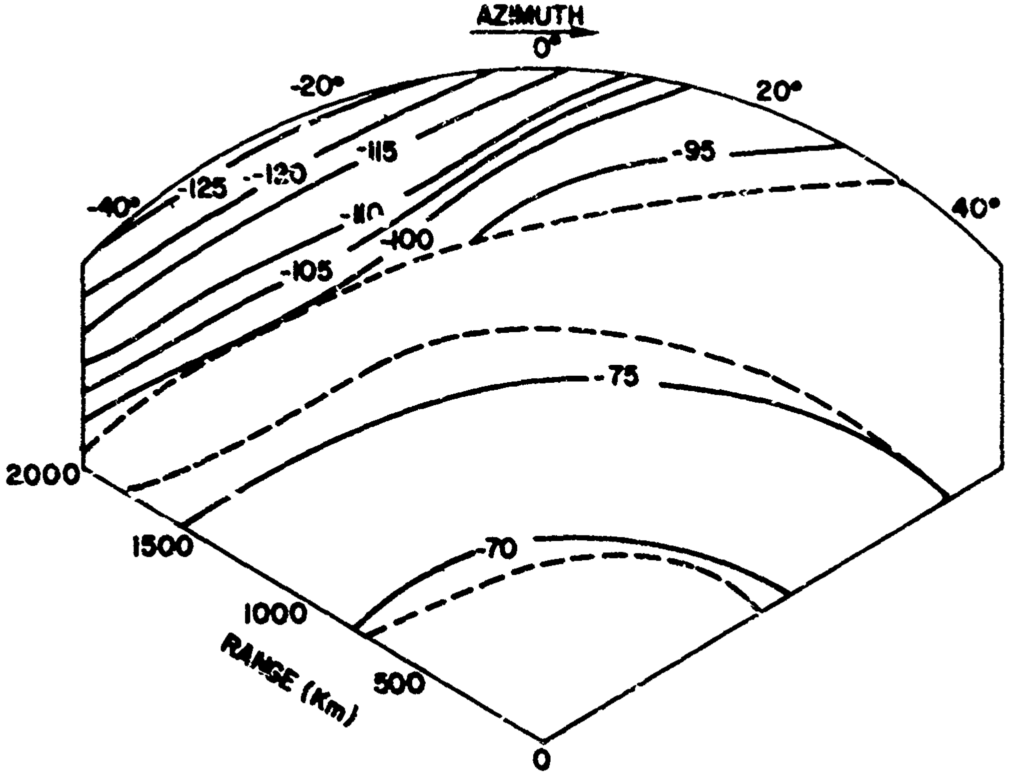

GEOMAGNETIC FIELD
Δε
R
τ/2
RADAR
$V_S = R^2 \cdot \Delta\epsilon \cdot \phi \cdot \frac{C\tau}{2}$
SCATTER VOLUME AND EFFECTIVE VERTICAL ANGLE

FIGURE 7: Illustrating scatter volume (V_S) in presence of strong refraction and geomagnetic field. Azimuthal antenna beamwidth is ϕ.



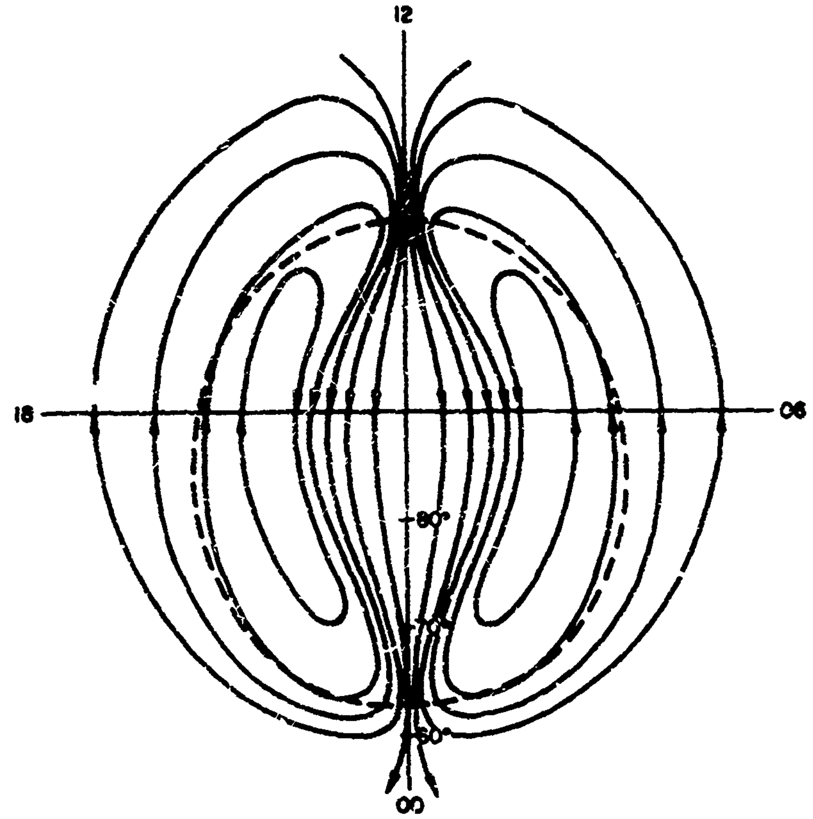
REGIONS OF MAGNETIC ORTHOGOALITY AT MIDNIGHT; JAN ; SSN 100

FIGURE 8: Regions of F-region geomagnetic orthogonality (shaded) at 6 MHz for local midnight in January at Sunspot Number 100. Radar located at 60° geomagnetic latitude.



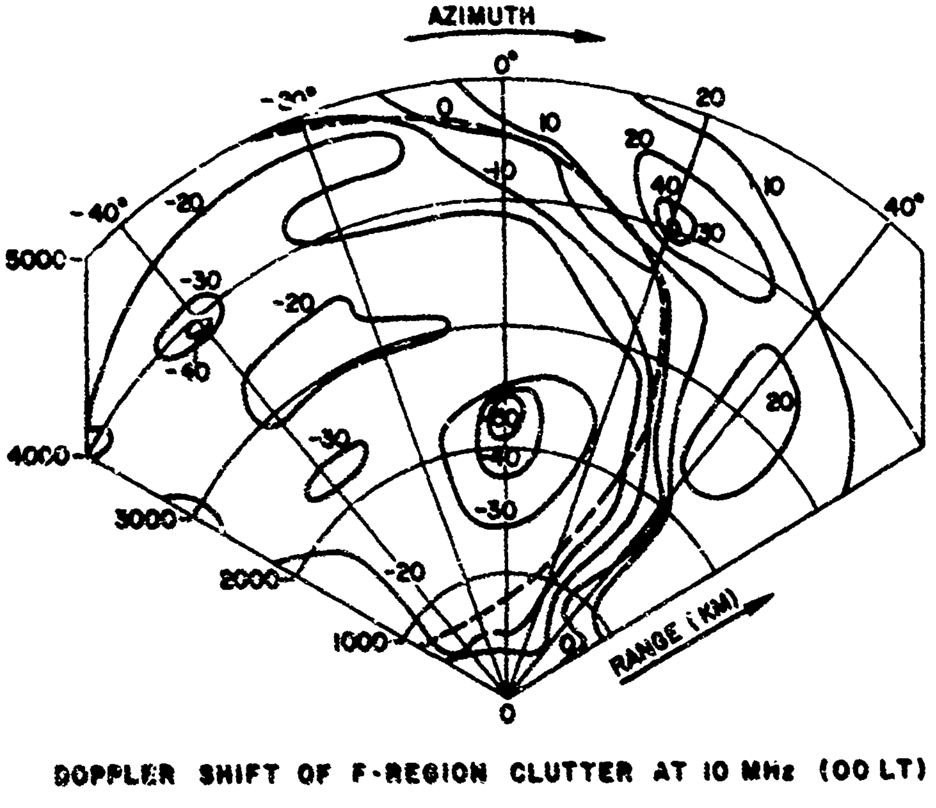
MEDIAN F-REGION CLUTTER POWER (dBW) FOR EXPERIMENTAL RADAR

FIGURE 9: Contours of median received F-region clutter power (dBW) for experimental radar. Dashed lines are boundaries of regions of orthogonality shown in Figure 8.



F-REGION PLASMA VELOCITY IN CORRECTED GEOMAGNETIC COORDINATES

FIGURE 10: Illustrating average F-region plasma velocity in CGS. Dashed line represents the equatorial boundary of the auroral oval.



DOPPLER SHIFT OF F-REGION CLUTTER AT 10 MHz (00 LT)

FIGURE 11: Doppler shift at 10 MHz for F-region clutter at local midnight for an OTH radar located at

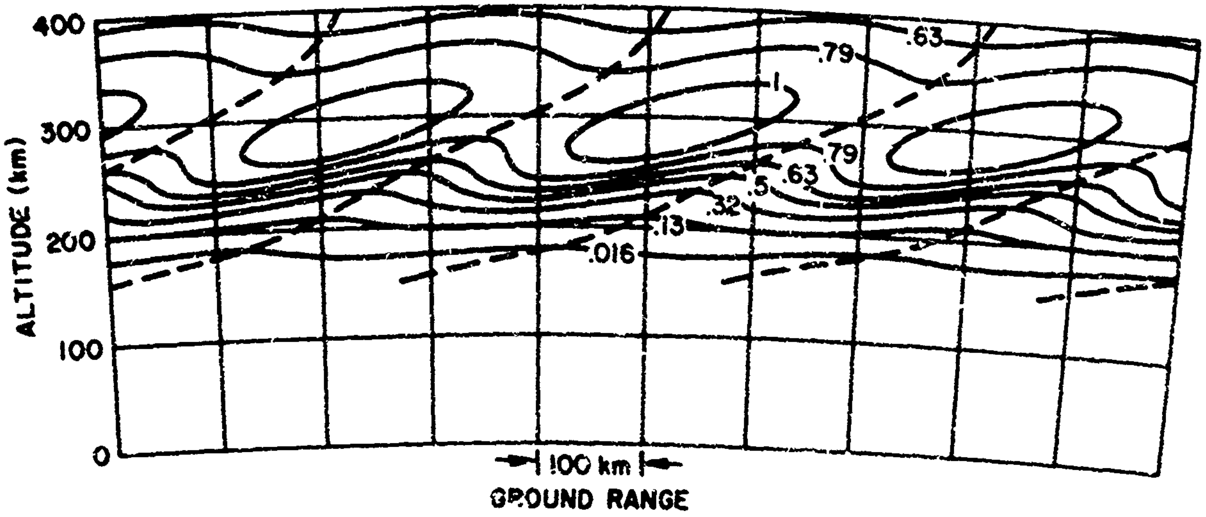


FIGURE 12: Contours of electron density, normalized to the maximum, for a Chapman layer perturbed by a TID (after Francis, private communication). Dashed lines represent phase fronts.

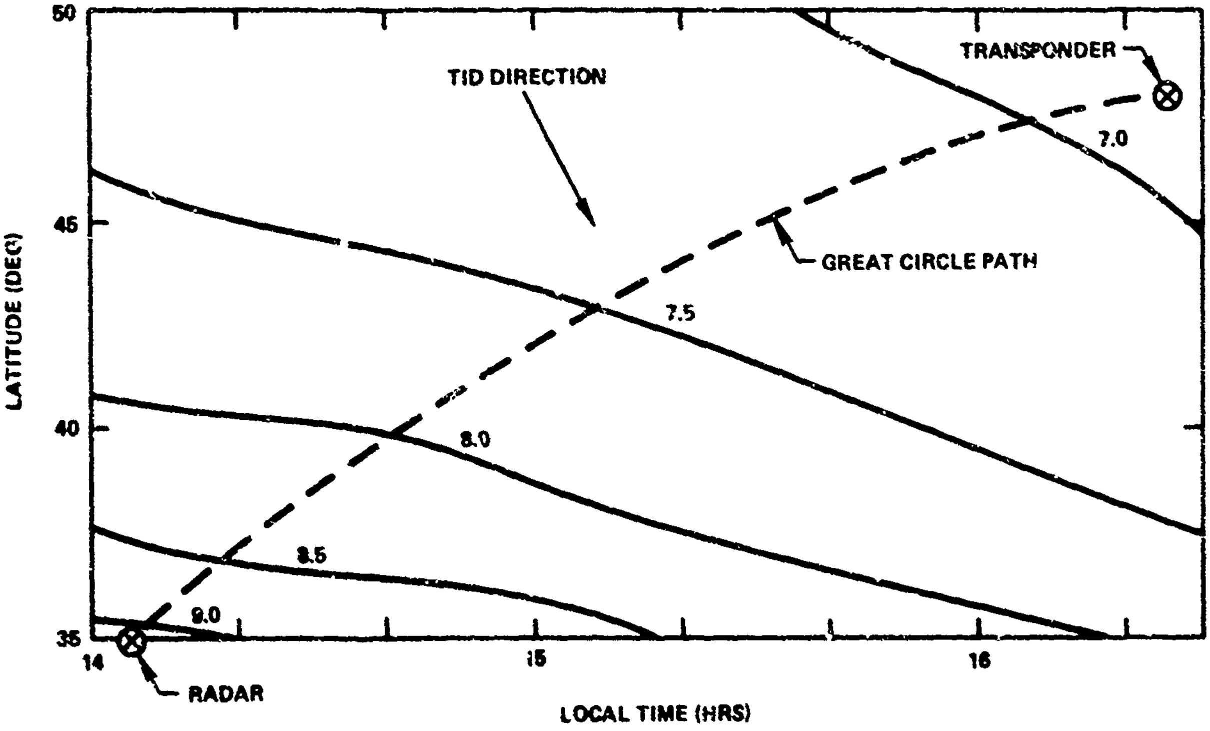


FIGURE 13: Experiment geometry showing OTH radar illuminating distant transponder while a TID crosses the propagation path at right angles. Contours of f_oF_2 for the ambient ionosphere are shown.

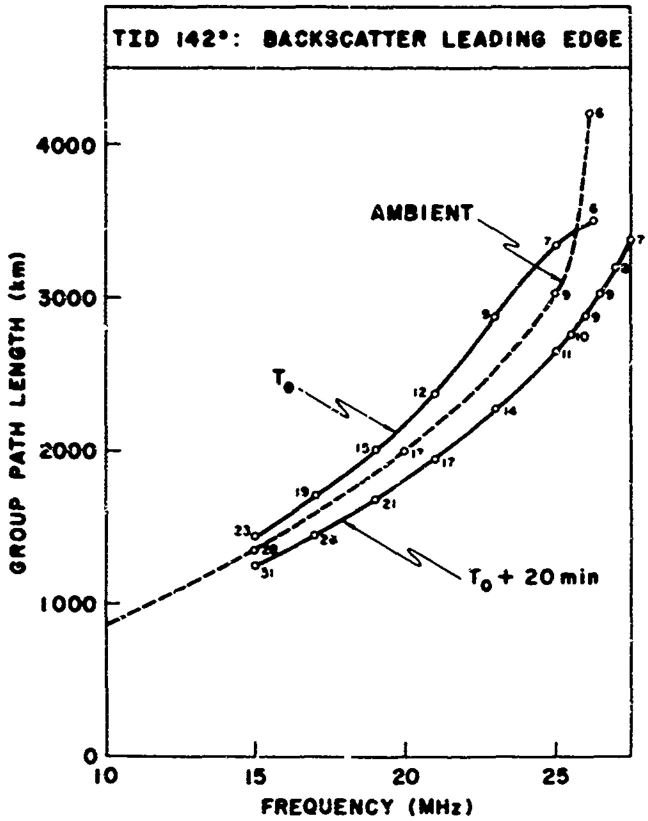


FIGURE 14: Simulated leading edges of backscatter ionograms in the ambient ionosphere and at two extreme phases of the TID. The small numbers represent take-off angles at the radar.

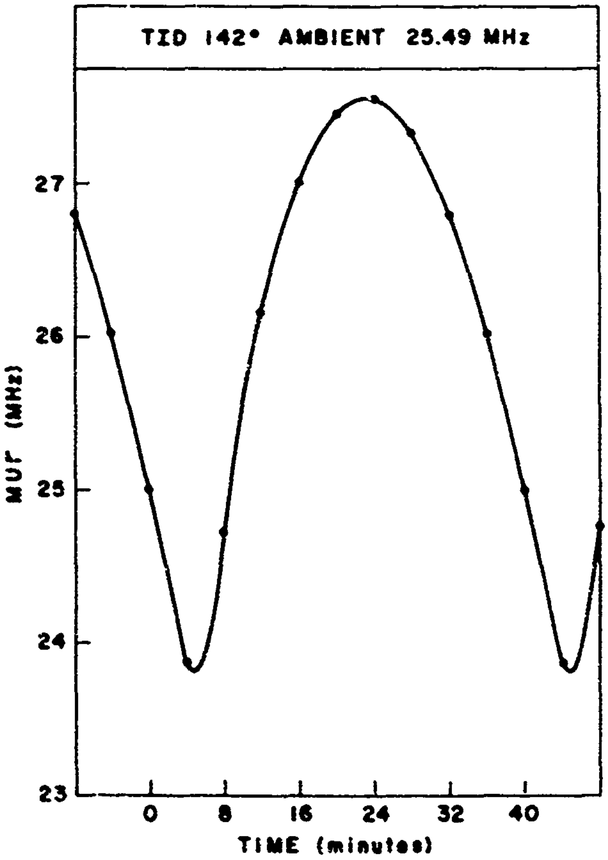


FIGURE 15: Variation with TID phase of the frequency required to maintain the transponder at the first hop skip range (for the ambient ionosphere, this frequency is 25.49 MHz).

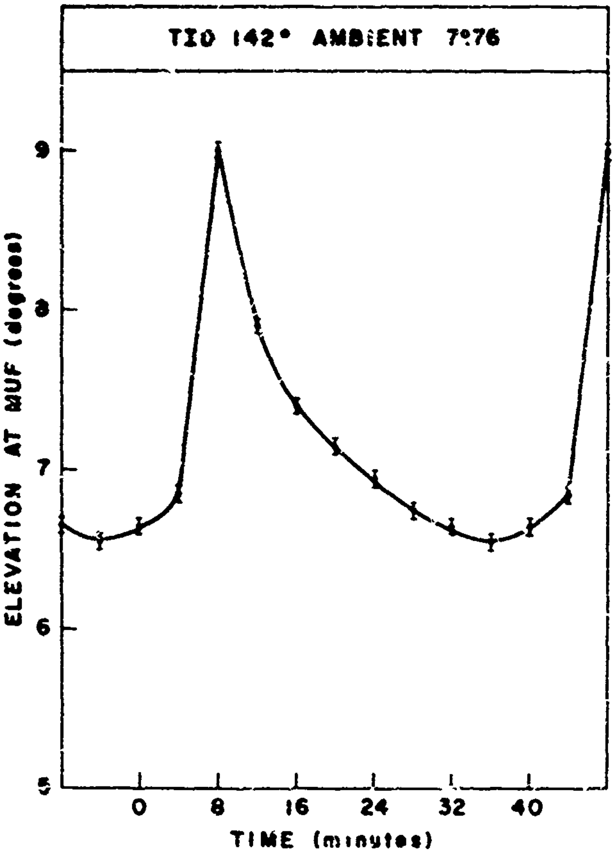


FIGURE 16: Variation with TID phase of take-off elevation at the optimum frequency (ref. Figure 15).

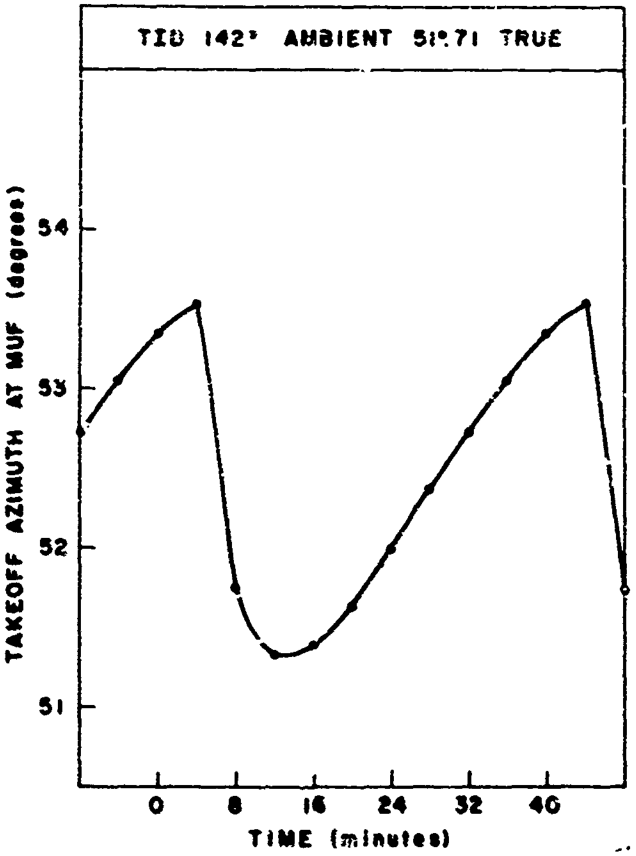


FIGURE 17: Variation with TID phase of take-off azimuth at the optimum frequency (ref. Figure 15).

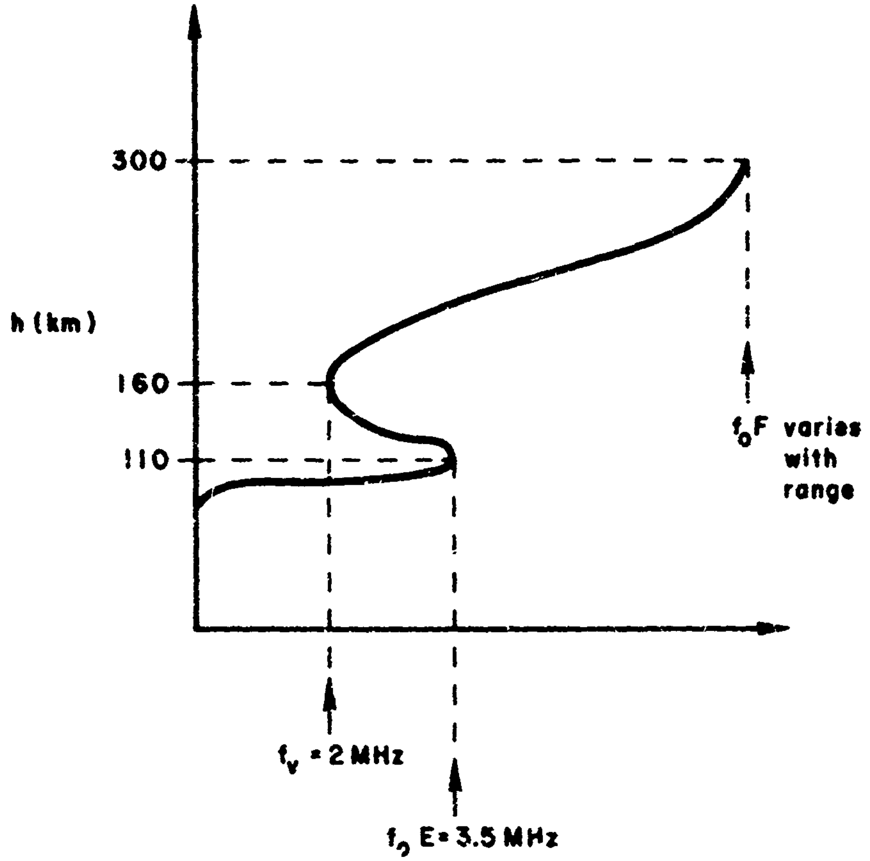

300
h (km)
160
110
f_oF varies with range
f_v = 2 MHz
f_o E = 3.5 MHz

FIGURE 18: Vertical ionospheric plasma frequency profile, showing E and F layers with valley between them.

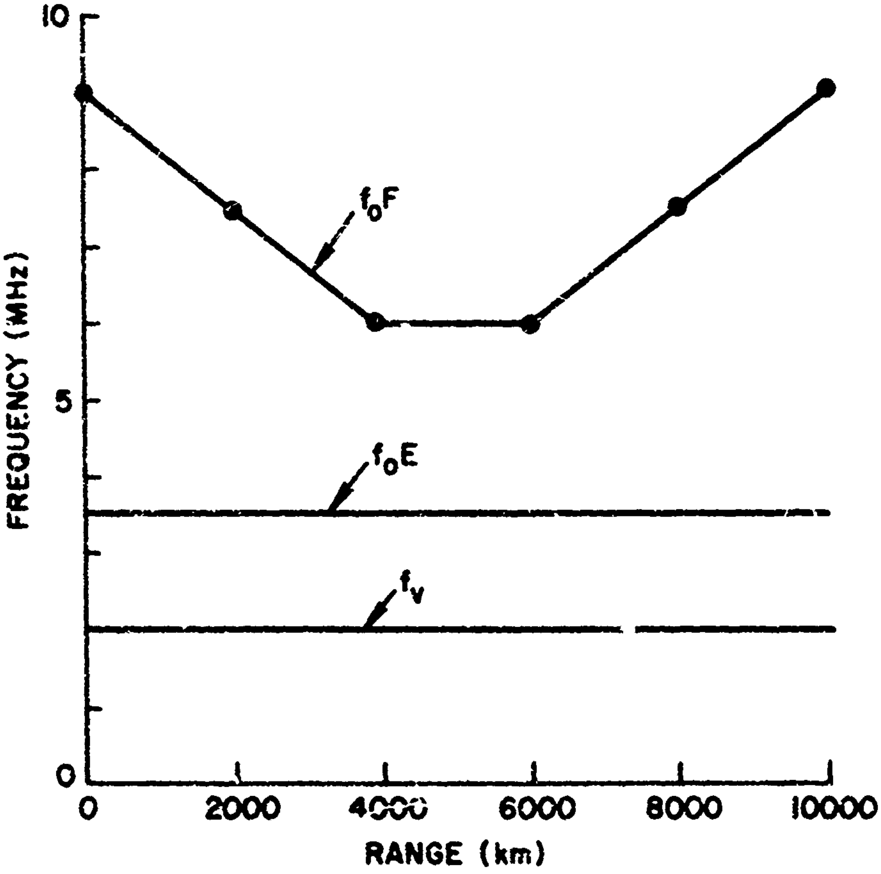


FIGURE 19: Range variation of parameters of the vertical profile shown in Figure 18.

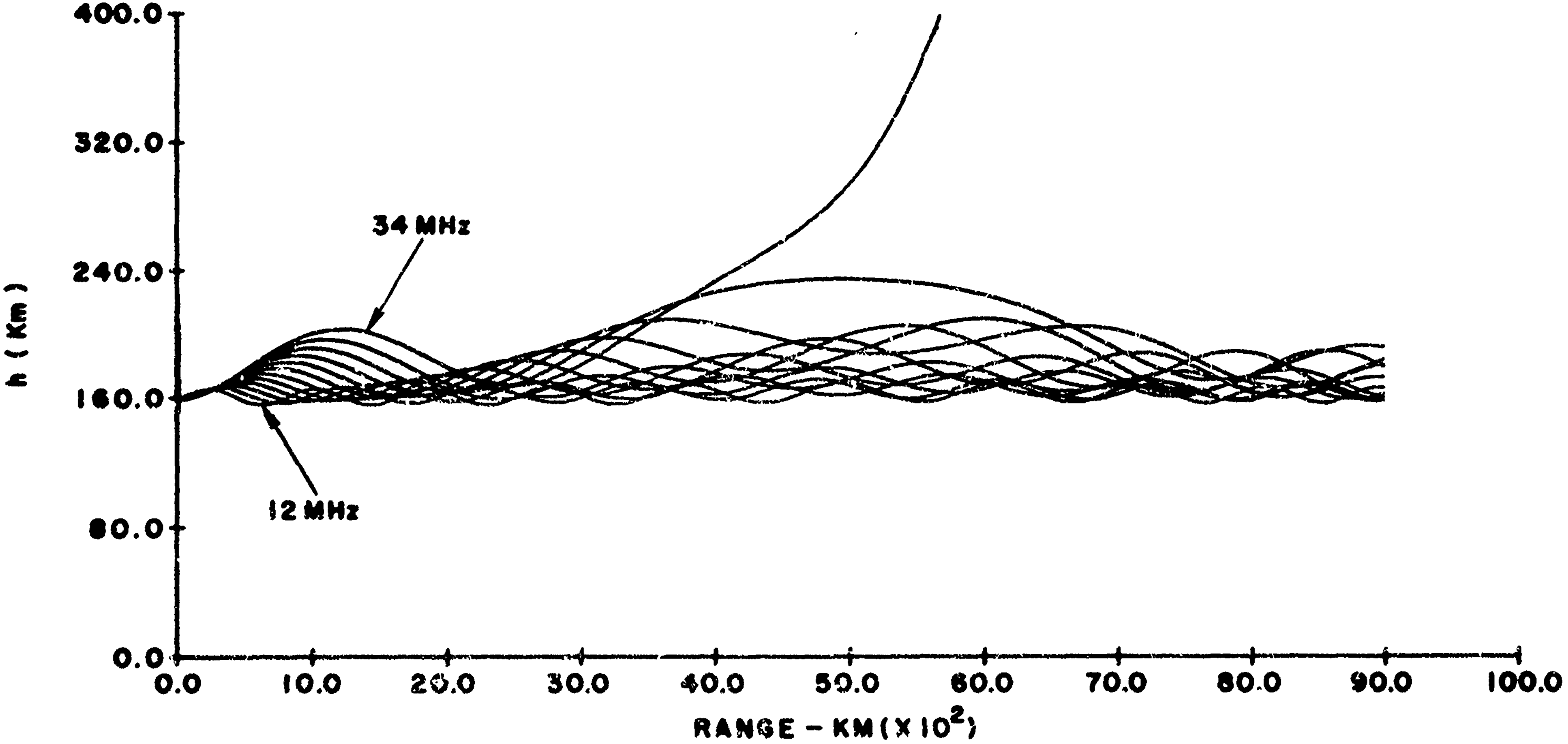


FIGURE 20: Rays traced from an elevated source, with initially horizontal direction, in the ionosphere shown in Figures 18 and 19. Frequency varies from 12 MHz to 34 MHz in 2 MHz increments.

DISCUSSION

C.S. Goutelard, Fr

La "cross-section" que vous avez calculée a-t-elle été donnée pour une détection particulière du rayon par rapport au champs magnétique terrestre ou bien considérez-vous qu'elle est, en HF, isotrapic?

La fréquence doppler que vous avez calculée et qui correspond à des vitesses de 2000km/H corrsepond-elle à une orientation normale du rayon de l'onde par rapport au champs magnétique ou de la vitesse du plasma?

Author's Reply

The cross-section is highly anisotropic (6-10dB/degrees off-perpendicularily) and this anisotropy can be taken into account in a more sophisticated computation of scatter volume. In the simplified treatment considered here, the cross section shown should be considered as an average value over n ± degree off-perpendicularily.

The Doppler model was developed from satellite measurements of plasma velocity auroral radar Doppler measurements and a theoretical analysis (by Maeda) of plasma velocity driven by magnetospheric electric fields. The velocities deduced are assumed to be normal to the geomagnetic field direction. Because of the highly anisotropic nature of auroral backscatter, it is assumed that velocities along the geomagnetic field direction (for example, associated with field-aligned currents) will not be observed by a HF radar.

SUMMARY OF SESSION IV

PROPAGATION THROUGH IRREGULARITIES

The propagation through irregularities session encompassed papers which observed a gamut of irregularity sizes at various locations over the globe. Large and small scale irregularities were studied. Equatorial and middle latitude observations have been reported on. One paper described the effects of artificially created irregularities.

There is an increased interest in equatorial irregularities, particularly the F layer nighttime large scale bubbles. A new theoretical study of this subject by **Heron** is included in this volume but was not given. **Rottger** gave new evidence for the generation of the nighttime bubbles by penetrative cumulus convection; **Lange-Hess** described the airglow observations at the end of the equatorial patches.

E layer observations at the magnetic equator were outlined by **Rottger** who recorded backscatter observations from the equatorial electrojet. **Merkel** and **Rubio** compared sporadic E communication reception over the globe with a prediction method developed from the vast accumulation of ionosounder data.

The propagation effects of artificially produced irregularities (by heating) were described by **Sales** and **Elkins** at the conference. Long range HF ducted paths yielded both predicted and as yet unexplained results.

The recently developed use of aperture synthesis, phased arrays and other methods of integrating individual signals received along a series of antennas has lead to the measurement of phase coherence (or lack thereof) along an aperture. At HF **Jones** and **Thomas** showed there were periods when phase coherence was not achieved across a series of antennas; at times this was due to multi-mode propagation but there were occasions when only a single mode was present.

We now understand the physics of irregularity generation and decay to a greater extent than 10 years ago. Description of the morphology of these regions is proceeding rapidly. Both the utility of the information and the discovery of the physics of equatorial irregularity bubbles has brought a new excitement to the field of ionospheric physics.

TRANSEQUATORIAL PROPAGATION THROUGH EQUATORIAL PLASMA BUBBLES - DISCRETE EVENTS.

by

M.L.HERON
Physics Department,
James Cook University,
Townsville, 4811
Australia.

SUMMARY

Recent results from the Jicamarca radar facility and satellite-borne plasma density probes have provided the general characteristics of rising depletions in the evening ionospheric plasma over the equator. The depletions are not spherical bubbles but extend along the magnetic field lines so that their feet are always in the lower F region. The bubbles are driven by E x B fields. Generally they move with the background ionosphere horizontally towards the east at about 125 ms^{-1}, but often they have a vertical component as well.

A propagation model has been developed, which uses the observed characteristics of the bubbles to predict radio wave guiding along the depleted tubes. The numerical model is set up to calculate power, Doppler shift, elevation and azimuth at the receiver of a transequatorial transmitter-receiver pair of stations. Individual bubbles cause a burst of power on the link which is initially a few hertz positively Doppler shifted and then moves into a negative shift with maximum power at around -10 to -20 Hz depending on radio frequency and bubble geometry. After that the power drops off and the frequency goes rapidly to negative values beyond -100 Hz.

The successful evaluation of the numerical model in terms of a few observations of transequatorial propagation events leads to the suggestion of an experiment which may elucidate some of the properties of equatorial plasma bubbles.

1. INTRODUCTION

The two main categories of transequatorial radio propagation have been labelled "afternoon" and evening" types. The afternoon transequatorial propagation occurs principally because of refraction at electron density gradients in the equatorial ionosphere. Gibson-Wilde (1969), for example, has shown that the high density equatorial anomaly regions, which straddle the equator during the day, have sufficiently large gradients to cause transequatorial propagation by a double (chordal) reflection off the bottomside of the F-layer. There appears to be a general consensus (Nielson, 1968a) that this mechanism works in the afternoon when the equatorial ionosphere is well developed, and for radio waves with frequencies up to about 60 MHz.

The second category of transequatorial propagation occurs in the evenings and includes frequencies in the VHF range, extending at least to the 2 metre amateur band (Reisert and Pfeffer, 1978). At these times and frequencies the bottomside reflection mechanism cannot work and some other explanation is required for the strong and frequent openings of transequatorial propagation. Previous theoretical approaches to the problem of finding a mechanism which will adequately explain the evening type of transequatorial propagation have included forward scattering from a volume of irregularities (Kuriki et al, 1972) and ducting between spread-F irregularities in a field aligned multi-scatter mode (Nielson, 1968a, 1969). These scatter modes encounter some difficulties in the explanation of off-great circle propagation and propagation between stations which are magnetically asymmetric. (Nielson and Crochet, 1974).

Experimental observations which require special attention are the Doppler shifts and the durations of VHF transequatorial openings. A series of spectra taken at 20 minute intervals at 54,72 and 90 MHz on the Oahu-Raratonga path by Nielson (1968b) showed that the incoming energy was Doppler shifted slightly positively at first and then increasingly negatively until the conclusion of the event after about an hour. Event durations of the order of an hour have also been observed on the 2m band. These results suggest a moving duct or scattering patch of irregularities in the F region.

This paper shows how the Doppler shift history, the off-great circle propagation and the approximately one hour event duration are consistent with transequatorial propagation along equatorial plasma bubbles.

2. EQUATORIAL PLASMA BUBBLE OBSERVATIONS

Plasma bubbles in the equatorial ionosphere have been reported and widely discussed in the literature since the appearance of rising plume-like structures on the Jicamarca 50 MHz radar records (Woodman and La Hoz, 1976). The radar records are derived from radio waves reflected off small scale structure of the scale size of 3 metres. A sketch of one such record is reproduced in figure 1 where the ordinate axis represents altitude. The graphic display is based on an abscissa scale which assumes that the background ionosphere is drifting eastwards at 125 ms^{-1}; a time period of one hour is equivalent to a distance of 450 km if the drift speed is 125 ms^{-1}. The plume structures are therefore drifting at a velocity near to 125 ms^{-1} and they rise to higher altitudes as time progresses. The apparent tilting of the structures may be due to an off-vertical ascent of the bubbles in the reference frame of the drifting background plasma, or to variations with height of the background plasma drift velocity.

Observations made on the Atmospheric Explorer Satellite, AE-C, flying in a near equatorial orbit show deep depressions in plasma density (McClure et al. 1977). These results show a wide range of bubbles in the equatorial plasma, typically 100 km in width and having one or two orders of magnitude less ionisation density than the background plasma. The bubble regions had steep plasma density gradients in the walls, of the order of 10^6 m^{-4} and localised gradients of 10^8 m^{-4} associated with fine structure of scale sizes down to less than half a kilometre. Indeed the fine structure must persist down to scale sizes of ~3 metres to give the strong radar returns on 50 MHz at Jicamarca.

In the F region, the diffusion time constant for a local plasma density perturbation to spread along a magnetic field line is less than one second. We would expect, therefore, that any bubbles seen in the equatorial plane would extend to occupy the complete magnetic field tube at least to the bottom of the F layer. Observations on the three dimensional aspect of the bubbles were made by Dyson and Benson (1978) using topside sounder data taken whilst the spacecraft was flying inside a bubble region. The echoes received were consistent with tubular field aligned regions of plasma density depletion. More direct observations of the field alignment of the bubbles were described by Heron and Dorling (1979) in terms of a plasma density probe on the ESRO-4 satellite flying in a low altitude polar orbit. At times, this spacecraft flew through depleted regions symmetrically placed about the dip equator and in the magnetic meridian. Observations of intertropical dark bands in 630 nm airglow emission by Weber et al (1978) also confirms the extension of the disturbed regions out of the equatorial plane.

3. EQUATORIAL PLASMA BUBBLE THEORY

The development of perturbations in the horizontal plasma density contours into rising bubbles of depleted density, depends upon zonal winds and high plasma density gradients. A simplified view of the growth mechanism is given in figure 2 where a perturbation in the height of a plasma density contour is shown on the underside of the F layer where the vertical grad'ent [illegible] a density is high. A horizontal wind, u, in the neutral medium is more strongly coupled to the [illegible] region A than in region C, and is less strongly coupled in B than in C. This coupling, by virtue [illegible] g, results in local polarisation electric fields as shown. In region A, which is an en[illegible] density region, the polarisation electric field and the earth's magnetic field drive the [illegible] density perturbation downwards into a higher loss regime. This contrasts with the less dense plasma [illegible] locality B which is driven upwards by an $\underline{E} \times \underline{B}$ force, into a stable growth situation.

These concepts were developed by Ossakow and Chaturvedi (1977) who used typical densities and conductivities over the equator, and assumed an initial Rayleigh-Taylor gravitational instability. The calculations gave vertical velocities which increased to several hundred metres per second at several scale heights above the F layer. Anderson and Haerendel (1979) have calculated the vertical rise of bubbles in terms of the electron content gradients and conductivities integrated along the field lines. Under this scheme the maximum vertical gradient in tube content is normally near or above the local F layer peak at the equator and suggests that bubbles may first appear higher up at about 500 km.

These theoretical approaches do not address the problem of what happens to the bubbles well above the F layer peak, at say 10,000 km, nor do they explain the observation of McClure et al. (1977) that there are many non-rising bubbles at a wide range of heights in the F layer. The calculations which follow in this paper, for a waveguide mode of transequatorial propagation through plasma bubbles, discreetly cover both the non-rising and the exponentially rising cases and an experiment is described which will discriminate between them.

4. NUMERICAL MODELLING OF EQUATORIAL PLASMA BUBBLES.

The objective of this project was to construct a numerical model of a plasma bubble which would comply with the observed or theoretically predicted characteristics of ionospheric phenomena, and to investigate the ensuing transequatorial propagation. The basic model is a field aligned tube with low plasma density and with its height in the equatorial plane being governed by a vertical velocity v_v of the form

$$v_v = v_o \exp((h - h_o)/H) \tag{1}$$

where v_o is the initial vertical velocity at height h_o and H is the scale height for the vertical velocity. In the numerical model, v_o could be varied but H was set to an arbitrary value of 74 km which arose in the calculations of Anderson and Haerendel (1979). A horizontal eastwards velocity component v_H was applied to the bubble. The bubble was given a diameter of 100 km and it was assumed to cease at an altitude of 250 km where the F layer ionisation gives way to essentially neutral gas at lower altitudes at night.

The bubble was initiated with given v_o and h_o values and subsequently developed by allowing it to move eastwards and upwards. Figure 3 shows this height development for bubbles initiated at 350 km and with a series of v_o values. Figure 4 shows the latitudes and longitudes of the 'feet' of the same bubbles at the 250 km altitude level where the open end of the waveguide would appear to a ground-based observer. As the altitude of the bubble increases in the equatorial plane the feet move out to high latitudes. The exponentially increasing vertical velocity generates an exponentially increasing height and this model predicts that the feet of the bubble pass very quickly through the mid-latitudes.

5. EQUATORIAL PLASMA BUBBLES AS WAVEGUIDES.

When the plasma density inside the bubble is less than that outside, there is a tendency for a radio wave propagating inside the bubble to reflect off the walls and remain trapped inside the bubble. In fact the condition for total internal reflection can be invoked as a criterion for deciding which rays are trapped and which are refracted outwards away from the waveguide as shown in figure 5. If the critical angle is θ_c and μ_o, μ_i are refractive indices outside and inside the wave guide respectively then

$$\cos \theta_c = \mu_o/\mu_i \; . \tag{2}$$

The refractive indices are frequency dependent according to the relation

$$\mu^2 = 1 - kN/f^2 \qquad (3)$$

where N is the electron density and k is a constant value of 80.5 in MKS units. In the model the value for θ_c is taken as the minimum value occurring along the waveguide, which in practice is the value at the apex of the magnetic field line. The angle θ_c then defines a cone of acceptance such that rays encountering the bubble end-face with $\theta < \theta_c$ will be transmitted to the magnetically conjugate zone, while rays with $\theta > \theta_c$ will be lost from the waveguide.

In reality this sharp criterion for acceptance of rays may not precisely describe transequatorial propagation characteristics because of high plasma density gradients at field-aligned needle-like fine-structure irregularities within the bubble system which may in general prevent rays with $\theta > \theta_c$ from escaping. Offsetting this is the possibility of energy loss through the bottom wall of the bubble if the depleted zone continues to low altitudes as the dark bands in the 630 nm airglow would suggest (Weber et al., 1978).

The cone of acceptance has an angle θ_c which is frequency dependent. For higher frequencies the angle is less, and for a conjugate pair of stations, the duration of a transequatorial communication link is predicted to be shorter.

Once the decision is made that a ray from a particular transmitter can propagate to a given receiver then the power calculation is made on the basis of the size of the solid angles subtended by the waveguide end-faces at the stations. The Doppler shift is calculated from the phase path length between the transmitter and the receiver via the axis of the bubble at two times spaced 10 seconds apart.

For the bubbles described in figure 3 and figure 4 the observable quantities of the transequatorial propagation between two conjugate stations were calculated. These observables are power (on a relative scale, not absolute), Doppler shift, elevation angle and azimuth of incoming rays at the receiver. They are all calculated as a function of the longitude of the bubble which is directly related to time by the horizontal eastwards drift velocity assumed in the model. For this calculation the longitudes of the stations have to be set relative to the initial longitude of the bubble, and for convenience and generality we designate the receiver and transmitter to have zero longitude. In these calculation the stations are symmetrically placed at ±20° latitude and all coordinates are on the magnetic frames of reference. The location of the stations is chosen to illustrate the relationship between plasma bubbles and transequatorial propagation, but the computer code is general and the calculations can be made for any pair of stations.

6. RESULTS

As a bubble drifts across the longitude sector of the radio stations, the transequatorial propagation path may open for a certain period which depends upon the geometric configuration of the bubble and the stations, and on the frequency of the radio wave. This is shown in figure 6 where the solid line refers to 50 MHz and the broken line refers to 100 MHz. The power axis in db has an arbitrary reference level. The asymmetrical signal enhancements for the bubbles with initial vertical velocity $v_o = 26\ \mathrm{ms}^{-1}$ (in equation 1) arise because the end-faces of the bubble sweep past quite close to the transmitter and receiver stations. This effect can be seen on figure 4. At the higher frequency the duration of the transequatorial propagation opening is reduced because the cone of acceptance for rays entering the waveguide is reduced. The fact that identical power levels are reached near the centre of the opening indicates that for a time, even at 100 MHz, all the rays in the solid angle which the bubble end-face subtended to the stations were acceptable on the $\theta < \theta_c$ criterion.

The angle of arrival of the ray which comes from the centre of the end-face of the bubble to the receiver is shown in figures 7 and 8. The azimuth (figure 7) reflects mainly the progression of the bubble in longitude while the elevation (figure 8) relates primarily to the changing height of the bubble at its apex in the equatorial plane.

The Doppler shifts are shown in figure 9. Initially when the transequatorial propagation link opens as the bubble approaches from the west, the Doppler shift is slightly positive. For rising bubbles the shift becomes increasingly negative as the bubble moves past the stations. Eventually, as the power on the link diminishes the Doppler shift reaches values around -20 to -40Hz at f = 50MHz. The Doppler shifts are proportional to f, the radio wave frequency, in this model.

7. DISCUSSION

A direct consequence of transequatorial propagation via the plasma bubble mode is the discrete nature of circuit openings. As the bubbles drift across the appropriate longitude sector the communication link is predicted to open for a period ranging from a few minutes up to the order of an hour depending on geometry and radio frequency. Most transequatorial propagation summaries appear in the form of occurrence statistics, latitude dependence and frequency dependence and Heron and McNamara (1979) have evaluated these statistical characteristics using the numerical model for transequatorial propagation via plasma bubbles.

Some experimental data can however be examined on the basis of discrete openings. The high frequency records of Kuriki et al. (1972) show event structure of this order of duration, though at lower frequencies (where event durations are predicted to be longer) they are not always separable. Successive plasma bubbles, as shown by the series of plumes on the Jicamarca radar record by Woodman and La Hoz in their figure 3e, are spaced less than half an hour apart and particularly at the lower end of the VHF band the cones of acceptance are likely to overlap. Records of S-meter levels on a 2-metre link between Rhodesia and Greece show openings lasting typically 40 minutes to 1 hour (Röttger, private communication based on logs by ZE2JV and SV1AB).

The results of Nielson (1968b) show the discrete nature of VHF transequatorial propagation events by highlighting the development of the Doppler shift during circuit openings. The Doppler shifts on four frequencies are shown in his figure 2 for the Oahu-Raratonga path. The series at 2200 to 2300 on 25 March 1968 show quite clearly the development of the Doppler shift with time. The values for 90 MHz were scaled off and are shown in figure 10. The trend from a slightly positive initial frequency shift to a highly negative shift follows the general pattern which is calculated for a rising bubble in figure 9 (v_o = 26 ms^{-1}). However the rapid change in the first twenty minutes would, in terms of the present model, indicate a more rapid rise of the bubble initially, followed by a slower ascent over the middle of the observation period. The magnitude of the Doppler shifts on Nielson's records are approximately proportional to the radio frequency, as suggested by the present model.

8. CONCLUSION

A model has been described which allows the calculation of certain observable parameters of transequatorial propagation based on the assumption that the energy is transmitted through equatorial plasma bubbles. The calculations here are for a particular pair of stations placed symmetrically 20° away from the magnetic equator, but this geometry is not a restriction of the computer code. The parameters calculated are quantities which are observable at the receiving station, being elevation, azimuth, power and Doppler shift. In the model these parameters are calculated as functions of the longitude of the bubble whereas the experimental observations would be made as functions of time.

The bubble waveguide approach adequately explains relative power levels and Doppler shift observations available in the literature. Correlations between transequatorial propagation and spread-F can be understood in terms of the plasma bubble mechanism for transequatorial propagation. The existence of a skip zone right at the equator (Nielson, 1968b) is consistent with the bubble waveguide concept. While the bubble mechanism adequately allows for off-great circle propagation at particular times, there is no obvious preference in the simple model for either side of the great circle path to be preferred for magnetically symmetric stations.

In the past, many transequatorial propagation observations have been reported on a statistical basis. A separate study by Heron and McNamara (1979) shows that the model for transequatorial propagation via plasma bubbles gives occurence statistics which, in general, agree with these observations of the evening type of transequatorial propagation. To test the results outlined in this paper we require more observations of parameters during the individual transequatorial propagation openings. The few such observations available not only support the bubble waveguide concept but also indicate that the data redundancy in the four basic observable parameters at a receiving station may be used to give accurate and continuous monitoring of the properties and development of equatorial plasma bubbles.

ACKNOWLEDGEMENTS

This project is supported by the Radio Research Board of Australia. I am grateful to L.F.McNamara for helpful comments on this work.

REFERENCES

ANDERSON, D.N. and G.HAERENDEL (1979), "The motion of depleted plasma regions in the equatorial ionosphere", J.Geophys. Res., submitted 1979

DYSON, P.L. and R.F.BENSON (1978), "Topside sounder observations of equatorial bubbles", Geophys. Res.Letters, 5, 795-798.

GIBSON-WILDE, B.C. (1969), "Relation between the equatorial anomaly and trans-equatorial VHF radio propagation", Radio Science, 4 , 797-802

HERON, M.L. and F.B.DORLING (1979), "Clefts due to equatorial ionospheric plasma density bubbles observed by ESRO-4", Planet. Space Sci., submitted.

HERON, M.L. and L.F.McNAMARA (1979) "Transequatorial VHF propagation through equatorial plasma bubbles", Radio Science, (in press)

KURIKI, I., K.TANOHATA, T. SAKAMOTO and M.IGUCHI (1972), "Propagational mode deducted from signal strengths in the VHF band on the transequatorial path", J.Rad. Res.Lab. (Japan), 19, 175-195.

McCLURE, J.P., W.B.HANSON and J.H.HOFFMAN (1977), "Plasma bubbles and irregularities in the equatorial ionosphere", J. Geophys. Res., 82, 2650-2656.

NIELSON, D.L. (1968a), "A review of VHF transequatorial propagation", paper No. 45 AGARD Conference Proceedings No. 37, part 2 "Scatter Propagation of Radio Waves". AD 685 666.

NIELSON, D.L. (1968b), "The importance of horizontal F-region drifts to transequatorial VHF propagation", paper No. 46, AGARD Conference Proceedings No. 37, "Scatter Propagation of Radio Waves". AD 685 666

NIELSON, D.L. (1969), "Long-range VHF propagation across the geomagnetic equator", Stanford Research Institute Report, Menlo Park, California, USA.

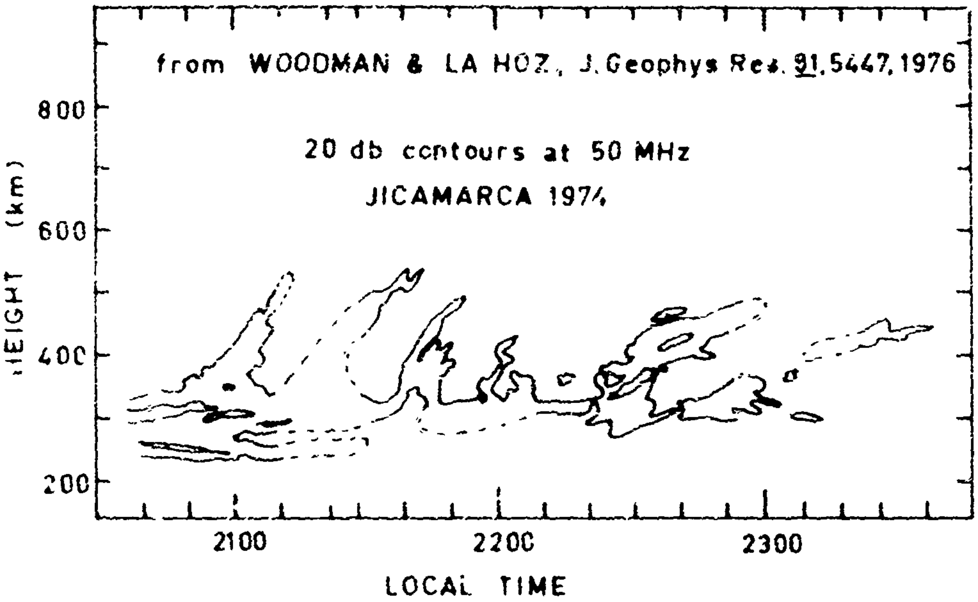
NIELSON, D.L. and M.CROCHET (1974), "Ionospheric propagation of HF and VHF radio waves across the geomagnetic equator", Rev. Geophys. and Space Phys., 12, 688-702

OSSAKOW, S.L. and P. CHATURVEDI (1977), "Morphological studies of rising equatorial spread F bubbles", NRL Mem. Rep. 3650, Naval Research Laboratory, Washington, D.C.

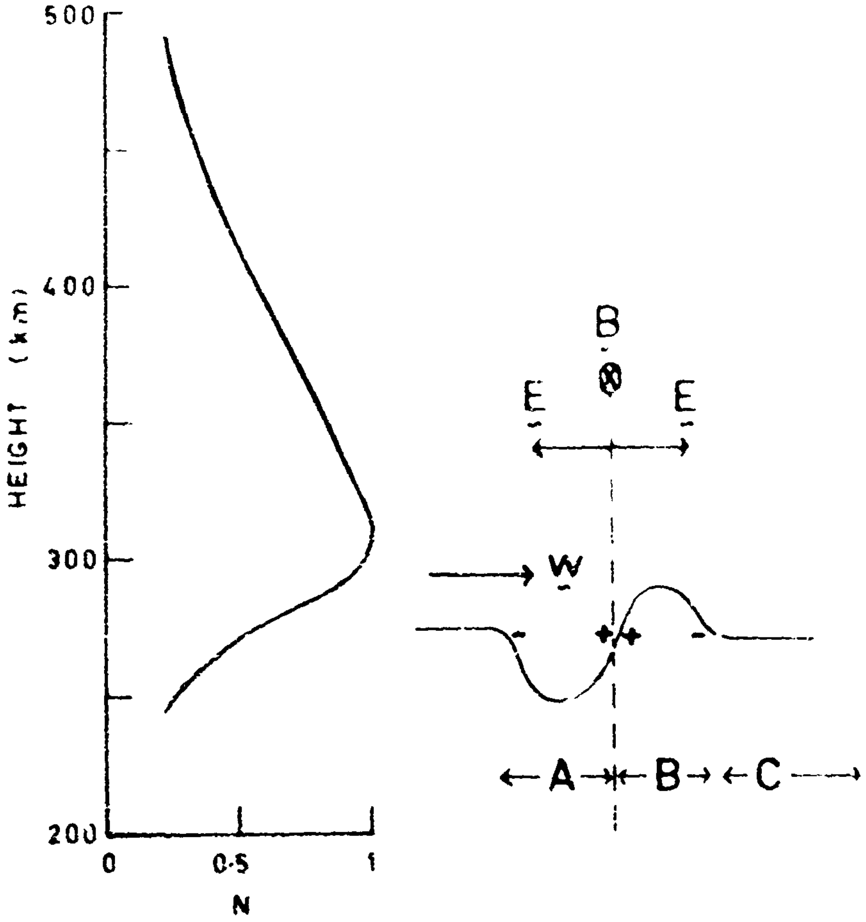
REISERT, J.H. and G. PFEFFER (1978), "A newly discovered mode of VHF Propagation", QST, October 1978, p 11.

WEBER, E.J., J. BUCHAU, R.H. EATHER and S.B. MENDE (1978), "North-south aligned equatorial airglow depletions". J. Geophys. Res., 83, 712-716.

WOODMAN, R.F. and C. La HOZ (1976), "Radar observations of F region equatorial irregularities", J. Geophys. Res., 81, 5447-5466.



[illegible]


500
400
300
200
HEIGHT (km)
0
0·5
1
N
B
E
E
W
+
+
-
-
A
B
C

[illegible] produce vertical [illegible]

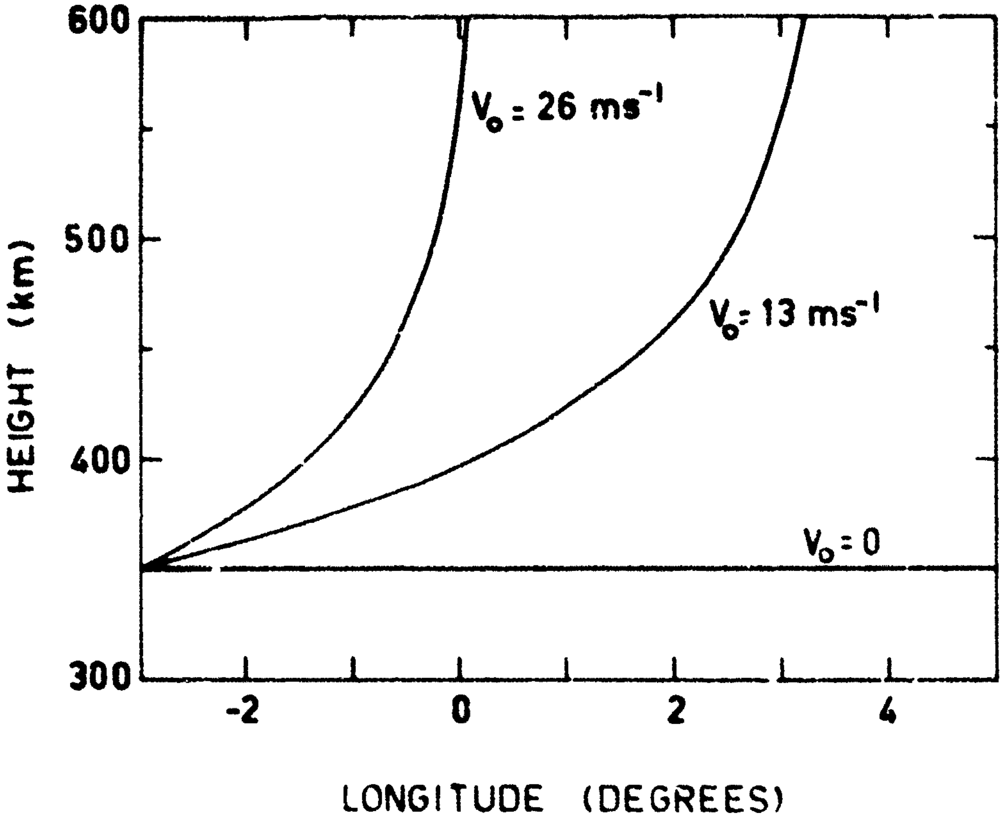


Figure 3. Bubble height, in the equatorial plane, as a function of time for a range of initial vertical velocity values V_o. The horizontal eastward velocity is constant at 125 ms^{-1}.

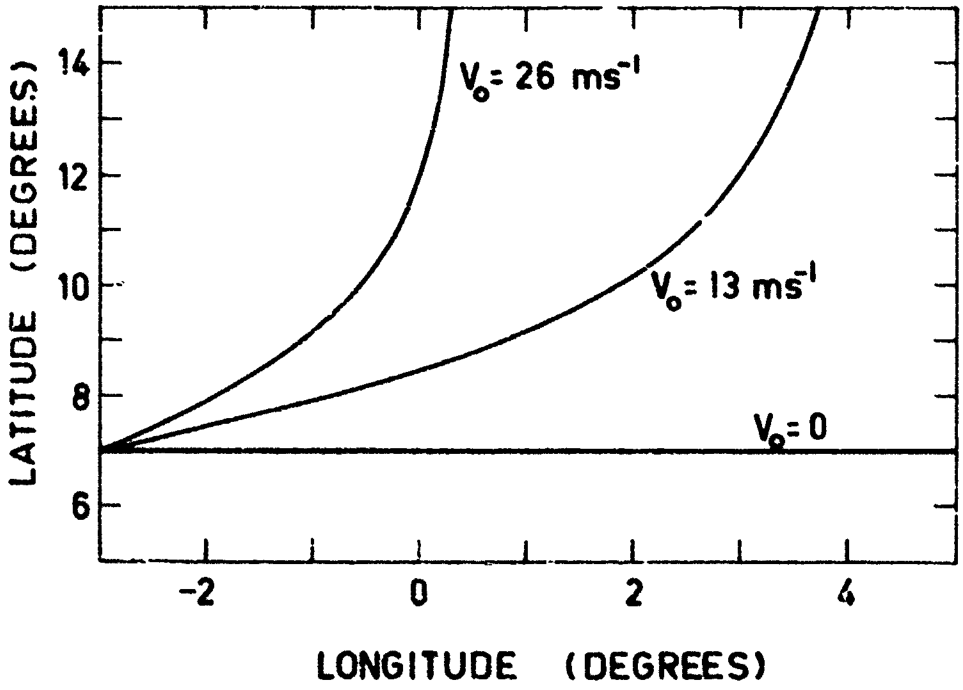


Figure 4. The same bubbles as in figure 3 spread out to higher latitudes as they rise up. The "feet" of the bubble are assumed to emerge below the F layer at 250 km altitude.

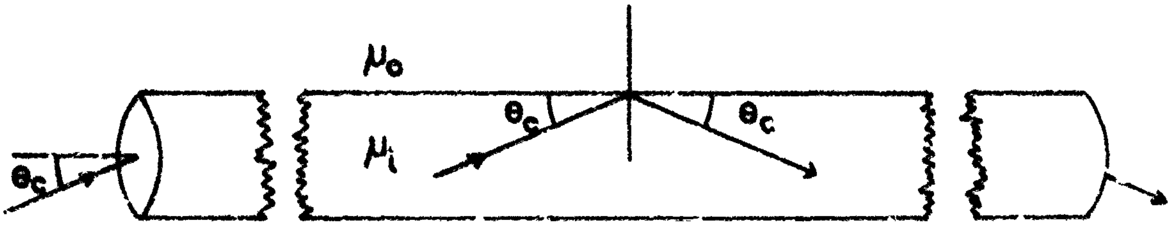


Figure 5. Illustration of the Snell's Law critical ray and its direct relationship to the cone of acceptance at the ends of the waveguide. Two transequatorial stations can establish a radio link if they simultaneously lie inside the cones of acceptance at opposite ends of the guide.

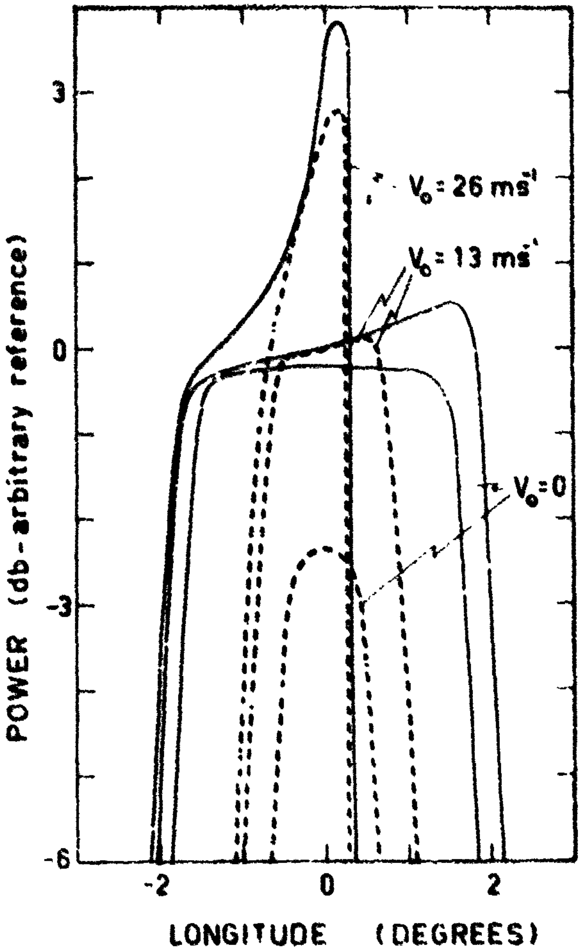


Figure 6. Relative power at the receiver for the bubbles depicted in figure 3. The solid lines are for a radio frequency of 50 MHz and the broken lines are for 100 MHz. The cones of acceptance are smaller in angle for higher frequencies.

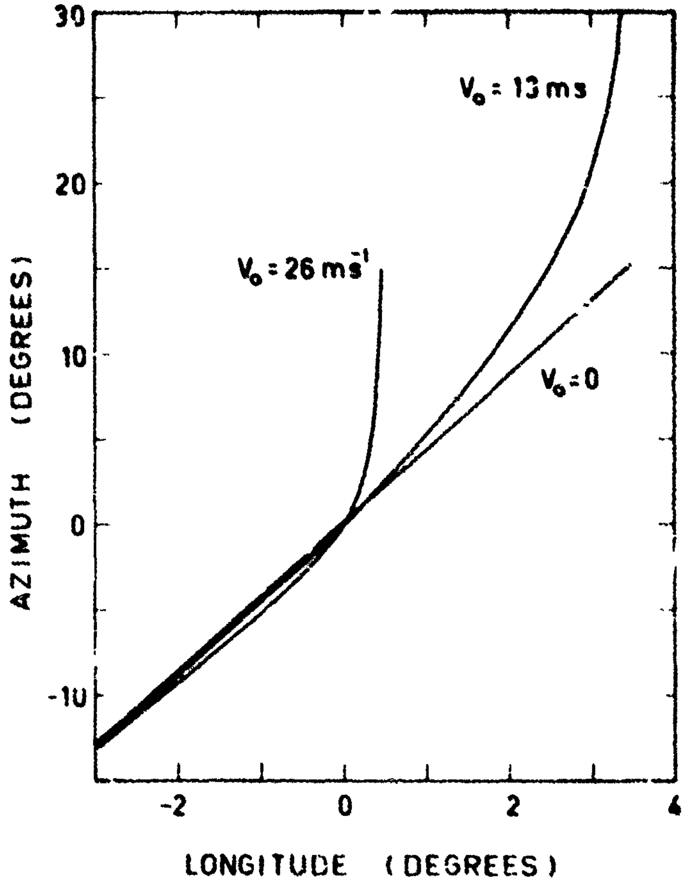


Figure 7. The azimuthal variations of the bubbles depicted in figure [illegible] as they sweep past radio stations at $\pm 20^\circ$ latitude and 0° longitude (magnetic). The end-face of the bubble is taken to be at 250 km altitude.

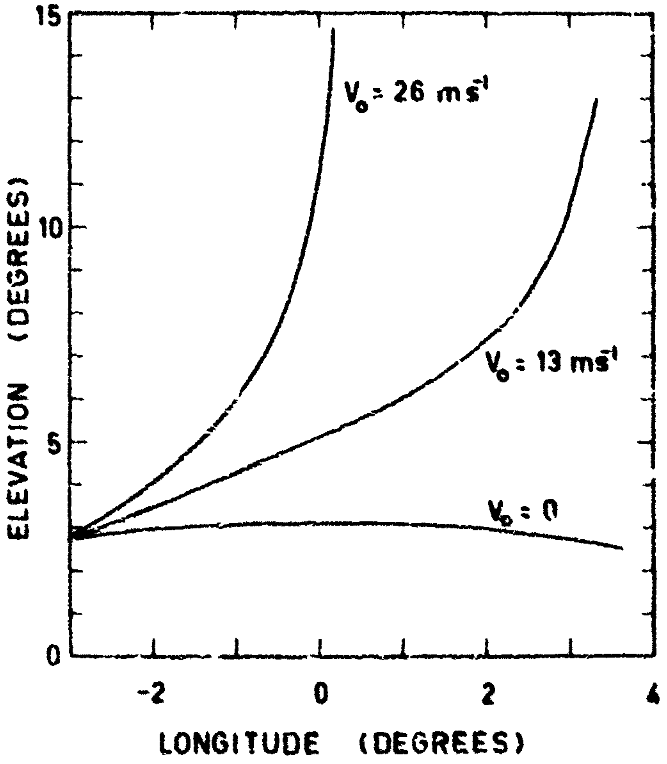


Figure 8. The elevation angles corresponding to figure 7 for stations at $\pm 20^\circ$ latitude and 0° longitude.

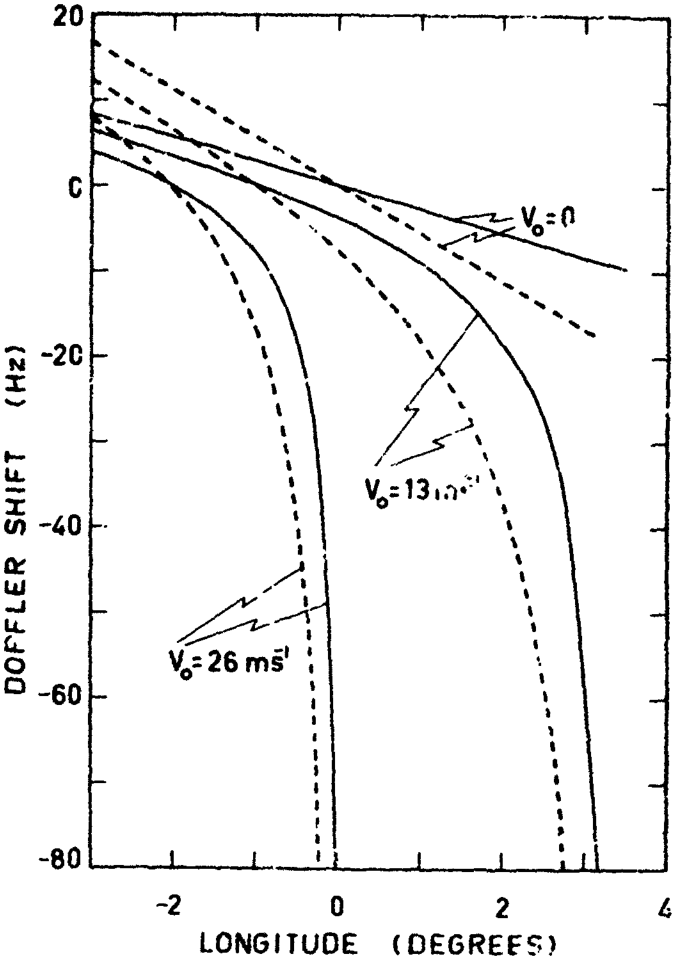


Figure 9. Doppler shifts for the bubbles described in figure 3 and stations at ±20° latitude, 0° longitude (magnetic). The solid lines are for a 50 MHz radio wave the broken lines are for 100 MHz.

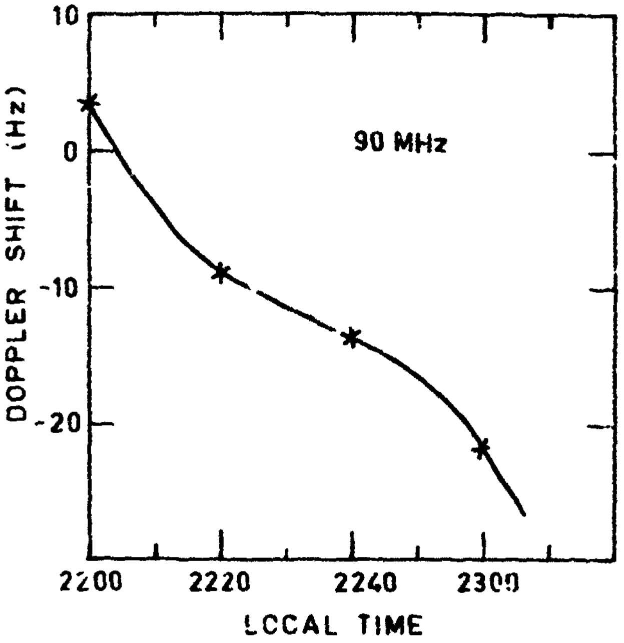


Figure 10. Observed Doppler shifts for a transequatorial propagation event on the path Oahu-Raratonga at 90 MHz by Nielson (1968b). The values were scaled from the published graphs and are to be compared with the theoretical curves in figure 9.

THE INFLUENCE OF E-REGION AND F-REGION IRREGULARITIES ON HF PROPAGATION IN THE EQUATORIAL ZONE

J. Röttger

Max-Planck-Institut für Aeronomie
3411 Katlenburg-Lindau 3, Fed. Rep. Germany

SUMMARY

Amplitude and phase variations of a 6-MHz radio signal on short paths (up to 200 km) in the equatorial zone are evaluated with respect to the influence of ionization irregularities in the E-region and F-region. Spectra of the 6-MHz signal observed by means of an HF-CW-Doppler system operated in Peru are investigated regarding the frequency spread and the frequency shift (RÖTTGER and BECKER, 1977).

The frequency spread observed during daylight hours is caused by the influence of Type II irregularities in the equatorial electrojet flowing in E-region height. The signals are reflected at the F-layer but are caused to scintillate due to the strong ionization irregularities in the passed E-layer (e.g. Figure 1). It is evident that the frequency spread, as expected for Type II irregularities, is strongly correlated to the variation of the H-component of the earth's magnetic field, i.e. the strength of the electrojet. The spread is more pronounced on longer paths (e.g. JIC) than on shorter ones and is also dependent on the path direction with respect to the earth's magnetic field.

The frequency spread observed during the evening hours is caused by equatorial spread-F irregularities. Figure 2 shows that just after sunset the F-layer reflections are split into distinct layers which move with different velocities. This phase is immediately followed by a rapid transition into strongly spread echoes. The strong phase and amplitude variations of the HF signals were evaluated by RÖTTGER and BECKER (1977) with respect to their spectral spread and spectral shift. The shift can either be caused by travelling ionospheric disturbances (RÖTTGER, 1977) and in this case mostly is periodical, or by electrical fields in the F-layer. The latter case is connected with variations of the earth's magnetic field.

The described signal characteristics of amplitude and phase variations can cause problems and limitations in radio communication as well as in direction finding systems. Their geophysical relevance is given by the correlation of signal characteristics with variations of the earth's magnetic field.

REFERENCES

RÖTTGER, J., 1977, "Travelling disturbances in the equatorial ionosphere and their association with penetrative cumulus convection", J. Atmos. Terr. Phys. 39, 987-998.

RÖTTGER, J. and H. BECKER, 1977, "Die HF-CW-Dopplermethode und ihre Anwendung in der Ionosphärenforschung", Kleinheubacher Berichte 20, 243-254.

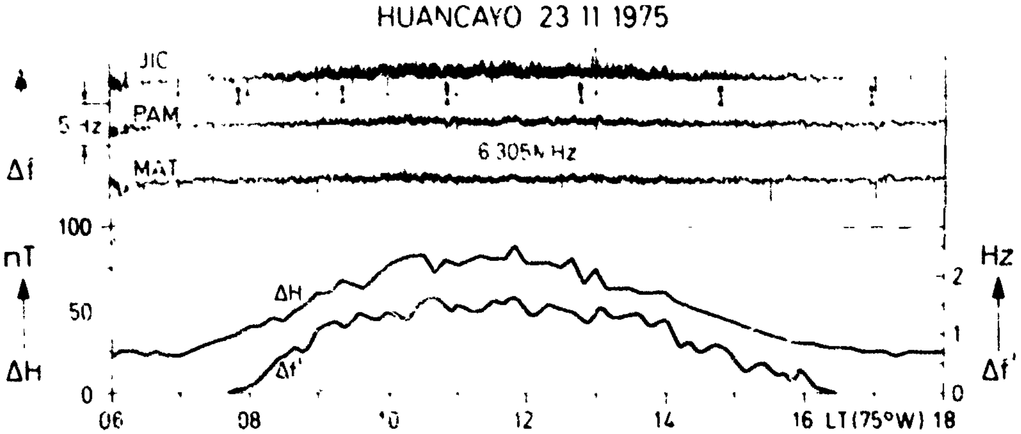


Fig. 1 Frequency scintillations of HF signals reflected in the F-region. The Doppler frequency Δf is measured on three paths in the area near Huancayo/Peru. The scintillations are due to Type II irregularities in the equatorial electrojet (E-region), which is passed by the HF signals. The frequency fluctuations Δf' are therefore strongly correlated to variations of the horizontal component ΔH of the earth's magnetic field.

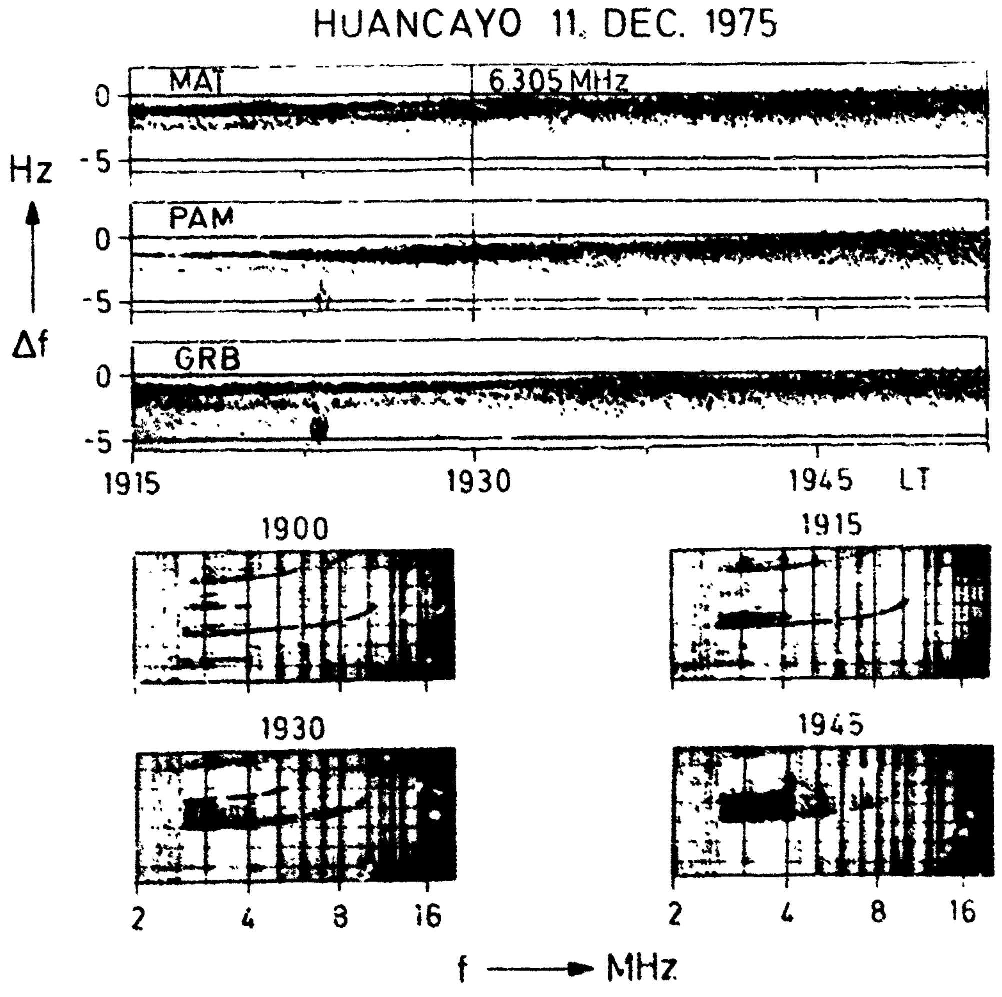


Fig. 2 Doppler frequency variations observed on three short radio paths (10–40 km) during the onset of equatorial spread-F. After splitting into two layers which move upward with different velocities (e.g. MAT), the F-region transits into a regime of ionization irregularities shown by the scattered Doppler traces and by satellite traces on the ionograms after 1930 LT.

DISCUSSION

G.H. Millman, US

With reference to slide No. 6, you discussed the correlation of the magnetic field variation with the Doppler frequency shift. On part of the curves, the opposite effect seems to take place. What is the explanation for this phenomenon?

Author's Reply

You are presumably referring to the times when the critical frequency approaches the operating frequency. This causes a moving of the reflection area resulting in a negative Doppler shift which in turn contaminates the measurement of the vertical drift due to current changes (i.e. magnetic field variation). A correlation, this, cannot be expected during this time period.

J. Aarons, US

F layer irregularities are of varying sizes, ranging from kilometer sizes dominating HF returns to 3 meter irregularities detected by the 50MHz backscatter radar. The term spread F irregularities is outmoded and not precise since they are large scale irregularities only as noted on ionograms. Even spread F irregularities are of two striking different types, range and frequency.

Author's Reply

According to the recent model of ionospheric irregularities proposed by Prof. H. Booker, one can assume that a statistical dependence exists between irregularities at large scales of some hundred kilometers and small scales of meters. This dependence can be understood as a decay of large scale into small scale ionosization irregularities (model of plasma turbulence), where the large scales are determined by typical scales of atmospheric gravity waves. This decay process of plasma turbulence is comparable to the decay of atmospheric turbulence described by the Kolmogoroe spectrum. We thus really observe the same phenomenon, which either may be called equational spread-F (for historical reasons) or equational F-region irregularities. However, the observation at HF and UHF are sensitive to different spatial scales of the irregularities spectrum. This even may hold for the different types of range and frequency-spread-F.

T.B. Jones, UK

Did you find that the Δf and ΔH disturbances were always in phase? There is evidence from temperature and high latitude Doppler experiments carried out by the Leicester group that time delays can occur between the Δf and ΔH disturbances.

Author's Reply

Due to the strong frequency fluctuations control by F-region irregularities such a phase shift cannot significantly be addressed for the presented example. It has to be checked with further data if a phase shift does also occur in our experiments.

Klaus Bibl, US

We hope that the analysis of about one year of digital ionogram data from the equatorial West Pacific station Kwajalein will resolve some of the discrepancies between measurements of VHF scintillation on satellite signals, of scientific radar and of analog ionosondes. Specifically, the opposite motion of background ioniation as seen in ionograms and of the plasma observed by the radars can be explained when the Doppler and arrival angle information on digital ionograms is considered. As seen on the figure (right side of page 4-4 of AGARD preprint) the lower ionogram presents the real amplitudes 4dB steps in the range/frequency display of the digital ionogram while in the upper ionogram the numbers representing Doppler and direction show the simultaneous upgrades motions of holes (electron density minima) and downwards motion of enhancement in addition to the horizontal motion of the average ionization. Previously analyses of the onset of the spread-F event agrees often very well in Dr Rottger's model.

THE PHENOMENOLOGY OF TRANSEQUATORIAL RADIO PROPAGATION UNDER SPREAD-F CONDITIONS

J. Röttger

Max-Planck-Institut für Aeronomie
3411 Katlenburg-Lindau 3, Fed. Rep. Germany

SUMMARY

Three special phenomena of transequatorial propagation are described: (1) The frequency dependence of off-great-circle path points to steep horizontal gradients of ionization, caused by the spatial resonance effect in the postsunset equatorial F-region. (2) The small-scale irregularities of the equatorial spread-F which are generated at the steep gradients give rise to strong amplitude fluctuations of transequatorial signals. (3) The upward rising depletions of ionization can duct and scatter VHF signals over distances of more than 6000 km.

1. INTRODUCTION

It was considered in former investigations (e.g. RÖTTGER, 1973) that the formation of equatorial spread-F irregularities which lead to considerable influence of transequatorial HF radio propagation takes place in three main phases. The purpose of this paper is to verify these assumptions by means of further results. The initial phase of the pre-midnight equatorial spread-F is the spatial resonance effect between plasma drifts and travelling ionospheric disturbances (BEER, 1973; RÖTTGER, 1978a; BOOKER, 1979) caused by atmospheric gravity waves. Due to this effect, ionospheric disturbances can be amplified so that steep electron density gradients are created (KLOSTERMEYER, 1978). In the second phase, electron density irregularities are generated at these gradients due to Rayleigh-Taylor instability. The depletions of ionization caused by the amplified disturbances can be accepted as source regions of rising equatorial spread-F bubbles (WOODMAN and LA HOZ, 1976) which in the third phase drift to high altitudes.

2. OBSERVATIONS OF QUASI-PERIODIC SPREAD-F PATCHES

During the developing phase of the equatorial spread-F often delayed echoes, so-called satellite traces (CLEMESHA and WRIGHT, 1966), are observed on ionograms which indicate that besides the normal vertical reflection also oblique reflections from electron density irregularities occur. This effect is known as range-spread-F. Since the critical frequency of the discrete satellite traces is observed to be not greater than the critical frequency at vertical incidence, it is assumed that these traces are caused by total reflection at sufficiently steep, large-scale electron density gradients rather than by scattering at small-scale electron density irregularities. The satellite traces often are observed to be periodical (Figure 1), which leads to the assumption that kinks in the electron density distribution vary periodically in space. As Figure 1 shows, these traces, which occur after the onset time of equatorial spread-F at about 1930 LT, change their position with time.

The moving kinks of large-scale equatorial spread-F irregularities also can be seen on Doppler records of DAVIES and CHANG (1968). They explained the observed structures by quasi-periodical kinks in the F-layer and reported that the kinks, which moved at a velocity between 100-200 m/s to the east, had a medium separation of 200-400 km. It can be noted from their records that during the lifetime of the observed traces the Doppler shift and the Doppler spread decrease.

Also results from transequatorial HF radio propagation indicate quasi-periodical structures in the occurrence of equatorial spread-F irregularities (RÖTTGER, 1973). It was reported that wave-trains of irregularity patches with a median wavelength of 380 km moved with a velocity of 110 m/s to the east, which is consistent with other observations (CLEMESHA and WRIGHT, 1966, DAVIES and CHANG, 1968; KELLEHER and SKINNER, 1971). It will be pointed out here that the developing phase of the observed transequatorial echoes supports the existence of strong horizontal electron density gradients during the commencement of the equatorial spread-F.

Considering the HF radio propagation measurements on the transequatorial path Lindau-Tsumeb (RÖTTGER, 1973), we obtain oblique-incidence sweep-frequency ionograms shown in Figure 2. On these records, the pulse propagation time t' is measured as a function of the transmitting frequency f. Figure 3a, which gives a simplified graphical display of the records (Figure 2), shows at 1908 UT the common echo traces characterizing the propagation conditions in the great-circle plane between transmitter and receiver. The corresponding propagation process is sketched in Figure 3b (see also Figure 1 of RÖTTGER, 1973). LI is the transmitting station Lindau and TS is the receiving station Tsumeb. Concerning the normal great-circle propagation between LI and TS, we measure the maximum usable frequency f'_o (commonly called MUF) and the corresponding propagation time t'_o, which are given by the ionospheric conditions along the great-circle path.

To investigate the frequency dependence of non-great-circle propagation, the transmitter antenna at Lindau was pointed about 30°W off the great-circle direction, and consequently a receiving antenna at Tsumeb was used pointing to the corresponding deviation angle at Tsumeb. This angle is slightly greater than the deviation angle at Lindau when considering that the side deflection takes place in the equatorial region. Examples of oblique-incidence ionograms obtained during this experiments are shown in Figure 2. By means of the graphical reproduction given in Figure 3a and the sketch of the propagation paths given in Figure 3b, we may understand the formation of the delayed echo traces shown on the record at the time t'_1 (Figure 2, 1912-1920 UT).

Due to the gradients of electron density in the ionosphere, a radio wave suffers a deflection from the original propagation direction. Gradients in the vertical plane lead to the well-known refraction or reflection of radio waves in the ionosphere. Correspondingly, gradients in the horizontal plane cause the ray path to be horizontally deflected, which commonly is known as azimuthal deviation of the ray path. By horizontal gradients (marked by arrows in Figure 3b) in the equatorial ionosphere (EQ in Figure 3b), a wave with frequency f_1 transmitted from LI at an angle $|\alpha| > 0$ will be deflected to the receiving station TS. This wave will reach EQ and propagate from there to TS if the following conditions are fulfilled:

$$f_1 \leq f'_o , \qquad (1)$$

where f'_o is the maximum usable frequency given by gradients in the vertical planes LI-EQ and EQ-TS. Considering for simplification comparable ionospheric conditions on the paths LI-EQ, EQ-TS and LI-TS, the maximum usable frequency f'_o is approximately the same on all three paths.

The next condition is stronger than condition (1), because this condition is necessary to provide the great-circle deviation:

$$f_1 \geq f'_1 , \qquad (2)$$

where f'_1 corresponds to the horizontal gradients in the EQ area. To fulfill both conditions (1) and (2), we obtain

$$f_1 \leq f'_1 \leq f'_o . \qquad (3)$$

Let us now consider the development of non-great-circle paths to demonstrate the realization of these conditions. The graphical display (Figure 3a) of the oblique-incidence records (Figure 2) indicates three different phases. At 1908 UT only normal great-circle propagation takes place, we observe three echo traces which probably indicate 3F, 4F and 5F propagation modes. In this first phase, the maximum usable frequency is f'_o and the corresponding propagation time is t'_o. Besides these great-circle paths additional echoes (1, 2 and 3) are observed in the second phase after 1912 UT which exhibit a larger propagation time t'_1 than the great-circle echoes. The maximum frequency f_{1max} of these signals is evidently lower than f'_o at 1912 UT.

Ray tracing calculations indicate that the long time delay of the signals observed around t'_1 excludes a propagation in the great-circle plane. The observed time delay $t'_1 - t'_o \approx 3$ ms strongly gives evidence that the delayed echoes are at least 20°W off the great circle. This fact incidentally should be expected due to the main lobes of the antennas pointing to west off the great circle.

Further evidence that these delayed signals propagated via non-great-circle paths is given by the following reason: As confirmed by several investigations (LYON et al., 1961, CLEMESHA and WRIGHT, 1966; SKINNER and KELLEHER, 1971), equatorial F-region irregularities are almost exclusively observed after sunset in F-region heights. The terminator crosses the F-region in that part of the equatorial zone (near the dip equator) which is seen at 20°W from the Lindau-Tsumeb path, at around 1910 UT. Since the delayed echoes on the record sequence of Figure 2 appear at about the same time 1910 UT, we consequently conclude that these echoes observed around 20°W are caused by postsunset ionization irregularities in the equatorial region.

The threefold split of the non-great-circle traces observed at about t'_1 is not due to multiple-hop propagation but is caused by three different irregularity areas, as demonstrated in the lower section of Figure 3b. This assumption is proved by ray tracing calculations which indicate that a delay of less than 1 ms between different hops is possible for non-great-circle paths at 20°-30°W. The maximum frequencies of the three off-great-circle paths have to be

$$f'_{11} > f'_{12} > f'_{13} \ (\text{and } t'_{11} < t'_{12} < t'_{13}) \qquad (4)$$

if only one irregularity area would cause the deviation. Since the records show that this condition does not hold, further support is given on the assumption of three irregularity areas causing these paths (see lower section of Figure 3b).

Concerning again the maximum frequency of the three non-great-circle traces, we conclude that steep horizontal gradients must be existent to get the paths deviated from the great-circle direction. Because

$$f'_1 \ (= f_{1max}) < f'_o , \qquad (5)$$

we conclude from condition (2), which is more decisive than condition (1), that the maximum frequency f_{1max} is controlled by the horizontal gradients in the irregularity areas (EQ).

As the horizontal gradients become steeper within a few minutes, we observe that $f_{1max} = f'_o$, which takes place at about 1916 UT. This observation does not exclude that $f'_1 > f'_o$. However, in this case condition (1) would be violated, thus, off-great-circle propagation is obviously not possible at frequencies $f_1 < f'_o$.

In order to explain the observed frequency dependence of side paths, it is roughly estimated that the perturbation amplitude has to be as strong as to produce horizontal gradients of the electron density which are comparable to the normal vertical gradients. This can be concluded from the record taken at 1912 UT, where the non-great-circle traces look like steep incidence vertical echoes, i.e. the entire layer has to be tilted off the horizontal plane up to about 30° (RÖTTGER, 1973) in order to provide the observed azimuth deflection of the ray path.

Steep horizontal electron density obviously has to be claimed to explain the frequency dependence of the non-great-circle paths observed at 1912-1916 UT. Following this initial perturbation phase, strong patchy-type irregularities are generated which still show a wave-like structure (RÖTTGER, 1973). The electron density variations in the small-scale irregularities (scale sizes up to some 100 m), which form large-scale irregularity patches (scale sizes of up to some 100 km), can reach some ten percent. This evidently can be concluded from amplitude distributions and spectra of HF waves propagating along non-great-circle paths (RÖTTGER, 1976).

It is interesting to note that the quasi-periodically occurring perturbation areas seem to come into existence almost simultaneously at different locations. Later on, the discrete traces deform into scattered and more or less non-resolved echoes caused by spread-F irregularity patches. This event is observed on the records around 1920 UT. The characterising feature, that the perturbation or irregularity areas simultaneously commence at separated locations, rather regularly is observed by means of transequatorial propagation and by direct radar observations in the equatorial region (KELLEHER and SKINNER, 1971).

As supposed by RÖTTGER (1973), these observations can be explained by a resonance mechanism of an ionization drift and gravity wave induced ionization perturbations (WHITEHEAD, 1971; BEER, 1973). Concerning the above mentioned observations, it follows that periodical horizontal variations of the electron density are precursors of periodically occurring patches of equatorial spread-F irregularities. The fact that perturbation areas in the F-region commence simultaneously at locations separated by a few hundred kilometers directs attention to periodical structures in the neutral or ionized atmosphere, which may trigger the generation of irregularities or will at least modulate the spatial distribution of irregularities.

3. HF SIGNAL AMPLITUDE DURING SPREAD-F CONDITIONS

In Section 2 the frequency dependence of transequatorial HF propagation during the commencing phase of spread-F was described. Before presenting in Section 4 new results which prove a considerable extension of the MUF due propagation via field-aligned ionization depletions, we will shortly outline the amplitude variations due to spread-F irregularities.

In Figure 4 examples of power or group delay intensity plots are shown which indicate the transition from normal propagation conditions in the late afternoon (1805 UT) to fully developed spread-F conditions in the evening (2024 UT). The plots are snapshots of 1-s intervals showing the short-term fading of pulsed signal amplitude on a fixed frequency. Around 1805 UT two traces (3F and 4F propagation along the great-circle path) are evident which show a rather low fading rate. Two hours later distinct traces can no more be identified since the propagation is strongly influenced by spread-F irregularities. The range of propagation time increased, which is explained by signals propagating along off-great-circle paths. It is also discernible that the fading rate is much faster. The amplitude coherence is a few hundred milliseconds only. Some further signal parameters of transequatorial HF radio propagation under spread-F conditions can be found in RÖTTGER (1976), where the coherence times as well as amplitude distributions and fading power spectra are treated in more detail.

4. MUF EXTENSIONS DUE TO SCATTERING AND DUCTING IN SPREAD-F BUBBLES

Observations of long-distance contacts on 144 MHz over more than 6000 km on transequatorial paths were reported recently by radio amateurs (TYNAN, 1978). After some stations in one hemisphere had heard 144 MHz signals from stations located several thousand kilometers apart in the other hemisphere, the first two-way transequatorial contacts in the 144-MHz VHF band were made at the end of October 1977 between several stations in Argentina and a station in Caracas over a distance of more than 5000 km. These were followed by contacts over almost 6400 km between stations in Puerto Rico and other stations in Argentina. In February 1978 contacts between stations in Darwin (Northern Australia) and Japan took place. A continuous series of contacts between Rhodesia in the southern and Cyprus in the northern hemisphere started in April 1978. An instructive picture of the continuity of these openings is given by the observations shown in Figure 5. These observations indicate maximum occurrence of transequatorial propagation on frequencies larger than the classical MUF between 1700 and 1900 GMT, which is 1900-2100 local time at the stations involved.

The general features of this type of propagation, as observed until now, can be summarized as follows: (1) highest probability and maximum signal strength is observed during the hours after sunset near the radio path; (2) north-south paths with the station locations approximately symmetrical to the earth's magnetic equator seem to be preferred; and (3) signals very frequently indicate strong flutter fading connected with perceptible frequency modulation comparable to those features of auroral backscatter.

It is known that transequatorial contacts on 50 MHz can be established regularly during years of reasonably high sunspot numbers. These contacts become possible due to the so-called supermodes, which are explained by a double reflection (better, perhaps, to say deflection, Figure 6) in the equatorial ionosphere showing large enhancements of the critical frequency north and south of the earth's magnetic equator. This enhancement is called "equatorial anomaly of the F-region". The supermodes (see dashed-dotted line in Figure 6) due to this anomaly are sometimes covered by the influence of strong spread-F irregularities. Since it is well known that the critical frequencies in the equatorial anomaly of the F-region do no reach values which allow frequencies greater than about 50 MHz to be reflected at oblique incidence, a new explanation was offered by RÖTTGER (1978b) for the well-established transequatorial contacts in the higher VHF band.

Observations with the high-power 50-MHz radar near Lima, Peru (WOODMAN and LA HOZ, 1976), transequatorial propagation experiments in the HF band (RÖTTGER, 1973; DIEMINGER and RÖTTGER, 1975), as well as observations with orbiting satellites (McCLURE et al., 1977) showed that in the night-time equatorial ionosphere rather turbulent fluctuations of irregularities exist which are connected with large depletions or bite-outs of ionization. These depletions in the ionosphere, which are generated at an altitude of about 400 km, can drift upwards in the ionosphere. It is observed that these bubbles, which have dimensions of some 10 to 100 km, occur in periodical patches. One may assume that this periodicity is due to atmospheric gravity waves caused by severe tropical thunderstorms (RÖTTGER, 1977). A further treatment of the connection between spread-F irregularities and thunderstorm activity in the tropics will be published elsewhere. Besides drifting upwards, these patches of bubbles and the corresponding spread-F irregularities also move in west-east direction. One can thus expect several bubbles to cross a transequatorial radio path during an evening at time intervals of 20-60 min. The ionization or plasma bubbles, which most frequently occur about 1-4 h after sunset at the equator, can even rise to altitudes larger than 1000 km. At the boundaries and in the wake of these rising bubbles, ionization irregularities occur which scatter HF and VHF radio signals.

This scattering process is comparable to aurora backscatter from field-aligned, almost vertically extended, irregularities. Because of the horizontal direction of the earth's magnetic field at the equator, the ionization irregularities causing transequatorial propagation, however, are almost horizontally extended. In the case of transequatorial propagation these high-altitude spread-F irregularities cause forward scattering (see dashed line in Figure 6). One may also regard transequatorial ducting of VHF signals in field-aligned tubes of ionization bite-outs. The term "scattering" used in this paper expresses a possible super-position of ducting or quasi-reflection at the bubble boundaries and pure forward scattering. More detailed calculations of this propagation phenomenon are given in the paper by HERON (1979).

It is noted that ray bending in the lower ionosphere along the paths to the bubbles can be neglected on high VHF (e.g. f_2 = 144 MHz in Figure 6), which means that these signals propagate approximately on a straight line from the transmitter to the scattering bubble and from there to the receiver. The lower the frequency, the stronger is the ray path deflection of radio signals in the ionosphere. This well-known phenomenon causes the ray paths of HF signals to be deflected or bent so strongly that they again reach the earth's surface. At frequencies larger than the MUF (maximum usable frequency for a given radio path) the ray is still somewhat deflected in the ionosphere but can no longer reach the earth's surface. On 50 MHz this ray bending still has to be taken into account, which can cause these signals not to reach the high-altitude irregularities in the bubbles (dotted line in Figure 6). This may be the reason for some reports indicating that 50-MHz transequatorial propagation was not observed during 144-MHz contacts.

During times of low critical frequencies of the ionosphere, the ray bending is not too strong. Thus 50-MHz signals can also reach high-rising spread-F bubbles so that long-distance transequatorial propagation via scattering can also take place on 50 MHz. The 50-MHz transequatorial signals then gain strong flutter fading. It is estimated that the bubble irregularities should also be capable of scattering frequencies even higher than 144 MHz, which in fact already appears to have been observed in the 430-MHz amateur radio band. Due to the moving bubble and the accompanying turbulent irregularities, the scattered signals gain strong amplitude and frequency distortions, which are known as severe flutter fading and frequency spread. Since bubbles can move in periodical patches across the transequatorial propagation path, one expects the signal to indicate long-term fading of 20-60 min (Figure 5). These bubbles commonly reach heights around 700 km, so that signals on high VHF were often observed over distances of 6000 km. Since the maximum altitude at which irregularities connected with these bubbles were observed exceeded 1000 km, and taking into account some slight ray bending in the F-layer, one should expect high VHF transequatorial propagation over distances up to 10000 km.

5. CONCLUSION

Spread-F irregularities occurring regularly after sunset in the equatorial ionosphere have a considerable influence on HF and VHF radio propagation. It was shown in this paper that strong side reflections due to horizontal gradients of electron density occur. These large-scale gradients, which are caused by the spatial resonance effect, are precursors of small-scale spread-F irregularities and upward rising depletions of ionization. The small-scale irregularities cause intense amplitude fluctuations of HF and VHF signals propagating on transequatorial paths. It is shown that the rising ionization depletions can give rise to forward scattering and ducting of VHF signals (up to 150 MHz) which can propagate over transequatorial distances of more than 6000 km.

REFERENCES

BEER, T., 1973, "Spatial resonance in the ionosphere", Planet. Space Sci. 21, 297-307.

BOOKER, H.G., 1979, "The role of acoustic gravity waves in the generation of spread F and ionospheric scintillation", J. Atmos. Terr. Phys. 41, 501-515.

CLEMESHA, B.R. and R.W.H. WRIGHT, 1966, "A survey of equatorial spread-F", AGARDograph 95, 3-27.

DAVIES, K. and N.J.F. CHANG, 1968, "Radio-Doppler observations of the ionosphere near the magnetic equator", AGARD-CP-37, No. 2, paper 52.

DIEMINGER, W. and J. RÖTTGER, 1975, "Transäquatoriale Kurzwellenfunkverbindungen", cq-DL, February 1975, pp. 84-86; March 1975, pp. 145-147.

HERON, M.L., 1979, "Transequatorial propagation through equatorial plasma bubbles - discrete events", AGARD-CP-263, paper 24.

KELLEHER, R.F. and N.J. SKINNER, 1971, "Studies of F region irregularities at Nairobi; II - By direct backscatter at 27.8 MHz", Ann. Géophys. 27, 195-200.

KLOSTERMEYER, J., 1978, "Nonlinear investigation of the spatial resonance effect in the nighttime equatorial F region", J. Geophys. Res. 83, 3753-3760.

LYON, A.J., N.J. SKINNER and R.W.H. WRIGHT, 1961, "Equatorial spread-F at Ibadan, Nigeria", J. Atmos. Terr. Phys. 21, 100-119.

McCLURE, J.P., W.B. HANSON and J.H. HOFFMAN, 1977, "Plasma bubbles and irregularities in the equatorial ionosphere", J. Geophys. Res. 82, 2650-2656.

RÖTTGER, J., 1973, "Some properties of large-scale equatorial spread-F irregularities interpreted by influences of atmospheric gravity waves", Ztschr. Geophys. 39, 799-818.

RÖTTGER, J., 1976, "Influence of spread-F on HF radio systems", AGARD-CP-173, paper 26.

RÖTTGER, J., 1977, "Travelling disturbances in the equatorial ionosphere and their association with penetrative cumulus convection", J. Atmos. Terr. Phys. 39, 987-998.

RÖTTGER, J., 1978a, "Drifting patches of equatorial spread-F irregularities - experimental support for the spatial resonance mechanism in the ionosphere", J. Atmos. Terr. Phys. 40, 1103-1112.

RÖTTGER, J., 1978b, "Transequatorial dx contacts on 144 MHz", Radio Comm., December 1978, pp. 1028-1029.

SKINNER, N.J. and R.F. KELLEHER, 1971, "Studies of F region irregularities at Nairobi; I - From spread F on ionograms 1964-1970", Ann. Géophys. 27, 181-194.

TYNAN, W.A., 1978, "The world above 50 MHz", QST, January 1978, pp. 73-74.

WHITEHEAD, J.D., 1971, "Ionization disturbances caused by gravity waves in the presence of an electrostatic field and background wind", J. Geophys. Res. 76, 238-241.

WOODMAN, R.F. and C. LA HOZ, 1976, "Radar observations of F region equatorial irregularities", J. Geophys. Res. 81, 5447-5466.

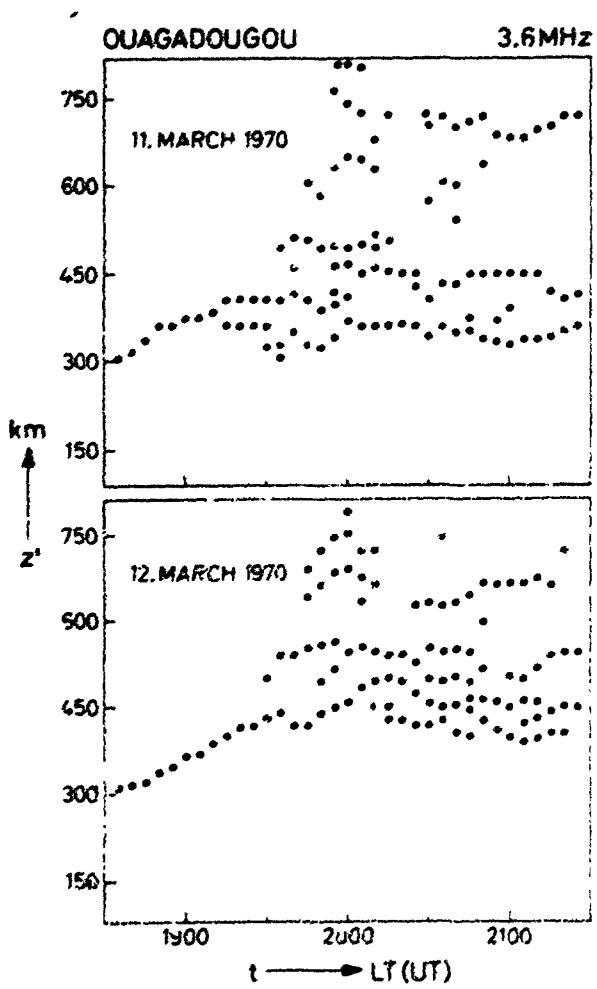


Fig. 1 Virtual height of ionosonde echoes on 3.6 MHz during spread-F conditions.

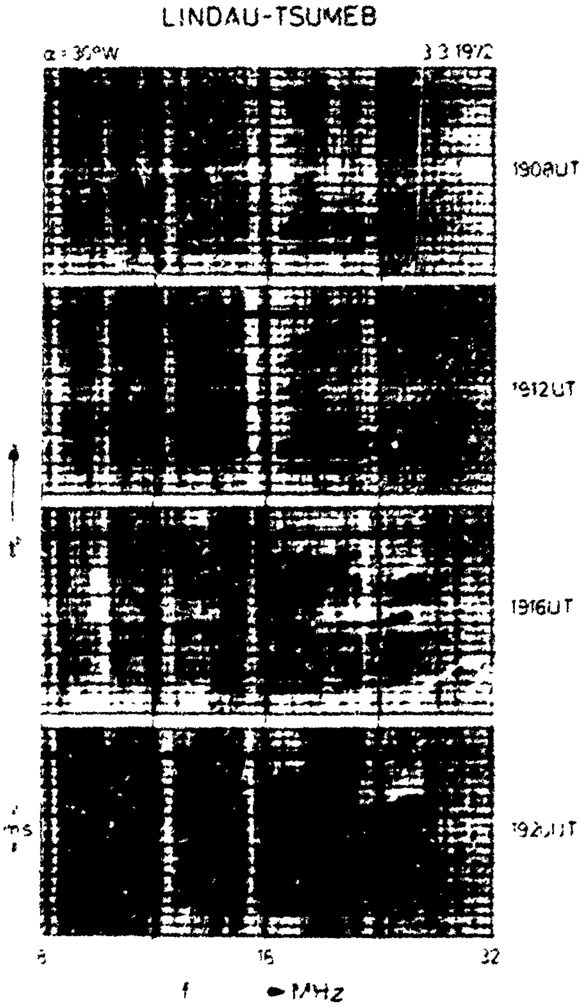


Fig. 2 Oblique-incidence ionograms recorded on the path Lindau-Tsumeb at an off-great-circle angle α = 30°W. The traces show the transition from normal (1908 UT) to spread-F (1920 UT) conditions.

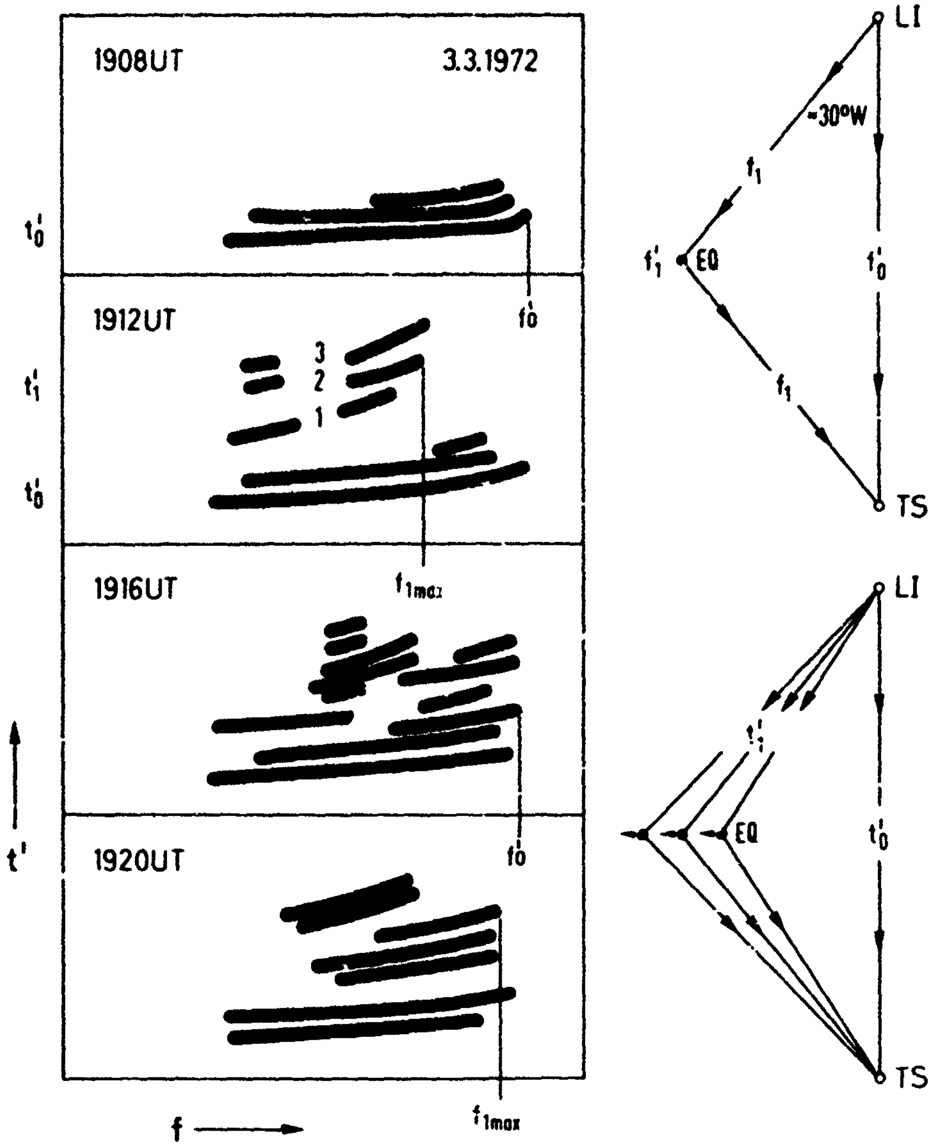

1908UT
3.3.1972
t'0
f'0
1912UT
3
2
1
t'1
t'0
1916UT
f1max
t'
1920UT
f'0
f1max
f
LI
≈30°W
f1
f'1
EQ
t'0
f1
TS
LI
f'1
EQ
t'0
TS

Fig. 3 Scheme of ionogram traces taken from Figure 2, and schematic view of the transequatorial path Lindau (LI) - Tsumeb (TS).

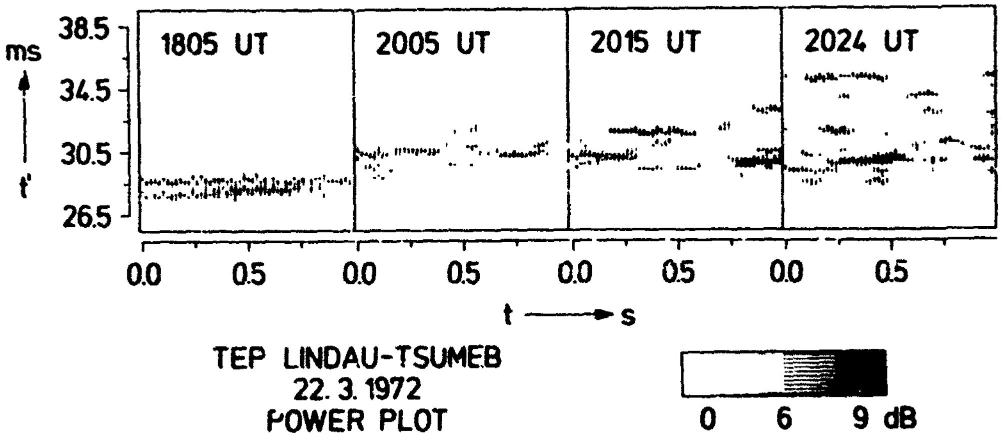


Fig. 4 Group-delay intensity plots recorded on 14.7 MHz on the transequatorial radio path Lindau-Tsumeb (7915 km). These records show the typical delay spread and fast fading of TEP signals influenced by equatorial spread-F irregularities.

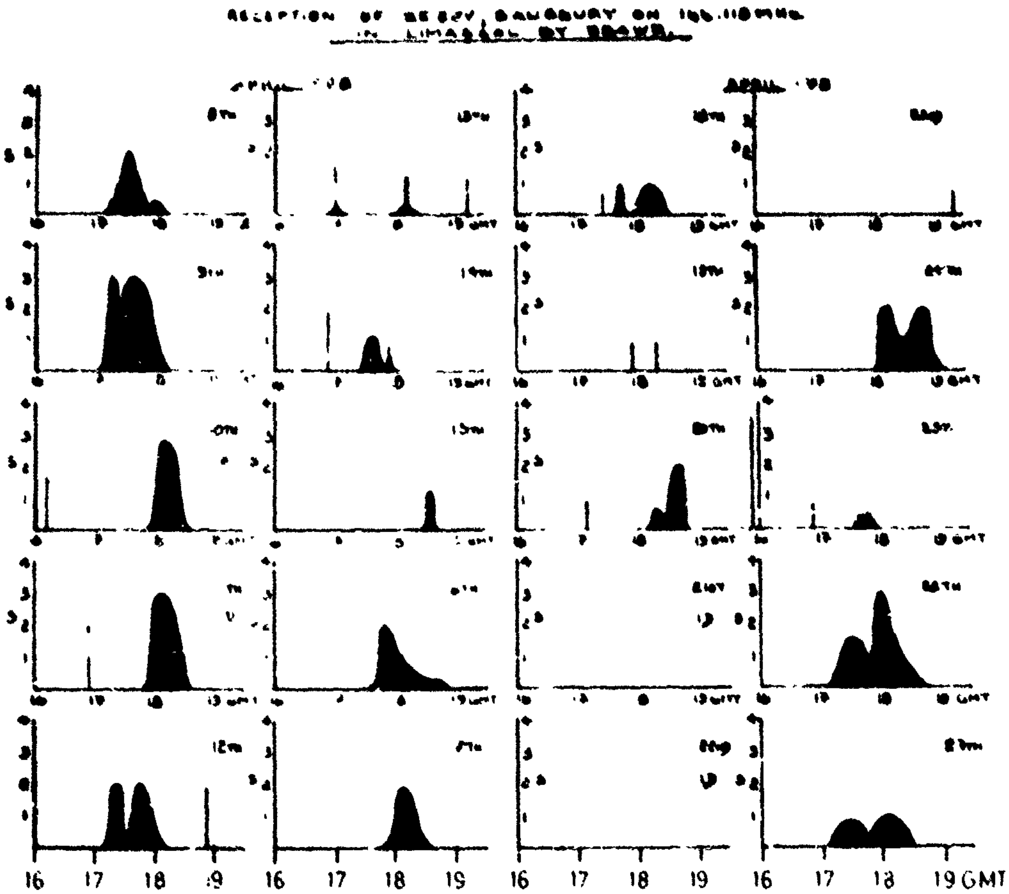


Fig. 5 Relative field strength of Salisbury beacon on 144,118 MHz, recorded in Cyprus during April 1978. The beacon transmitter had 50 W RF and an 11-el Yagi. A 10-el Yagi and a receiver with noise figure of 1,8 dB was used for reception. (From RÖTTGER, 1978b.)

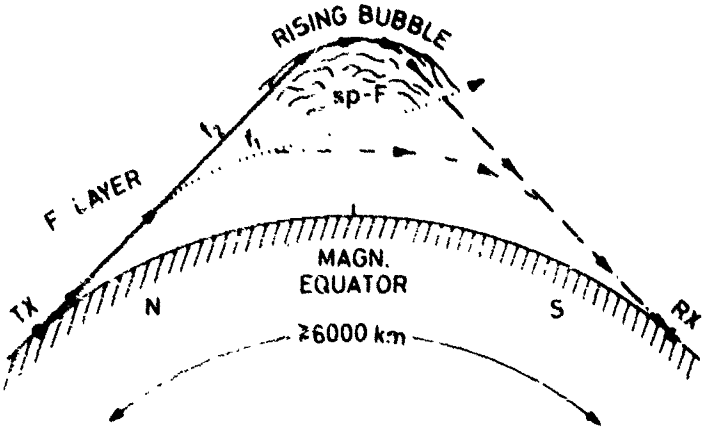


Fig. 6 Scheme of transequatorial propagation on low (f_1 = 50 MHz) and high (f_2 = 145 MHz) frequency. During the postsunset hours, spread-F irregularities (sp-F) rather regularly occur in the ionospheric F-layer near the magnetic equator. Characteristic features of the equatorial spread-F phenomenon are plasma bubbles which rise to altitudes of about 400-1000 km. The accompanying high-altitude ionospheric irregularities scatter and duct VHF signals so that TE propagation over distances of more than 6000 km becomes possible (dashed line). Low VHF signals are deflected north and south of the magnetic equator. Mainly during daytime hours this can give reason for transequatorial propagation ("supermode" indicated by dashed-dotted line). The deflection, on the other hand, can cause the low VHF signals not to reach the high-altitude irregularities (dotted line), which explains the observed difference of postsunset transequatorial propagation on low VHF and high VHF.

DISCUSSION

Jules Aarons, US

Have you validated your hypothesis of thunderstorm related bubble initiation on an individual (day to day) basis?

Author's Reply

One can assume that some triggering effect of equatorial F-region irregularities does exist due to gravity waves which are generated by thunderstorms. Except of proving this by means of statistics it appears to be very difficult to do case studies in tracing gravity waves between the source in the troposphere and the hemisphere. One has to take into account that thunderstorm systems in a rather large area may contribute to the gravity wave spectrum in the thermosphere. A rathe. extended observation network therefore, would be necessary to prove a day-to-day variability which at this time is not available.

DETECTION, RANGING AND DRIFTSPEED MEASUREMENTS OF EQUATORIAL IONOSPHERIC IRREGULARITIES BY MEANS OF AIRGLOW OBSERVATIONS

G.Lange-Hesse
H. Lauche
Max-Planck-Institut für Aeronomie
Postfach 20
D-3411 Katlenburg-Lindau 3
Federal Republic of Germany

SUMMARY

Measurements of the emission rate distribution of the red oxygen line λ = 630 nm (= 6300 Å) in the night airglow with a panorama photometer at Tsumeb, Namibia (former called South West Africa) have shown the occurrence of distinct airglow irregularities in the equatorial region. These irregularities are accociated with irregularities in the electron density of the ionospheric F2-layer (also called spread F) and show a nearly constant eastward drift with a average speed of about 100 m/s. The drift to the east is superimposed by an irregular movement to the south and north with a speed of about 200 m/s and more.

These equatorial electron density irregularities were identified by RÖTTGER, 1973 by means of transequatorial H.F. propagation from Germany to Namibia. The calculation of the eastward drift speed from this radio method has led to the same value as that deduced from airglow observations. However, the radio method does not permit information to be obtained about the north-south drift of the irregularities.

INTRODUCTION

The emission rate distribution of the red oxygen line λ = 630 nm (= 6300 Å) in the night airglow has been measured with a panorama photometer at Tsumeb, Namibia (former called South West Africa) from 1971 to 1973. Fig. 1 shows such an emission rate distribution for June 11, 1972, 1915 UT (about 2025 local time). The observation point is located at the center of the diagram. The photometer scans the sky at 9 different elevation angles, each rotating with 360° in azimuth. The elevation angles are destinated in such a way that the sky is scanned in equidistant circles with scarcely 200 km horizontal distance between them at about 250 km height. The outer circle in the Figure therefore has a radius of scarcely 1600 km. In the representation is north at the top and east to the right. The emission rate is proportional to the length of each dash. The calibration and the distance scale is to the right.

The Figure shows a pronounced increase of the emission rate to the north of Tsumeb at a distance of about 1500 km. This increase is the so called "tropical red arc" which was first observed by BARBIER, 1960 at Tamanrasset in the South of Algeria. The arc is extended along the geomagnetic parallels of latitude and occurs on both sides of the geomagnetic equator at a geomagnetic latitude of about 15-20° N and S. The southern arc therefore is still located in the range of the panorama photometer at Tsumeb (geomagnetic and geographic coordinates of Tsumeb see text Fig. 1). According to Fig. 1 the emission rate of the tropical red arc is about ten times higher than the emission rate over Tsumeb. Both arcs fit together with the so called "APPLETON anomaly" in the electron density of the F2 layer. This is a pronounced increase in the electron concentration at about 300 km height of the ionosphere. Both optical arcs and the APPLETON anomalies on the northern and southern Hemisphere are separated by a stripe along the geomagnetic equator with pronounced lower emission rate and lower electron density respectively.

The APPLETON anomaly represents an accumulation of electrons which have been transported by the HALL Effect of the equatorial ringcurrent in the geomagnetic field to the upper levels of the F-layer. From there they have drifted by the influence of gravity forces along the lines of force of the geomagnetic field to the north and south to lower heights and cause an increase of the electron concentration: the APPLETON anomaly. This increase there leads to an increase of the emission rate caused by enhanced dissociative recombination of O_2^+ according to the equation

$$O_2^+ + e \rightarrow O + O' + 6.96\ \mathrm{eV}$$

O_2^+ is generated at these heights by charge exchange with O according to

$$O_2 + O^+ \rightarrow O_2^+ + O$$

The energy of 6.96 eV released by the dissociative recombination is sufficient to excite the $O(^1D)$ level of atomic oxygen. This state leads to the excitation of the red line (630 nm = 6300 Å).

RESULTS OF THE OBSERVATIONS

The following Figures 2 to 12 show the development and variation of the tropical red arc in the first three hours after sunset. Fig. 2 shows the arc at 1945 UT. Compared with Fig. 1 the emission rate has decreased a little. The state of development 15 minutes later is shown in Fig. 3 (2000 UT). The emission rate maximum now has shifted to the south from the outer circle to the next inner circle and the extension in east-west direction is restricted, that means, that an irregularity in the emission rate has been generated. The following Figures will show that this irregularity drifts to the east.

Fig. 4 shows the start of the generation of a second irregularity to the left of the first. In the following Figures 5 and 6 this second one is already well developed. Fig. 7 —15 minutes later— clearly shows the drift of the irregularity to the east and a shift back to the north to the outer circle. The movement to the north over a distance of about 200 km —the radial distance between two circles— takes place within a maximum time intervall of 15 minutes— time between two measurements. This corresponds to a minimum north south speed of about 220 m/s.

In the following Figures 8 (2115 UT) to 12 (2215 UT) the continuation of the east west drift of the irregularities and an irregular alternating movement between the outer and the next inner circle can be seen. The calculation of the east west drift velocity by consecutive measurements gives a value of 91 m/s. This is less than the velocity in north south direction.

The airglow irregularities shown in Fugures 2 to 12 are associated with irregularities in the electron density in the ionospheric F2-layer (also called spred F) which give rise to radio wave propagation interference. These electron density irregularities in the APPLETON anomaly were identified by RÖTTGER (1973 and in this volume) by means of transequatorial H.F. propagation from Germany to Namibia. The calculation of the eastward drift speed from these observation has led to an average value of about 100 m/s and is of the same order as the speed deduced from optical observations.

From the transequatorial HF propagation information about the speed of the eastward drift can be received only. However, the radio method does not permit information to be obtained about the north south drift of the irregularities.

According to calculations of RÖTTGER 1973 the generation of these irregularities are released by spacial resonance effects between atmospheric gravity waves and plasma drift in the equatorial region. The largescale occurrence of these irregularities therefore is quasi periodical as can be seen from the Figures.

As shown in Fig. 1 to 3 the tropical red arc occurs in the early evening hours. This is in accordance with satellite observations with OGO 4 and 6 (REED, E.I. et al. 1973; THUILLIER, G. and BLAMONT, J.E. 1973). The OGO 4 takes 98 minutes for one orbit. The fine structure in the variation of the location and the emission rate of the arc revealed by the panorama photometer with 15 minutes observation sequence therefore can not be detected by the satellite. The technical construction of the panorama photometer shown here allows the production of emission rate distributions as shown in Figures 1 to 12 in 4 minutes intervalls. The results shown here therefore represents a good exemple that airglow observations by satellites can be supplemented very succesfully by ground based photometer observations.

REFERENCES

Barbier, D. Glaume, J.	1960	Ann. Geophys. 16, 319
Reed, E.I. Fowler, W.B. Blamont, J.E.	1973	J. Geophys. Res. 78, 5658
Röttger, J.	1973	Zeitschrift für Geophys. 39, 799
Thuillier, G. Blamont, J.E.	1973	Physics and Chemistry of Upper Atmospheres, edited by B.M. McCormac p. 219, Dordrecht-Holland/Boston-USA

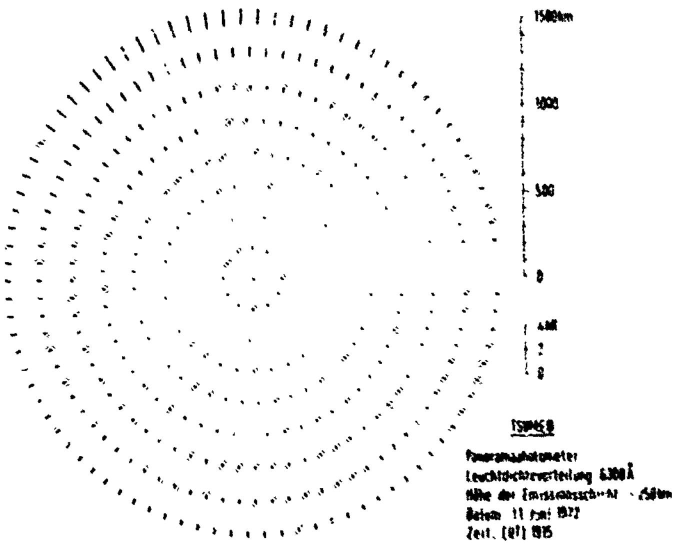


Fig. 1. Emission rate distribution of the red oxygen line over Tsumeb on June 11, 1972, 1915 UT. The observation point is located at the center of the diagram. The photometer scans the sky at 9 different elevation angles, each rotating with 360° in azimuth. The elevation angles are destinated in such a way that the sky is scanned in equidistant circles with scarcely 200 km horizontal distance between them at about 250 km height. The outer circle in the Figure therefore has a radius of scarcely 1600 km. In the representation is north at the top and east to the right. The emission rate is proportional to the length of each dash. The calibration and the distance scale is to the right. The geographic coordinates of Tsumeb are:

$\varphi = 19°14'S$, $\lambda = 17°43'E$; and the geomagnetic:

$\Phi = 18°11'S$, $\Lambda = 82°48'E$.

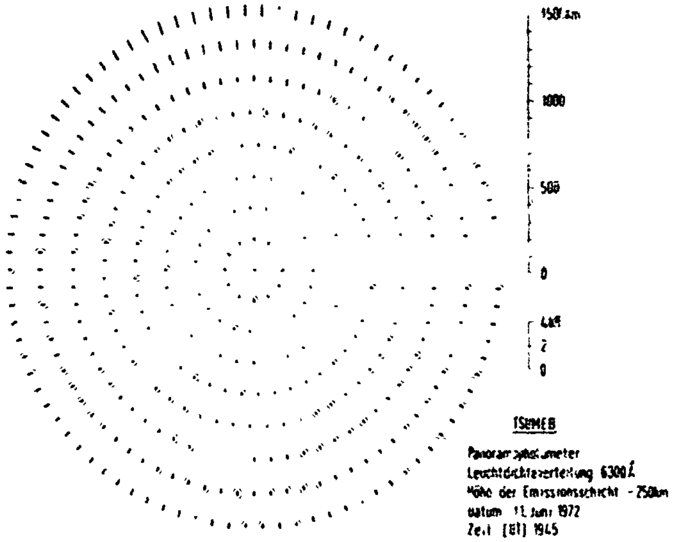


Fig. 2 Similar representation as in Fig. 1 but at 1945 UT the same day

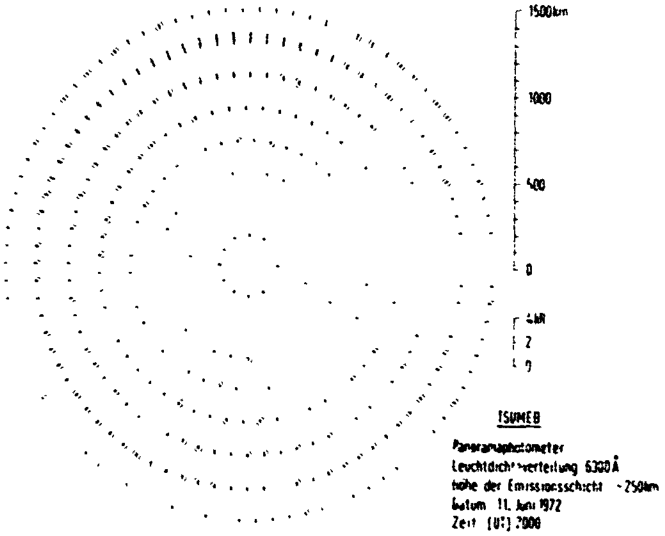


Fig. 3.Similar representation as in Fig. 1 but at 2000 UT the same day

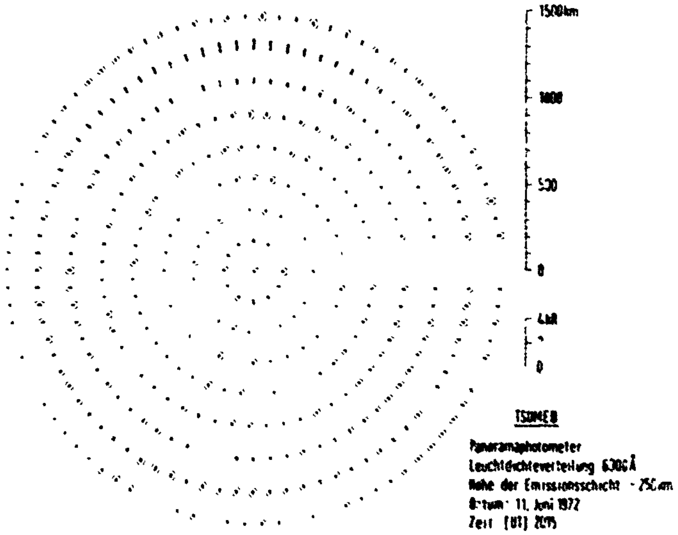


Fig. 4.Similar representation as in Fig. 1 but at 2015 UT the same day

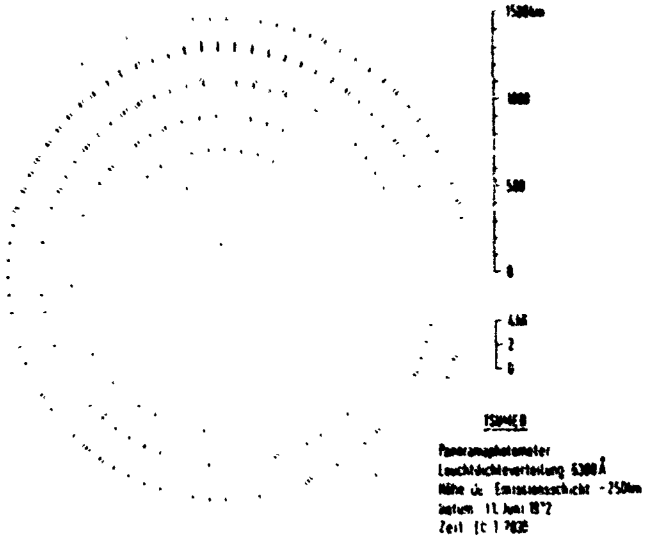


Fig. 5.Similar representation as in Fig. 1 but at 2030 UT the same day

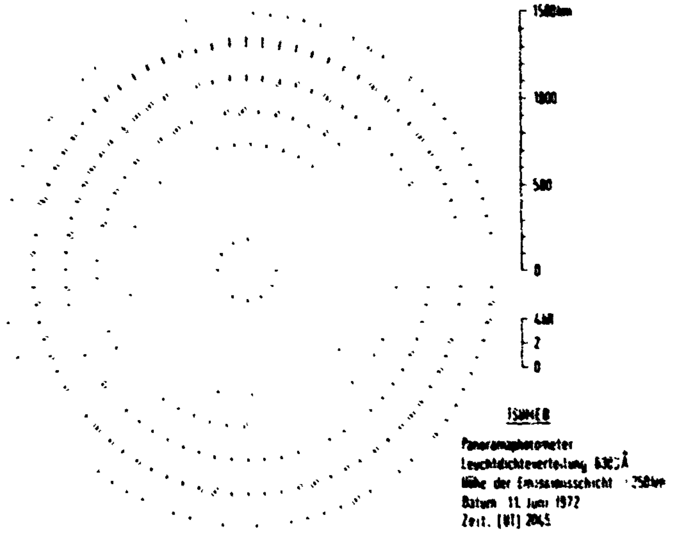


Fig. 6.Similar representation as in Fig. 1 but at 2045 UT the same day

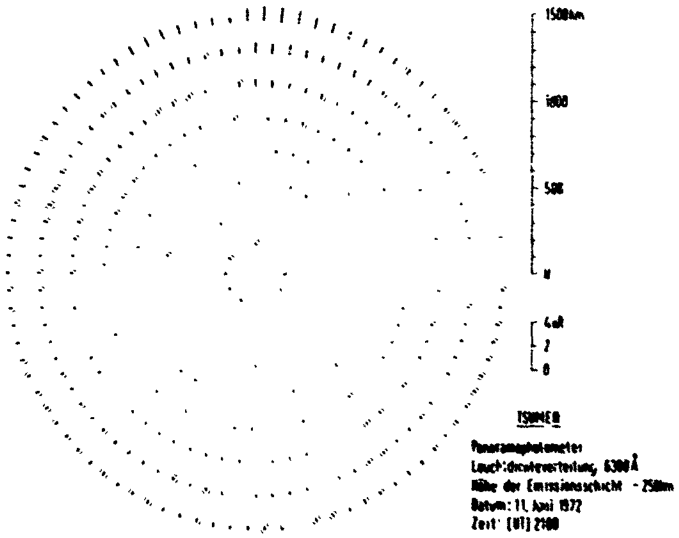


Fig. 7

Similar representation as in Fig. 1 but at 2100 UT the same day

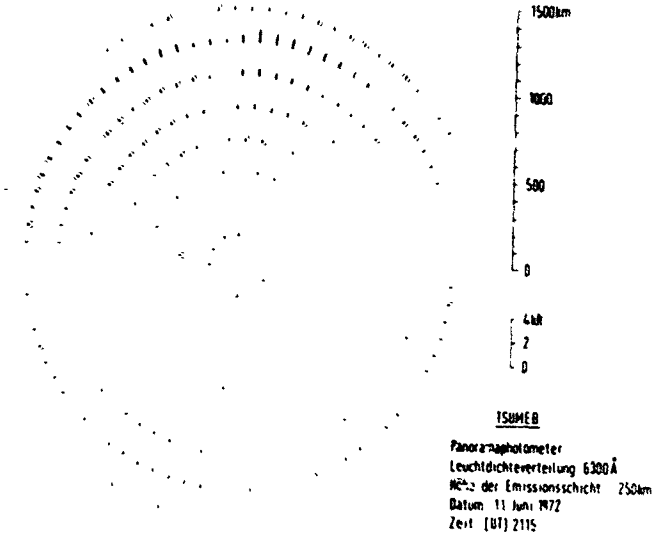


Fig. 8. Similar representation as in Fig. 1 but at 2115 UT the same day

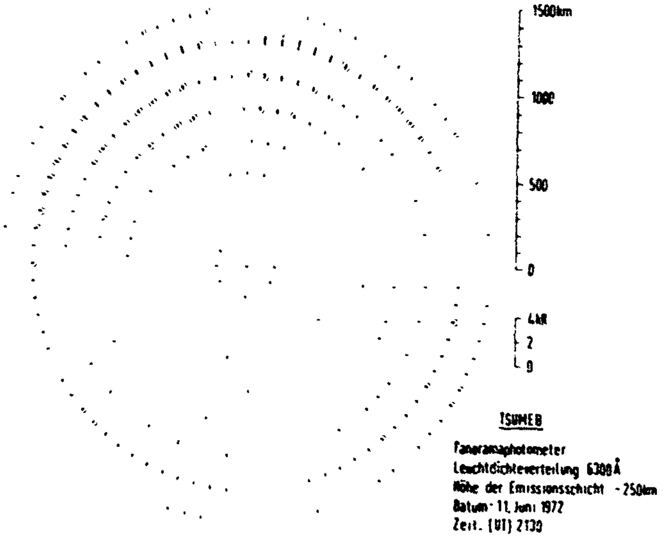


Fig. 9. Similar representation as in Fig. 1 but at 2130 UT the same day

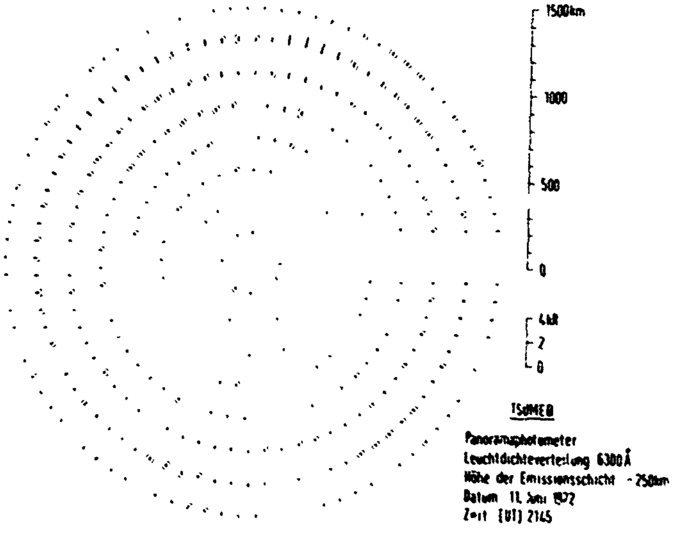


Fig. 10

Similar representation as in Fig. 1 but at 2145 UT the same day

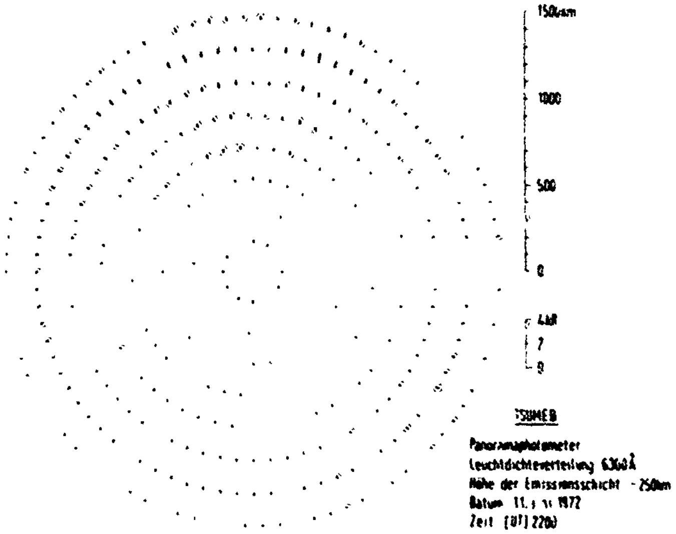


Fig.11.Similar representation as in Fig. 1 but at 2200 UT the same day

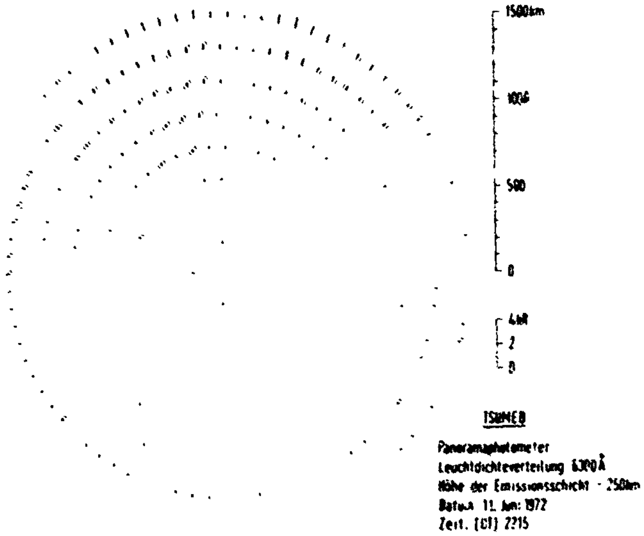


Fig.12.Similar representation as in Fig. 1 but at 2215 UT the same day

HF WAVEFRONT IRREGULARITIES OBSERVED ON A LARGE APERTURE RECEIVING ARRAY

T.B. Jones and E.C. Thomas
Department of Physics
University of Leicester
Leicester, U.K.

SUMMARY

In recent years considerable attention has been given to improving HF system performance by the use of large multi-element arrays which have adaptive capabilities. Such arrays normally require the phase front of the received signal to be linear across the array. There is evidence which suggests that the ionosphere produces distortions in the linearity of the reflected wavefront and this imposes a fundamental limitation on the antenna performance.

A new receiving array has been constructed consisting of seven elements distributed along the arms of an L in a minimum redundancy configuration. The maximum dimension, across the ends of the arms is 1.4 km. The phase and amplitude at each element can be measured simultaneously every 0.1 sec thus enabling rapid changes in the wavefront to be investigated. The paper describes some of the preliminary results obtained with the array for both ground and sky waves. The wavefront curvature can be determined for the ground wave case and hence the location of transmitter derived. For sky waves, wavefront irregularities frequently occur even during conditions when propagation appears to be single moded. The spatial and temporal extent of these irregularities is investigated since their presence is likely to be a major limiting factor in the performance of multi-element adaptive arrays.

1. INTRODUCTION

The optimum performance of a large antenna array is frequently not achieved because of non-linearities in the phase front of the received radio signal. In addition to the phase deviations, amplitude fluctuations in the received wavefront can also occur. The limitations in performance imposed by wavefront non-linearities have long been recognised and these effects are discussed by several authors, e.g. Gething (1978), Rice (1975). In the high frequency (HF) band, particular attention has been given to the errors observed in the bearing of the received signals when large aperture direction finders are employed (Hayden, E.C., 1961: Jones and Reynolds, 1975).

Irregularities in the received wavefront can be produced by a number of mechanisms. When the signal is received via two independent propagation paths, destructive interference can occur at points along the antenna array. The interference pattern will vary in both space and time because the various propagation paths involved are usually time varying. An alternative mechanism involves the reception of the signal from a rough reflector (e.g. the ionosphere) when a large number of near equivalent rays are incident on the receiving antenna. In this case a "diffraction pattern" is formed on the array and corresponding changes in phase and amplitude are observed across the array (Booker et al, 1950). Since the reflecting medium is usually time varying, the diffraction pattern is also time varying as is the apparent wavefront measured by the antenna array.

This paper reports on an investigation of the wavefronts of HF signals received on a large receiving array for a range of propagation conditions. The array consists of seven vertical elements spaced in a minimum redundancy configuration. The phase and amplitude on each element are measured and the whole array scanned every 0.1 sec. Signals received via (a) ground waves, and (b) sky waves, have been investigated. The curvature of the wave front of the ground wave signals enables the location of the transmitter to be determined. The sky wave signals exhibit a wide variety of disturbance features and an attempt is made to relate these to the corresponding reflection conditions in the ionosphere.

2. EXPERIMENTAL ARRANGEMENT

The antenna array consists of seven elevated feed vertical monoples arranged along the arms of an L as indicated in Figure 1. The spacings along each arm of the array are in the ratio 1 : 3 : 2, with each unit corresponding to 200 meters. This allows the maximum number of element separations to be achieved for a given number of elements in the array. The signals from each element are brought to the central laboratory via equal length feeders to preserve the phase information and are then fed to a seven channel phase matched receiver. This arrangement is shown in Figure 2 and consists of seven Racal RA 17 HF receivers which have common mixer signals derived from highly stable synthesised sources. Suitable selection of these mixer frequencies provides tuning for all the receivers. An eighth similar channel is provided and supplied with a synthesised RF reference signal at a frequency 2 Hz lower than the required frequency. This down-converted reference signal is mixed with the IF outputs of the other seven receivers, in quadrature phase detectors. The arrangement effectively provides down-conversion of the RF signals in the seven channels to 2 Hz. The in-phase and quadrature outputs of the phase detectors are sampled by an A-D converter, the complete arrangement being scanned ten times a second. A PDP 8 mini-computer converts this information to amplitude and phase values to allow either analysis in real time or storage on magnetic tape for later processing.

The receiver in each channel of the system introduces an arbitrary phase shift into the signal and in addition imposes a different amplitude response. A calibration facility is therefore required to determine the magnitude of these effects. This is achieved by switching the receiver inputs to a reference signal at the required input frequency such that the phase of the reference is the same at each receiver input. The amplitude of this calibration signal is varied by means of a programmable attenuator. The PDP 8 controls the calibration procedure and stores the data. This allows correction of the phase and amplitude information subsequently recorded. The analysis facility can produce graphical representation of the phase and amplitude variation on each element displayed as a function of time

3. OBSERVATIONS OF GROUND WAVE SIGNALS

To test the correct operation of the phase and amplitude measuring system, observations were undertaken on ground wave signals which are assumed not to vary with time. Measurements were undertaken on two BBC transmitters of frequency 1.34 and 0.693 MHz situated at distances of 11.5 and 75.5 km respectively from the receiving array. A sample of the results obtained is reproduced in Figure 3, and clearly shows the characteristically constant phase and amplitude recorded on each of the seven antenna elements. Similar results were obtained for other ground wave signals.

The ability to record the phase at three or more known points in the ground wave phase front, intrinsically suggests that it should be possible to determine the position of the transmitter from a knowledge of the wave front curvature. This situation is illustrated in Figure 4 where it is evident that the difference in phase between the points AB, BC and AC depend on their distance from the transmitter. The phase difference between elements of the array can be measured to an accuracy of 3 degrees and this, together with the maximum spacing available and the wavelength of the received signal, limits the accuracy to which the position of the transmitter can be determined. The higher the signal frequency the greater accuracy for a given size of array aperture and phase resolution. Figure 5 illustrates the increase in maximum range with frequency at which a transmitter can be located for an aperture of 1 km and a phase resolution of 3 degrees. It should also be noted that the ground wave propagation range decreases with increasing frequency.

Excellent results have been obtained using the 1.34 and 0.693 MHz transmissions. As a cross check, the phase differences produced over the array were calculated for a knowledge of the transmitter locations. In Table 1 these calculated phases are compared with the measured values and agreement to better than 10 degrees is obtained in most cases

	Channel	1	2	3	4	5	6	7
BBC Radio 2 0.693 MHz	Expected	148	330	259	0	258	203	339
Dist. = 76 Km θ = 349°	Measured	155	328	251	0	260	209	324
BBC Radio 4 1.34 MHz	Expected	[illegible]	165	43	0	277	85	317
Dist. = 11 Km θ = 141°	Measured	211	160	42	0	260	83	315

TABLE 1 Comparison of measured and calculated phases on each element of the array for 0.693 and 1.34 MHz ground wave signals

It is evident that in these examples accurate location of the signal source could be achieved from the phase observations. The occurrence of interfering signals considerably disturbs the phase measurements and in general it was not possible to locate the transmitters accurately in the presence of interference.

4. OBSERVATIONS OF SKY WAVE SIGNALS

Signals from distant HF transmitters are received via the ionosphere, consequently, changes in the electron density distribution within the ionosphere play an important role in determining the phase and amplitude distribution across the received wavefront. The radio waves frequently travel between transmitter and receiver by means of more than one path and several 'modes' of propagation are possible, e.g. 1 hop E and 1 hop F. Thus, two or more independently varying components can be incident on the receiving array. It is important therefore to determine which propagation modes are active over a particular path at any given time. Ideally an oblique ionospheric sounder would be employed but since such an instrument is not available, the active modes have been estimated from vertical incidence ionograms taken near the receiving site and the appropriate transmission curves. This method is well established and details are given in the literature (Davies, 1965; Dieminger, 1968). A further estimate of the mode activity over a particular path has been derived from standard prediction programmes, such as the ITS/OT programme (Lucas and Haydon, 1966) which utilise models of average ionospheric conditions.

Even when propagation is nominally single moded, localised disturbances can occur in the ionosphere which greatly affect the characteristics of the received signal. Medium scale travelling disturbances (TIDs) are examples of such events. These can distort the isoionic contours so that both the amplitude and phase of the received signal are considerably modified. The phase and amplitude of the signal vary with time as the TID moves through the radio wave reflection zone. It should be emphasised that the extent of the localised disturbances is quite variable and ranges from tens of meters to hundreds of kilometers. The smaller disturbances often produce a "roughness" in the ionosphere and true specular reflection is not achieved under these conditions. Quite frequently a cone of rays [illegible] from the transmitter to the receiver and mutual interference can occur between constituent rays within the cone, a condition referred to as 'in mode' interference. Even though large scale symmetry is not achieved, the small irregularities can act as a phase diffracting screen and so produce a distribution of phase and amplitude over the receiving array not unlike a diffraction pattern. The multiplicity of propagation conditions complicates the study of the wave fronts of sky wave signals measured on the large array. To assist the interpretation quiet and disturbed conditions are treated separately.

4.1 Undisturbed Conditions

On no occasion were the phase and amplitude of the sky wave signals observed to be as constant as for ground wave propagation. A typical quiet situation is shown in Figure 6 for the 9.535 MHz transmission from Beromunster, Switzerland, (path length 850 km). At the time of the observations the ionogram/transmission curve, indicated a 1F mode with the possibility of $1E_s$ propagation if the E_s layer was present in the region of the reflection point of the wave.

During the period shown, approximately 2 min, three interference events occur with minima at 18.55.10, 18.56.20 and 18.57.10 GMT. These features are characterised by amplitude changes of some 10-12 dB and phase changes of approximately 75 degrees. The amplitude minimum occurs at the time of the most rapid change in phase. This suggests interference between two quasi-sinusoidal wave components, the amplitude of one being about five times that of the other. This factor is consistent with the greater attenuation suffered by a 1E mode during its passage through the absorbing D-region. Between the rather pronounced minima, the wavefront is approximately linear across the array, and it is at these times that optimum performance could be expected from a wide aperture array. Several examples of this type of behaviour have been observed.

4.2 Disturbed Conditions

The ionosphere is in general a turbulent medium and it is therefore expected that appreciable disturbances of phase and amplitude will arise due to localised turbulences within the Fresnel zone of the reflected wave. An example of a moderately disturbed period is reproduced in Figure 7. Between 17.28.30 and 17.29.15 GMT there is a gradual phase retardation which indicates that the optical path length between transmitter and receiver is increasing. From 17.29.15 to 17.30.20 GMT the phase change recovers and the phase advance corresponds to a decrease in optical path. This situation arises from a bulk movement of the reflecting region such as might be produced by vertical drift of the plasma. Superimposed on to the gradual phase change are a number of interference events with a quasi period of about 10 sec. During the amplitude minima of these events the phase exhibits its most rapid change of phase. This suggests interference between two or more components. The ionogram and transmission curves indicate that the propagation is single moded (1F) at this time. It would appear therefore that the 10 sec period events arise from mutual interference between the ordinary and extraordinary magneto-ionic components. It seems unlikely that high angle rays or E modes are sufficiently strong at this time to play a significant role.

A more extreme example of short period interference events, superimposed on gradual changes is reproduced in Figure 8. Very rapid fluctuations in phase and amplitude are evident with periods of approximately 1 sec. This corresponds to rapid changes in the various components of the signal such as would be associated with non specular reflection of the signal. The ionograms for this period suggest that the F region may be breaking up during the onset of spread F. Moreover, both 1F and $1E_s$ modes appear to be active at this time. During conditions such as these considerable distortions in the wavefront of the received signals can be expected.

Spectrum analysis of the signal indicates the existence of multiple Doppler shifted frequency components which again provides evidence for a 'non specular-spread' type of reflection mechanism. Similar spread of the frequency spectrum of the received signals for spread F conditions has been reported (e.g. Jones and Spracklen, 1973).

An interesting feature of these disturbances is the time delays between their observation on the various elements of the array (see Figure 6 and Figure 7 for example). In Figure 6 the events occur first on element 1 and progressively later on each element up to the 7th. Figure 7 indicates a similar time progression, but in this case the irregularities are first noted on element 7 and sequentially later down to element 1. These results correspond to events propagating from West to East and East to West respectively. Since the time delays between observations of the events on each of the various antenna elements are known, together with the spatial location of the elements, the apparent horizontal velocity of the irregularities can be determined by triangulation. Only three spaced observing locations are necessary to obtain the vector velocity. The availability of 7 locations provides a useful cross-check facility and enables the change in the form of the events as they move across the array to be investigated. The speeds and directions calculated for some of the disturbances shown in Figure 6 are reproduced in Table 2 below.

Antenna Triangle (element numbers)			Speed (m/s)	Direction (degrees)
1	4	7	286	247
2	4	6	280	252
1	4	5	265	247
1	4	6	276	251
2	[illegible]	7	312	247
2	4	5	2[illegible]	251
1	5	7	285	245
2	5	7	293	246

TABLE 2 Speed and direction of the disturbances shown in Figure 6. Calculations are made for various triangles within the array limits

The measured speeds across the array of approximately 250 meters sec^{-1} are consistent with other observations of F-region travelling irregularities. No systematic study has yet been undertaken of the distribution of speeds and directions but many events appear to move towards the South-East.

5. CONCLUSIONS

This study has examined the distribution of phase and amplitude on widely spaced elements of a large antenna array. Measurements for both ground waves and sky waves are reported and these show completely different characteristics. In the case of ground wave propagation, it is possible to measure the curvature of the wavefront and hence determine the transmitter's location. The limitation imposed on the present investigation by the wide spacing of the array elements and the resulting ambiguity in the measurements of the wavefronts is recognised.

The principal characteristic of the sky wave observations is the interference phenomena which appear to be present to some extent at all times. It has proved difficult, without oblique sounding facilities, to judge when propagation is single moded, although the use of ionograms and transmission curves is of some help. A wide range of interference phenomena are reported together with more slowly varying changes associated with bulk movements of the ionosphere. All but the most rapid interference features exhibit time delays as they propagate across the array. This enables their horizontal velocity to be determined and the results obtained are consistent with the accepted value for F-region travelling irregularities.

The variability in the periodicity of the interference effects suggests that some times are more suitable for optimum operation of large arrays than others. Further studies are being undertaken to relate array performance with ionospheric conditions. Hopefully it will be possible to recognise those conditions when optimum performance will be achieved.

REFERENCES

BOOKER, H.G., Ratcliffe, J.A. and Shinn, D.H. (1950) Diffraction from an irregular screen with applications to ionospheric problems. Phil. Trans. Roy. Soc. A242, 579.

DAVIES, K. (1965) Ionospheric Radio Propagation. N.B.S. Monograph 80.

DIEMINGER, W. (1968) Ionospheric propagation of HF radio waves. AGARD Lecture Series 29.

GETHING, P.J.D. (1978) Radio Direction Finding. I.E.E. Electromagnetic Wave Series 4.

HAYDEN, E.C. (1961) Propagation studies using direction finders. Radio Science 65D, 197.

JONES, T.B. and Reynolds, J.S.B. (1975) Ionospheric perturbations and their effect on the accuracy of HF direction finders. Radio and Electronic Engineer 45, 63.

JONES, T.B. and Spracklen, C.T. (1978) Technique for assessing the bearing accuracy of distant HF transmitters. Proc. I.E.E. 125, 469.

LUCAS, D.L. and Haydon, G.W. (1966) Predicting statistical performance for HF ionospheric telecommunications systems. E.S.S.A. Tech. Rep. IERI-ITSA1.

RICE, D.W. (1975) High resolution measurements of time delay and angle of arrival over a [illegible] km path. AGARD Conf. Proc. CP-173, chapter 33.

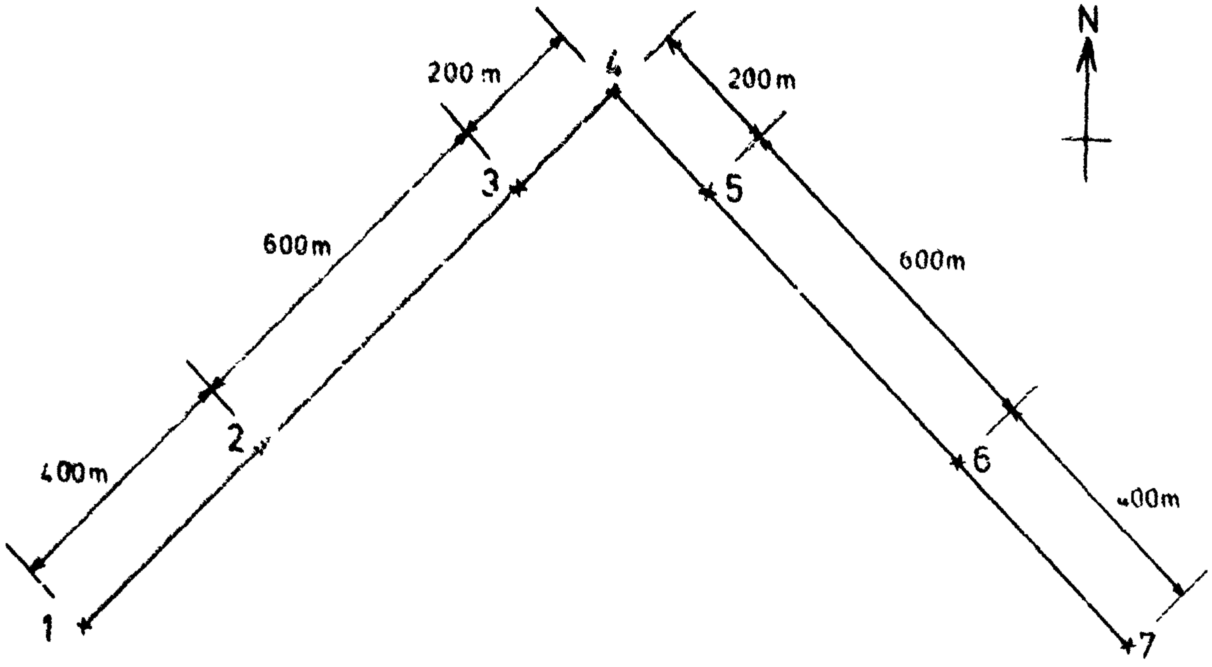

N
200 m
200m
4
3
5
600m
600m
2
6
400m
400m
1
7

FIGURE 1 Element positions in the minimum redundancy array. Unit spacing 200 m.

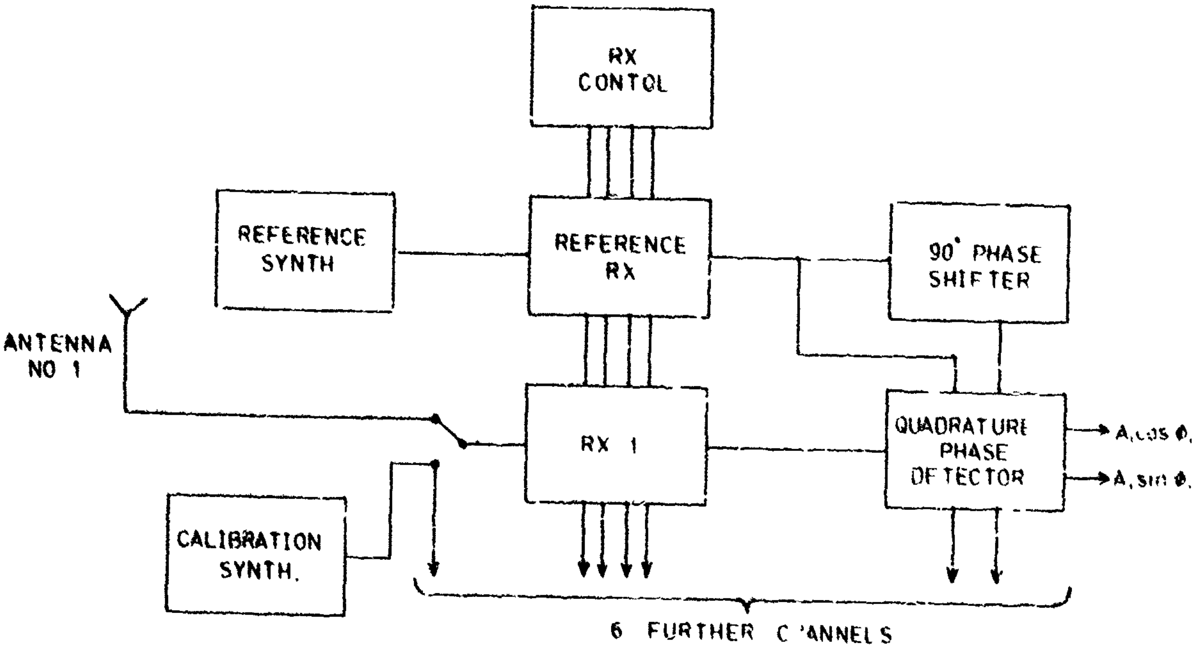

RX
CONTOL
REFERENCE
SYNTH
REFERENCE
RX
90° PHASE
SHIFTER
ANTENNA
NO 1
RX 1
QUADRATURE
PHASE
DETECTOR
A₁cos ϕ₁
A₁sin ϕ₁
CALIBRATION
SYNTH.
6 FURTHER CHANNELS

FIGURE 2 Diagram of receiving and phase measuring equipment.

START TIME 15 42 59 END TIME 15 45 20 FIRST BLOCK NO. 15 3 NO. OF DATA BLOCKS 10 LINEAR TREND REMOVED AT 50.375 DEG. PER POINT

DATA FOR RADIO 2 ON 0.693 MHZ. LOGGED ON 19.1.78. RECEIVERS ON MANUAL.

DATA IS CONTAINED IN FILE NO. 59 ON TAPE NO. 000895.

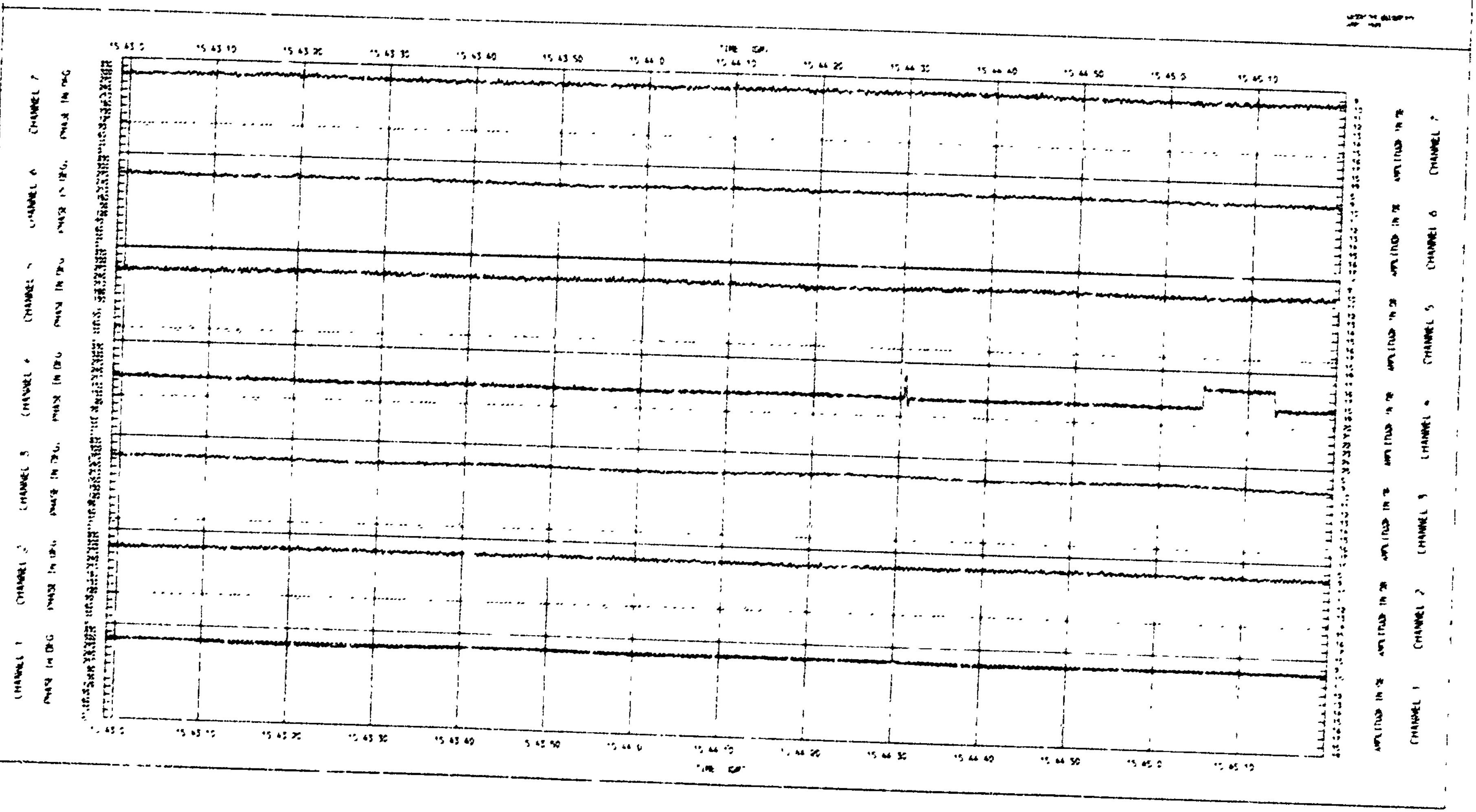


FIGURE 3 Phase and amplitude recorded on each of the 7 elements for the 0.693 MHz ground wave received over a distance of 75 km.

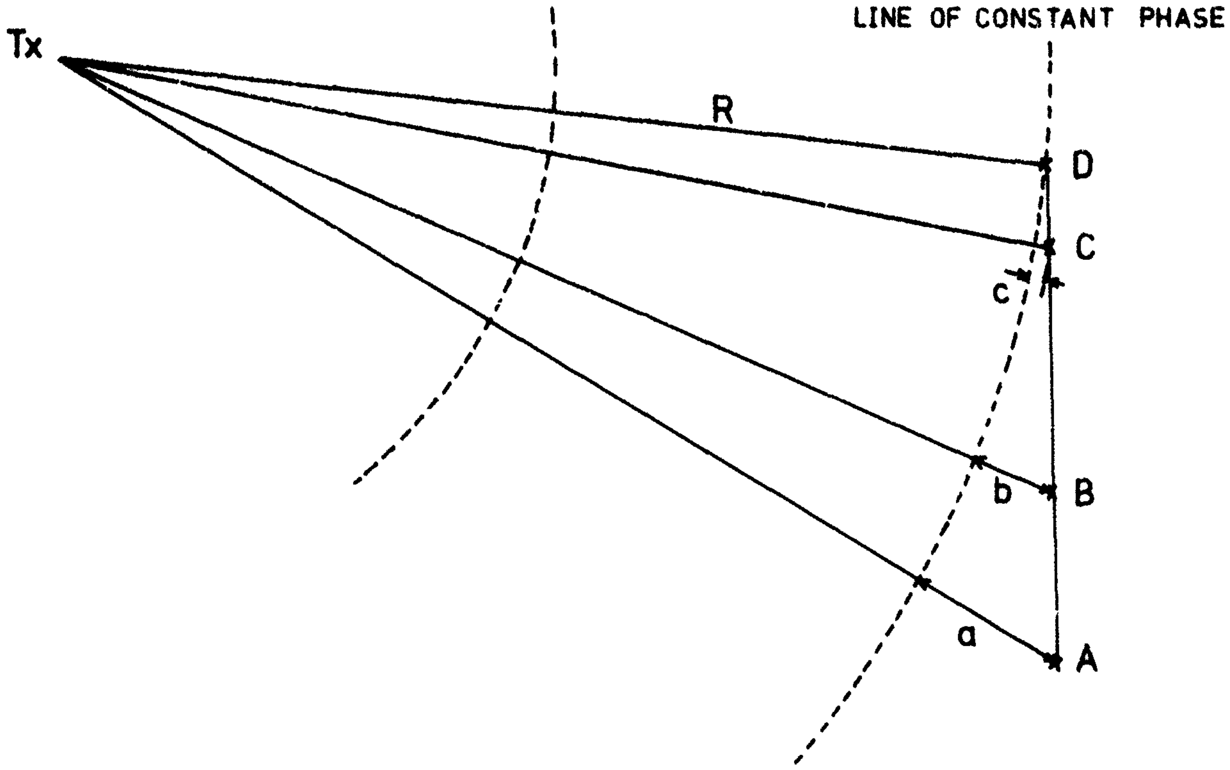


FIGURE 4 Illustration of the change in wavefront curvature as the wave travels away from the transmitter. If the phase is measured at locations A, B and C then location of the wave source can be determined.

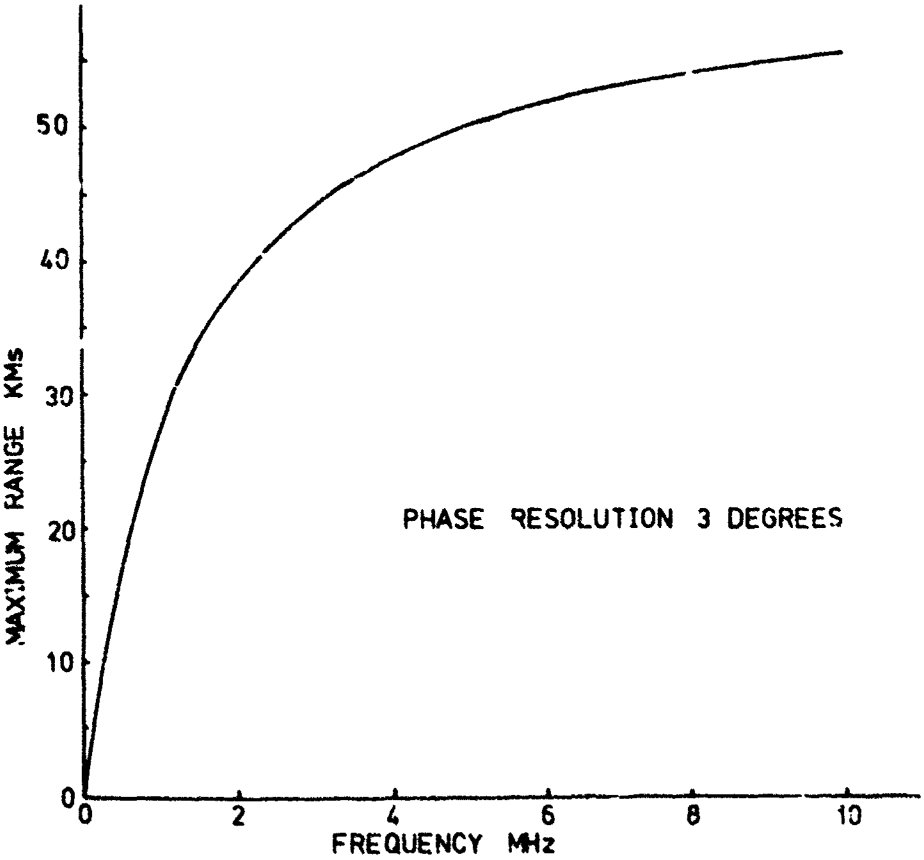


FIGURE 5 Variation of distance at which transmitter can be located with frequency for the present array and a phase measuring accuracy of 3 degrees.

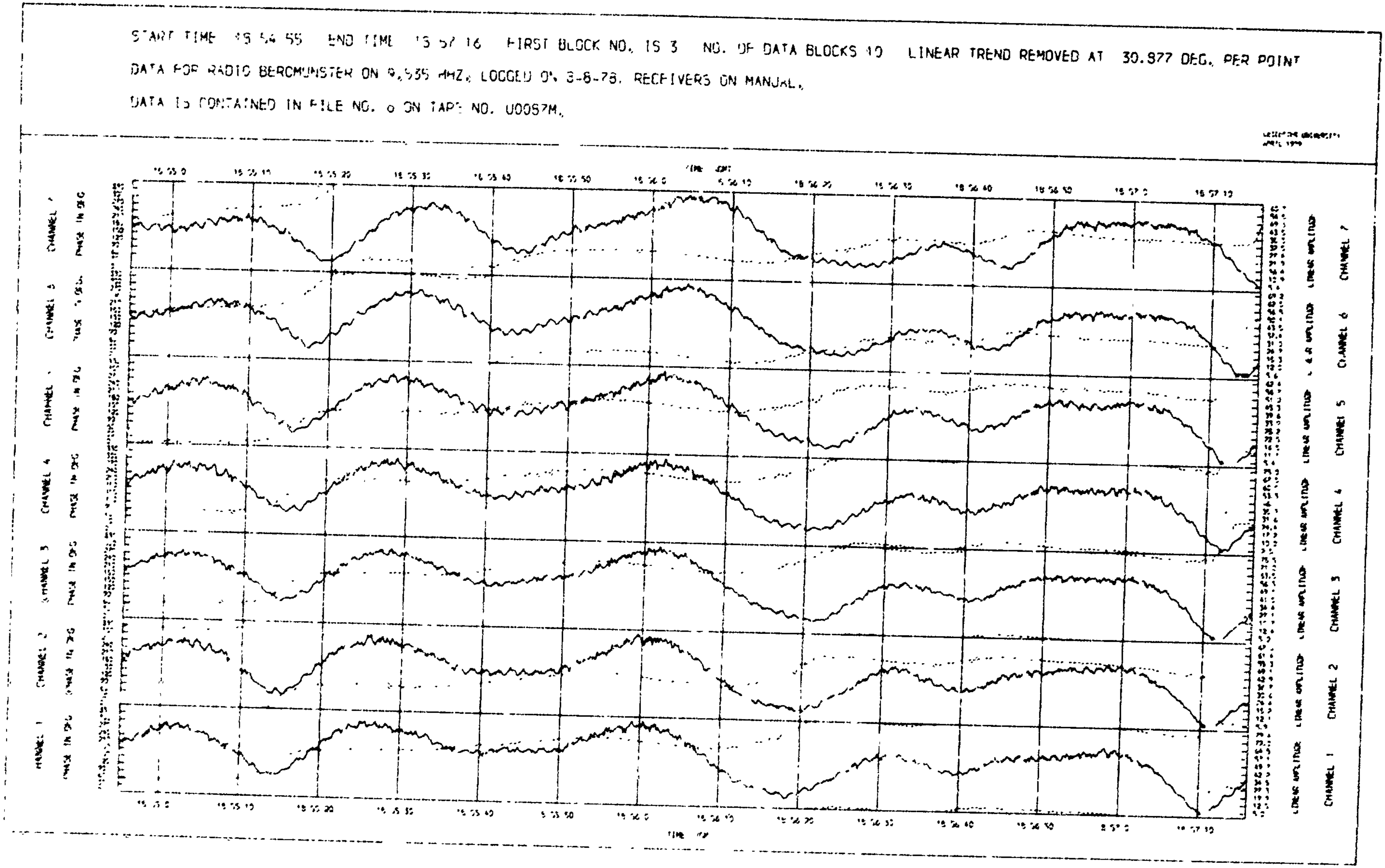
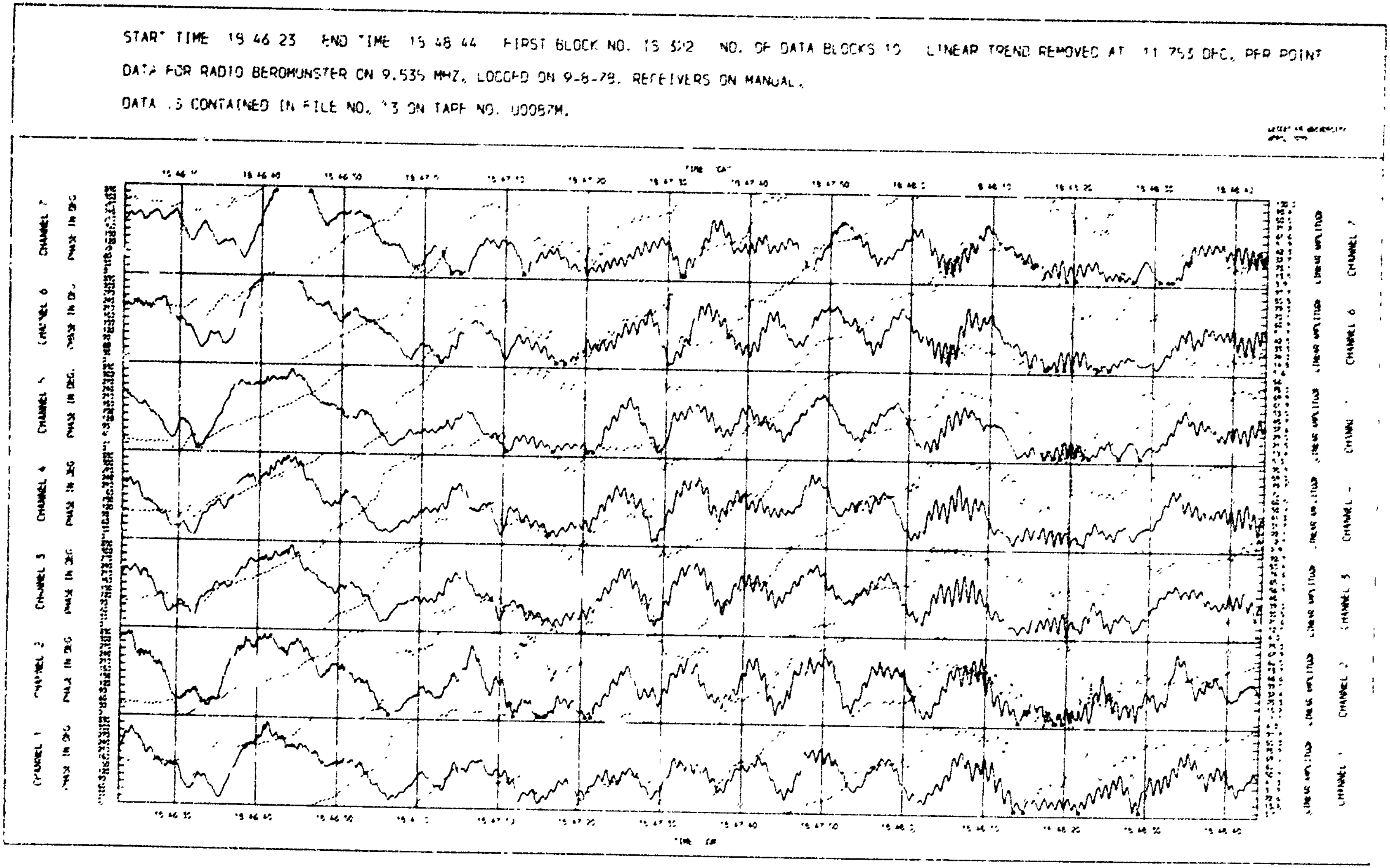


FIGURE 6 Phase and amplitude recorded on each of the 7 elements of the array for a 9.535 MHz sky wave received over a 850 km path from Beromunster, Switzerland. Quiet conditions.

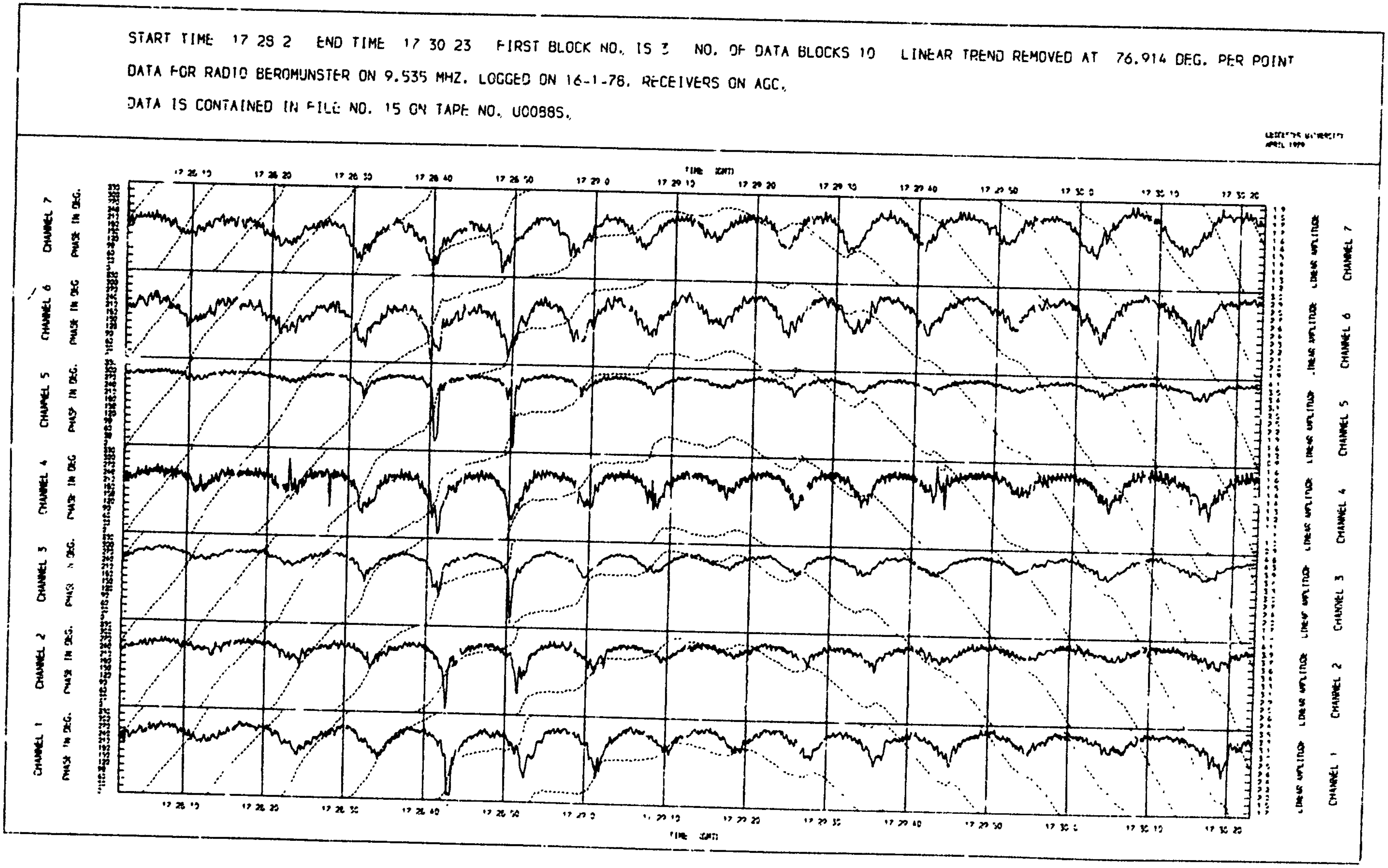


START TIME 19 46 23 END TIME 19 48 44 FIRST BLOCK NO. 19 502 NO. OF DATA BLOCKS 10 LINEAR TREND REMOVED AT 11 753 DEG. PER POINT

DATA FOR RADIO BEROMUNSTER ON 9.535 MHZ, LOGGED ON 9-8-78. RECEIVERS ON MANUAL.

DATA IS CONTAINED IN FILE NO. 13 ON TAPE NO. U0087M.

FIGURE 7 Phase and amplitude recorded on each of the 7 elements of the array for a 9.535 MHz sky wave received over a 850 km path from Beromunster, Switzerland. Moderately disturbed conditions.



START TIME 17 28 2 END TIME 17 30 23 FIRST BLOCK NO. IS 3 NO. OF DATA BLOCKS 10 LINEAR TREND REMOVED AT 76.914 DEG. PER POINT

DATA FOR RADIO BEROMUNSTER ON 9.535 MHZ. LOGGED ON 16-1-78. RECEIVERS ON AGC.

DATA IS CONTAINED IN FILE NO. 15 ON TAPE NO. U0088S.

FIGURE 8 Phase and amplitude recorded on each of the 7 elements of the array for a 9.535 MHz sky wave received over a 850 km path from Beromunster, Switzerland. Very disturbed conditions.

DISCUSSION

J. Röttger, Ge

Your results of drift speed and direction obtained from different triangles to scatter presumbaly due to changes of the pattern drifting through the antenna array. Have you applied the full correlation or cross spectrum analysis to compensate at least partly for changes in pattern?

Author's Reply

We have used the full correlation technique for some examples of these time displaced irregularities. The results shown in Table 2 are however, obtained by means of simple triangulation only. There are problems in interpreting these kind of data when more than one travelling wave is present in the viewing area of the array at the same time. I know of no way of resolving the ambiguities which can occur when more than one travelling disturbance occurs simultaneously in the ionosphere.

E.N. Bramley, UK

I suggest that care is needed in the interpretation of the drift speeds and directions indicated by the "similar fades" method of analysis used. In the case of a simple two-wave interference pattern the only motion that can be detected is in a direction perpendicular to the mean direction of the lines of constant amplitude, and may therefore not give much information about ionospheric movements.

Author's Reply

I agree with this comment. When there are two or more propagation waves present simultaneously in the field of view it is virtually impossible to separate them when using "full correlation" methods The only method that occurs to me is the one of a large "filled in" array such as employed by Dr Briggs at the University of Adelaide, Australia.

E.D.R. Shearman, UK

Dr Jones has referred to the use of large aperture array for measuring the curvature of a wavefront for determination of range, he has referred to the limitations of Fourier analysis for resolution of the different modes before measuring the phase gradients of individual modes.

At the University of Birmingham, UK, Mr S.S Ng has been making an experimental study of this technique and has shown that on a sample of occasions it is possible to separate modes on the basis of differential Doppler shift and to measure wavefront curvature, work to establish the fraction of time this will be possible is proceeding. Clearly it will fail when the differential Doppler is within the resolution of the analysis for the coherent dwell time permitted by the stationarity of the process.

In our technique we use a calibration oscillator in the field and consider that this shall improve on the accuracy of phase measurement of 3°.

Author's Reply

The limitation of the Fourier analysis is associated with the number of sampling points and the duration of the sample available. It was not my intention to imply that FFT techniques will not work, indeed they work well when only frequency information is required. It is only when the relative phases of the various frequency components in the spectrum are required that special care must be taken, otherwise the phase relationships are destroyed.

The error of 3° of phase is the present limitation of our phase detectors. We clearly have the opportunity to better this resolution if we record the data directly and process it later on our mainframe computers. We have not done this as yet since we wish to keep the system in a real-time configuration at the field site at least for the present.

Mode reduction by means of Doppler technique has been demonstrated on several occasions (e.g. Jones & Sprackler IEE 1978). However, I suggest that there are often ionospheric conditions which are so disturbed that the resulting Doppler spectrum shows considerable spread and moreover is rapidly changing, with time. In these cases mode separation is not possible.

A Sporadic E Prediction Technique

Miles A. Merkel
U.S. Army Communications Command
ATTN: CCC-EMEO
FT. Huachuca, Arizona 85613

Roberto R
U.S. Army Atmospheric Sc ces Laboratory
White Sands Mis ile Range
New Mexico 88002

Summary

A unique Sporadic E assessment technique has been developed and compared with a quantity of empirical data. Sporadic E occurs in patchy clouds at an elevation of approximately 100 kilometers above the earth. It has significant impact on communications circuits because of the density of formation. It can enhance or degrade communications on a given circuit depending on the geometry of the path and location of the Sporadic E clouds. The uniqueness of the technique lies in the treatment of the scaling of the available data and utilization of parametric relations developed by Dolukahanov to convert from the relationships privided by the World Data Center to the operating frequency of interest. The technique was implemented on electronic computers to facilitate computations and comparisons, which show very good agreement with the data in the mid-latitudes. Recommendations are made for approaches which should improve the accuracy and reduce computer run-time.

Introduction

Crowding of the electromagnetic spectrum is a spectre haunting all of us, haunting us from the standpoint of not enough frequencies to go around. Haunting us from the standpoint of interference even on assigned frequencies. One of the most popular regions of the spectrum in this respect is that supported by the ionosphere. As the spectrum has become increasingly crowded, we have become more sophisticated in our hardware and in our definition of the medium through which we propagate our electromagnetic energies.

We now examine so-called "anomalous" propagation mechanisms in detail in attempts to remove them from the anamalous category. One such mechanism is Sporadic E. Many of our learned colleagues have examined this phenomenon in some detail. What is presented here is no break-through in the state-of-the-art, but rather operational application of their works. As such, it is hoped that successful application will be highlighted, but more importantly, that unsuccessful applications will highlight the need for further investigation in specific areas which hold promise for immediate payoffs in terms of improved communications. The technique employed here represents a slight deviation from the conventional prediction technique and incorporates suggestions put forward by our Russian Colleague, Dolukhanov. World Data Isopleth maps, which depict the probability of the Sporadic E layer vertical critical frequency being equal to or greater than three megahertz, have been scaled and converted to Fourier coefficents for implementation on electronic computers. This technique of scaling the available data insures a smoothly varying continuous function of $P\{F_oE_s \geq f\}$ down to $f=0$ whereas others have used scaling techniques where $P\{F_oE_s \geq 0\}=1$ for several values of $f>0$. This resulted in inflated probabilities at several values of $f>0$, arising from scaling conventions in which F_oE or F_{min} were reported as F_oE_s when no Sporadic E existed, thus biasing predicted frequencies to higher values. The conventional Secant ϕ law was used to convert the desired operating frequencies on the communication link to an equivalent vertical frequency at the point of reflection. Extrapolation to higher and lower operating frequencies, given the probability that $F_oE_s \geq 3$ megahertz, was accomplished through use of relationships borrowed from Dolukhanov for temperate and polar regions. Comparisons of resultant predictions against empirical data at latitudes of approximately 30° and 60° North latitudes show promising results, however, the technique appears to be confined in usefulness to mid-latitudes.

General Characteristics of Sporadic E

Sporadic E (Es) is one of a number of heavily ionized layers occurring above the earth's surface. It derives its name from the fact that the scientific community has so far been unable to define precisely what physical laws govern its formation and from the fact that it occurs at approximately the same elevation (90 to 120 km) as the relatively well behaved "regular" E layer.

The Es layer is a world-wide phenomenon, and some difficulty has been encountered in categorizing it because of its differing characteristics with geographical location. A thorough survey for the neophyte is contained in the paper by Whitehead [5], which identifies many of the problems encountered to that time. A few general observations made over the past few decades are presented in the following paragraphs.

Ionospheric propagation is usually thought to be confined to the high-frequency range (3-30 MHz); however, very strong support for operationally usable periods of time have been observed for frequencies up to 100 MHz, and numerous reports have been recorded of effects to 150 MHz. The problem is one of being able to determine, in advance, when the Es layer will be enhancing or degrading operational performance of a given link.

The occurrence of Es sometimes enhances and sometimes degrades communication system performance. An example of enhancement is shown in Figure 1.

Two passes through the highly absorbent D region are avoided, thus increasing the signal strength arriving at the receiver. An example of degradation is shown in Figure 2. The two extra passes through the D layer causes a resultant decrease in the signal strength arriving at the communication system performance.

The Es layer has sometimes been observed to be opaque, blanketing the upper layers as observed in the N type propagation mode, sometimes the upper layers can be seen through the Es layer suggesting that the electromagnetic energy is penetrating through patchy gaps. The Es layer is commonly referred to as consisting of clouds of electrons. These clouds form for periods of minutes to hours and can extend from tens to hundreds of kolometers.

A band, or belt, of Es approximately 1000 km in width is centered on the Magnetic Equator and occurs during local daylight hours. This phenomenon, referred to by the Australians as Magnetic Equatorial Sporadic E (or MEQUES) has bery little sporadicity to it and is reasonably predictable. There is bery little discernable seasonal variability.

In the Auroral zone Es mostly occurs during nightime, with only some discernable seasonal variability during midday.

The least world-wide occurrence of Es is observed at the middle latitudes and is characterized by both diurnal and seasonal variability. At these latitudes, Es occurs most frequently in local summer during daylight hours.

Prediction Models

Given the geographical location of a Transmitter and intended Receiver, path parameter geometry is determined. The equivalent vertical frequency is determined at the mid path reflection point using the operating frequency and the Scant ϕ law.

Having determined the equivalent vertical frequency it is next determined whether the critical frequency of the Es layer is greater than or equal to 3 MHz.

The probability that the vertical critical frequency of the Es layer will exceed 3 MHz is obtained from values scaled off maps which were obtained from the World Data Center at Boulder, Colorado. These data are for March, August, September, October, and November of 1963 and for January, February, March, April, and May of 1964. These data therefore represent low solar activity periods

The scaled data were utilized to develop Fourier coefficents to facilitate electronic computer extraction of world wide values of probability that the vertical critical frequency of the Es layer will exceed 3 MHz. The mapping technique utilized was adopted from the method derived by Lucas and Harper [4].

For each month specified in the input, a set of 25 X 29 + 29 + 2 Fourier mapping coefficients is used to determine $P\{F_oE_s \geq 3 \text{ MHz}\}$. The general form of this numerical [illegible] os the Fourier time series:

$$\tau(\lambda,\theta,t)=a_0(\lambda,\theta)+\sum_{J=1}^{H}[a_j(\lambda,\theta)\cos_j t=b_j(\lambda,\theta)\sin_j \bar{t}],$$

where

$\tau = P\{F_oE_s \geq 3$ MHz

λ = Latitude of reflection point

t = Local time

H = Harmonics retained to represent diurnal variations.

The Fourier coefficents a_j and b_j, giving the latitude and longitude variations are defined as:

$$\sum_{k=o}^{K} D_{sk}\ G_k\ (\lambda,\theta),$$

where $G_k\ (\lambda,\theta)$ are geographic coordinate functions and D_{sk} is a coefficient dependent upon the resultant statistics of empirical data. Further detail may be obtained from CCIR Report 252-2 of the 14th Plenary Assembly held in Kyoto, Japan in 1978.

Having determined the probability of occurrence $F_oE_s \geq 3$ MHz, a set of probability of occurrence versus operating frequency parametric relationships is referenced. These relationships were developed from recommendations put forth by Dolukhanov. His technique consists of two families of curves covering the globe in the temperate (+45° to -45°) and the polar (90° to 45°, N & S) regions. From these curves the probability is determined of whether an equivalent vertical operating frequency will be less than or equal to the vertical critical frequency. These curves, given in terms of the probability that Es-2000 MUF (Maximum Usable Frequency) will equal or exceed the critical frequency of the Es layer, were normalized to the probability that F_0E_s will equal or exceed any desired operating frequency. These curves are entered after conversion of the desired oblique operating frequency to an equivalent vertical frequency via the Scant ϕ law.

Having determined the probability of support of the desired operating frequency on the desired path, the relatively simple task of computing transmission losses and signal-to-noise ratios were implemented using techniques which have been well documented in the technical literature.

Empirical Verification

The Sporadic E prediction technique described in this paper has been implemented in the computer program RUSSES. Extensive comparisons have been made of predictions from this program against empirical data obtained from hourly values of F_oE_s scaled from C-3 and C-4 Ionospheric sounders. Data were collected from 4 stations:

Station Name	Lat	Long	Lat	Long	Years
Narssarssuaq, Greenland	61.20	314.60	71.14	37.72	58,62,64,68
Boulder, Colorado	40.00	254.70	48.89	317.00	63,64,68
WSMR, N.M.	32.30	253.50	41.14	317.47	58,62,64,68
Huancayo, Peru	-12.00	284.70	-0.59	354.30	58,64,68

Narssarsuaq data were obtained from the Ionosphere Laboratory, Technical University of Denmark; Boulder data were obtained from the National Oceanic and Atmospheric Administration, U.S. Department of Commerce; WSMR (White Sands Missile Range) data were obtained from the Atmospheric Sciences Laboratory, US Department of the Army; and Huancayo data were obtained from the Instituo Geofisico Del Peru. Representative examples of graphical comparisons of predictions vs empirical data are contained in Figures 3 through 8.

Observations and Conclusions

Better agreement has been reached, on the average, with the described technique than has been achieved with any other known technique. The same general agreement was also achieved for 70% and 90% probabilities of frequency support and for the spring and fall months Largest discrepancies have been found for Huancayo, Peru in December as seen in Figure 8. It is not known whether the predictions will have the same utility during periods of high sunspot activity without the addition of data for these periods.

References

[1] Bazarzhapov, A.D., 1967, "Analytic Description of the Planetary Distribution of Ionospheric Parameters", Institute of Cosmophysical Investigations and Aeronomy. Yakutsk Section, Siberian Branch, USSR Academy of Sciences.

[2] Chermyshev, O.V., 1968, "F_0E_S Distribution Curves", Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation, USSR Academy of sciences.

[3] Dolukhanov, M.P., 1960, "Rasprostraneniye Radiovoln" Second Edition, Gosudarstvennoye Izdatel'stvo Literatury, PO Voprosam Svyazi i Radio, Moskva.

[4] Lucas, D.L., and Harper, J.D. JR.,"A Numerical Representation of CCIR Report 322 High Frequency (3-30 MC/S Atmospheric Radio Noise Data".

[5] Whitehead, J.D., 1970,"Production and Prediction of Sporadic E", Review of Geophysics and Space Physics, Vol 8,No 1

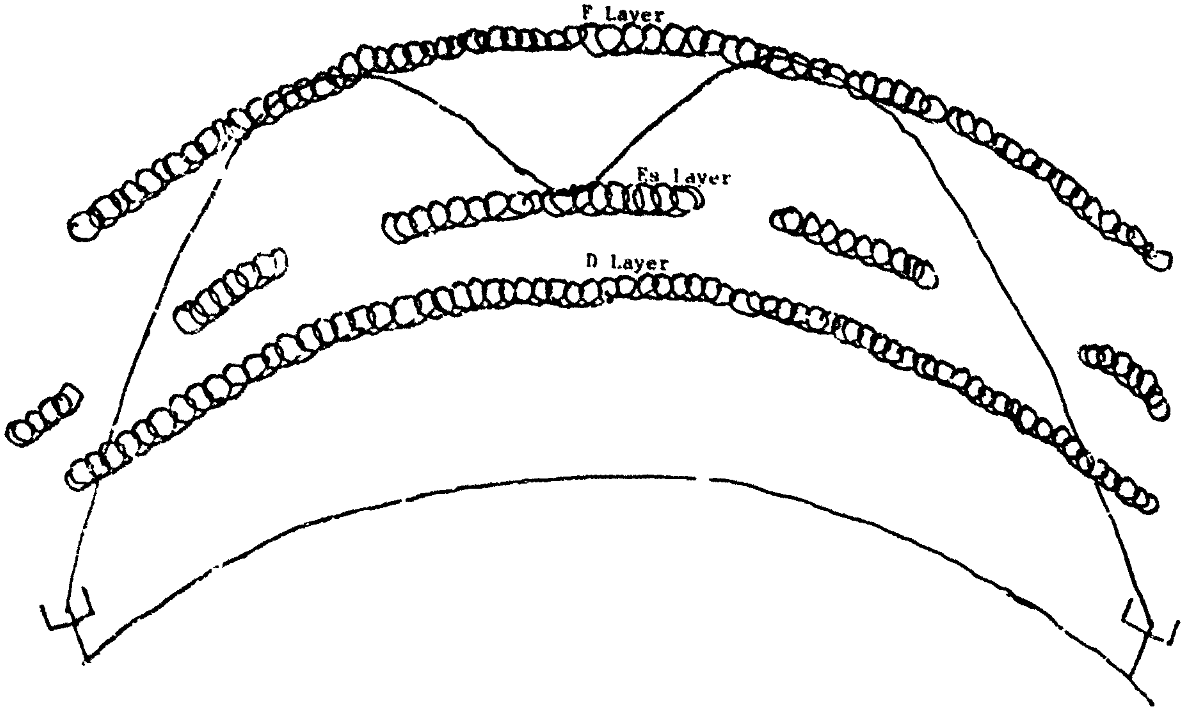

F Layer
Es Layer
D Layer

Fig.1 M type propagation mode

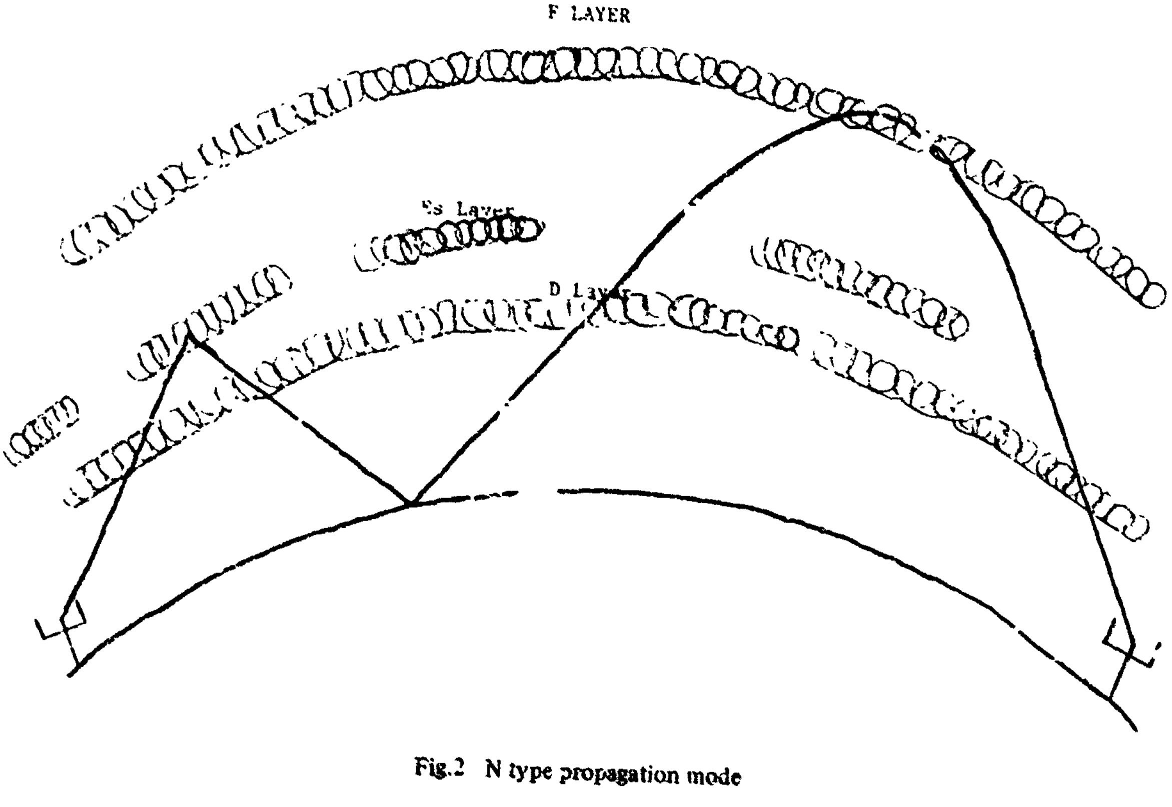

F LAYER
Es Layer
D Layer

Fig.2 N type propagation mode

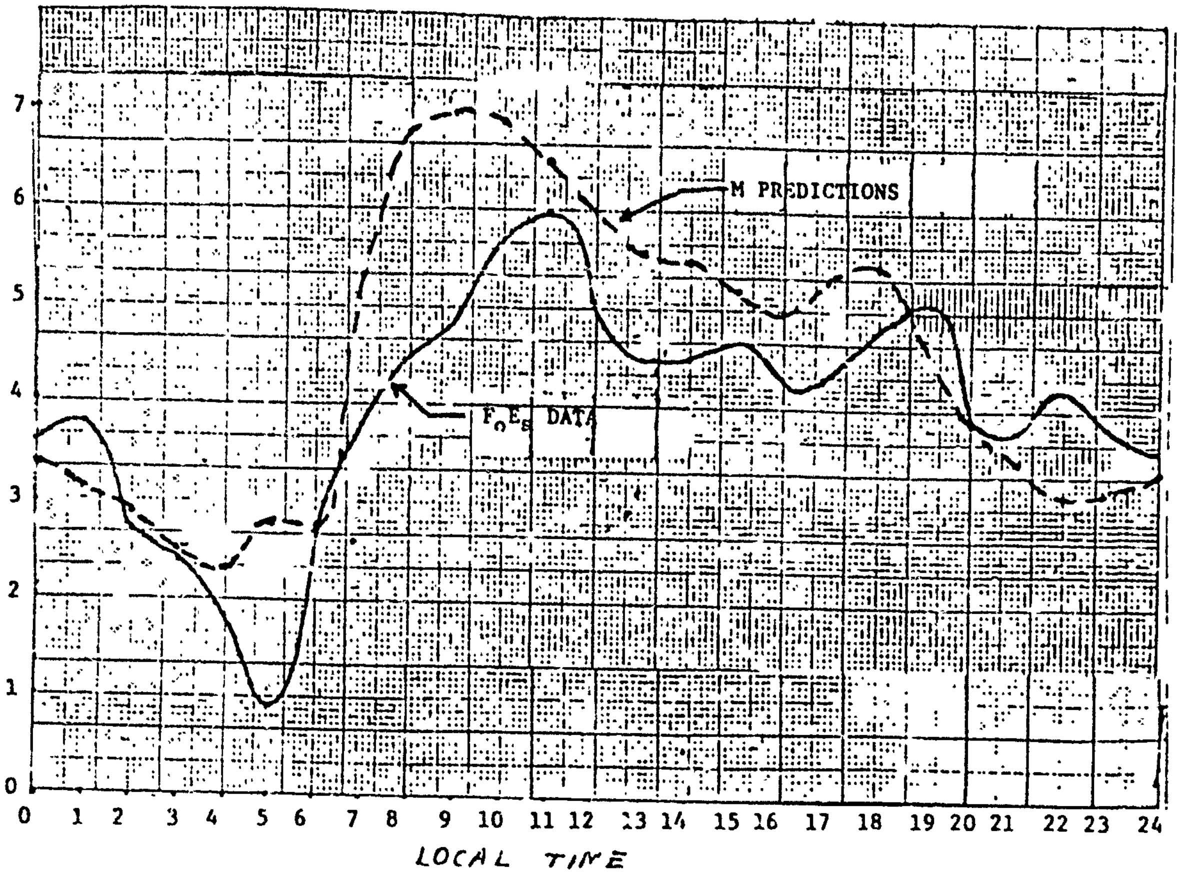


Fig.3 White Sands, NM 253E 32N medians June 1963

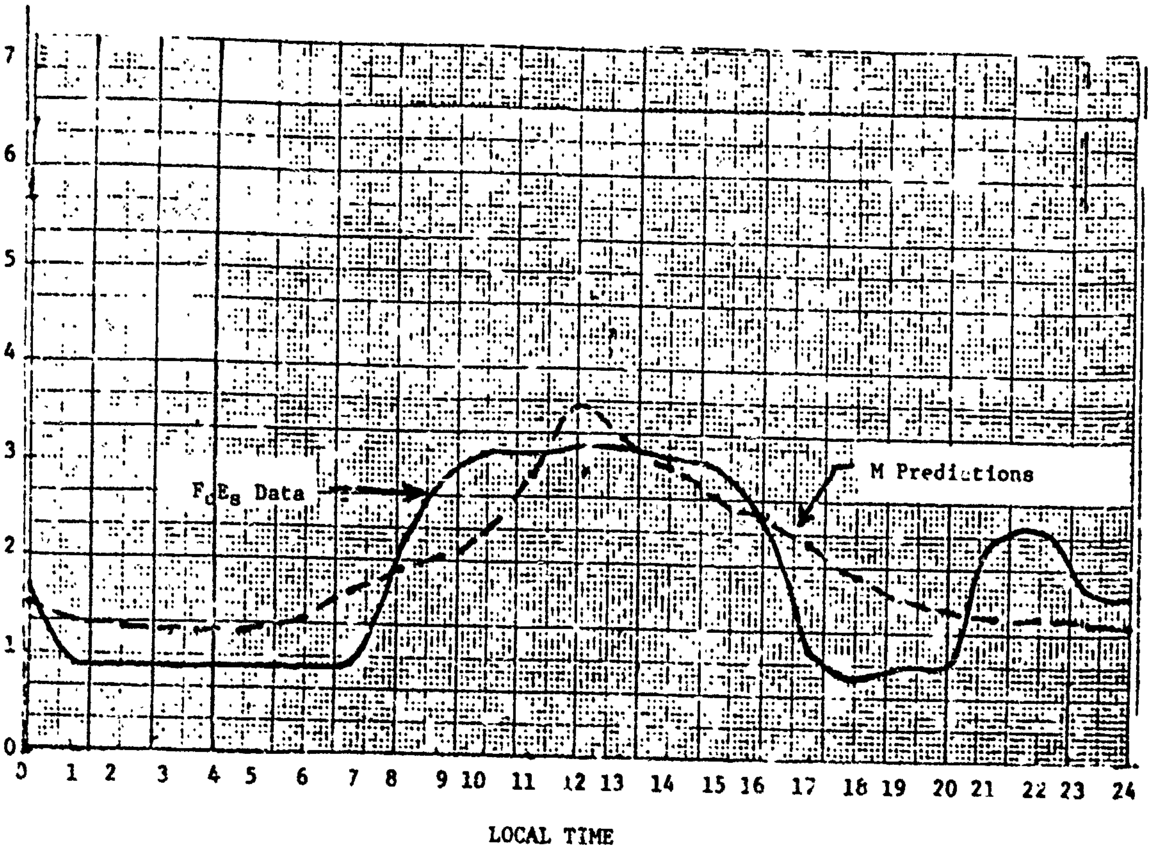


Fig.4 White Sands NW 253E 32N medians December 1963

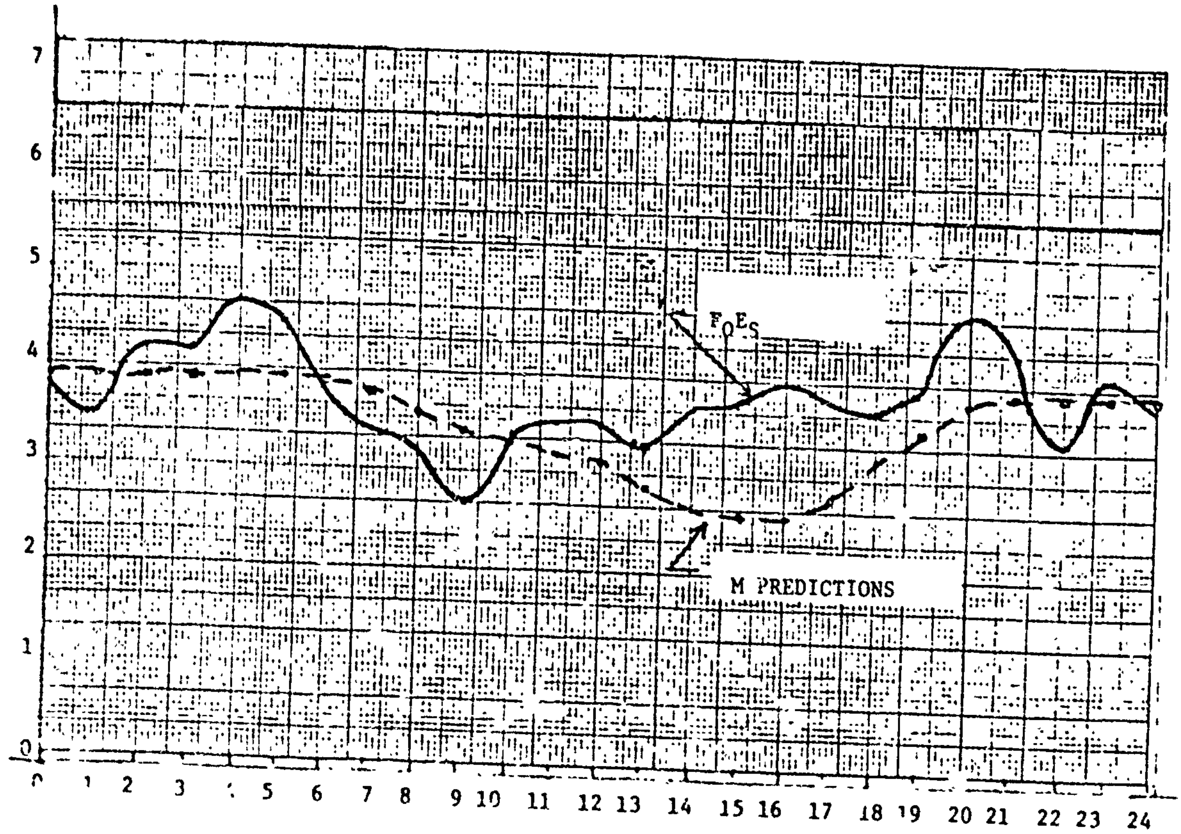


Fig 5 Narssarssuaq, GD 315E 61N medians June 1962

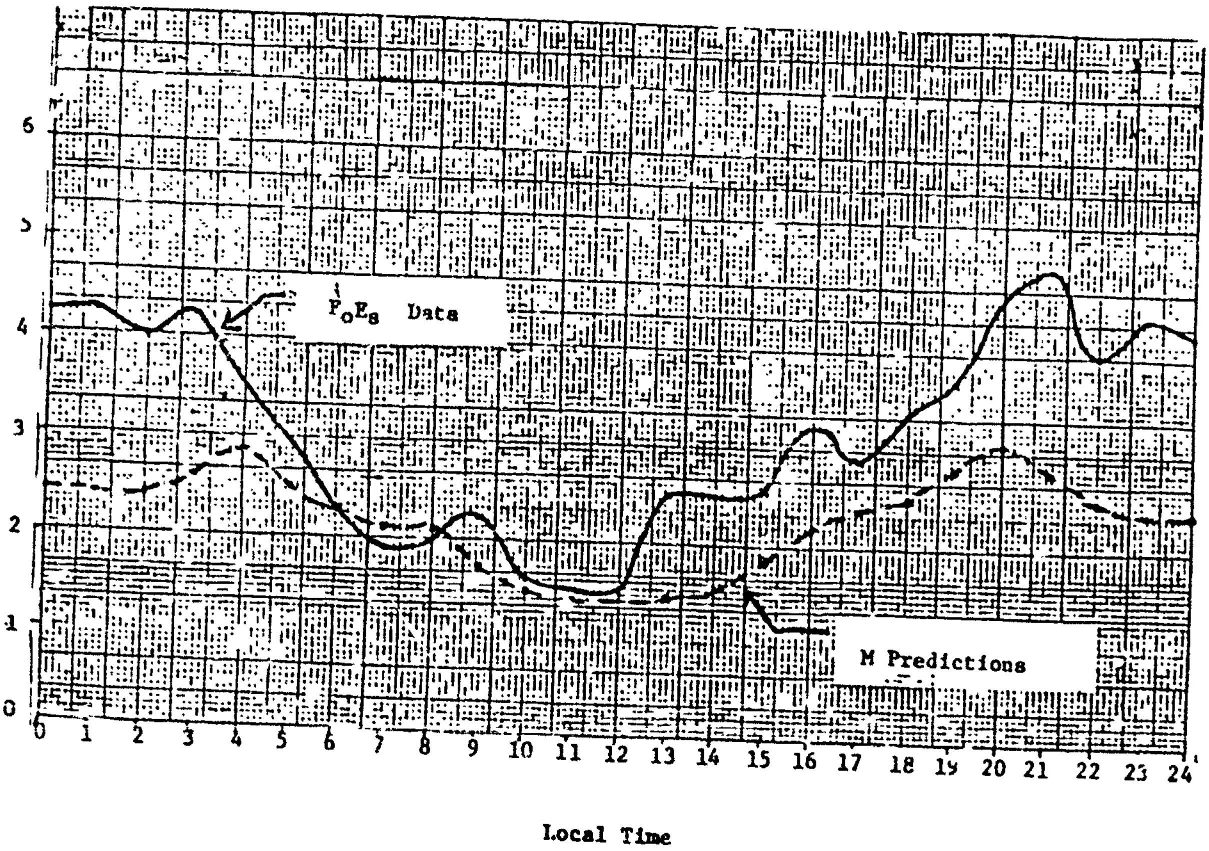


Fig.6 Narssarssuaq, GD 315E 61N medians December 1962

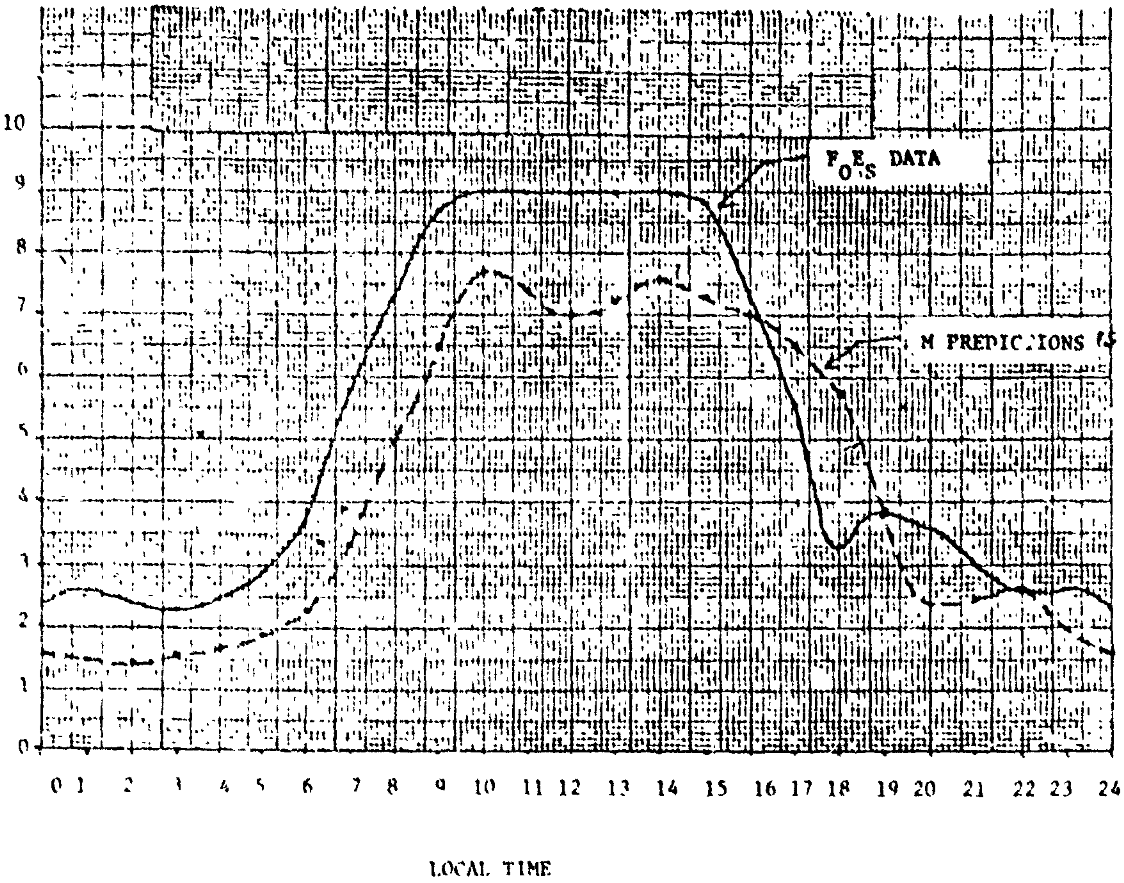


Fig 7 Huancayo, Peru 28SL 12S medians June 1964

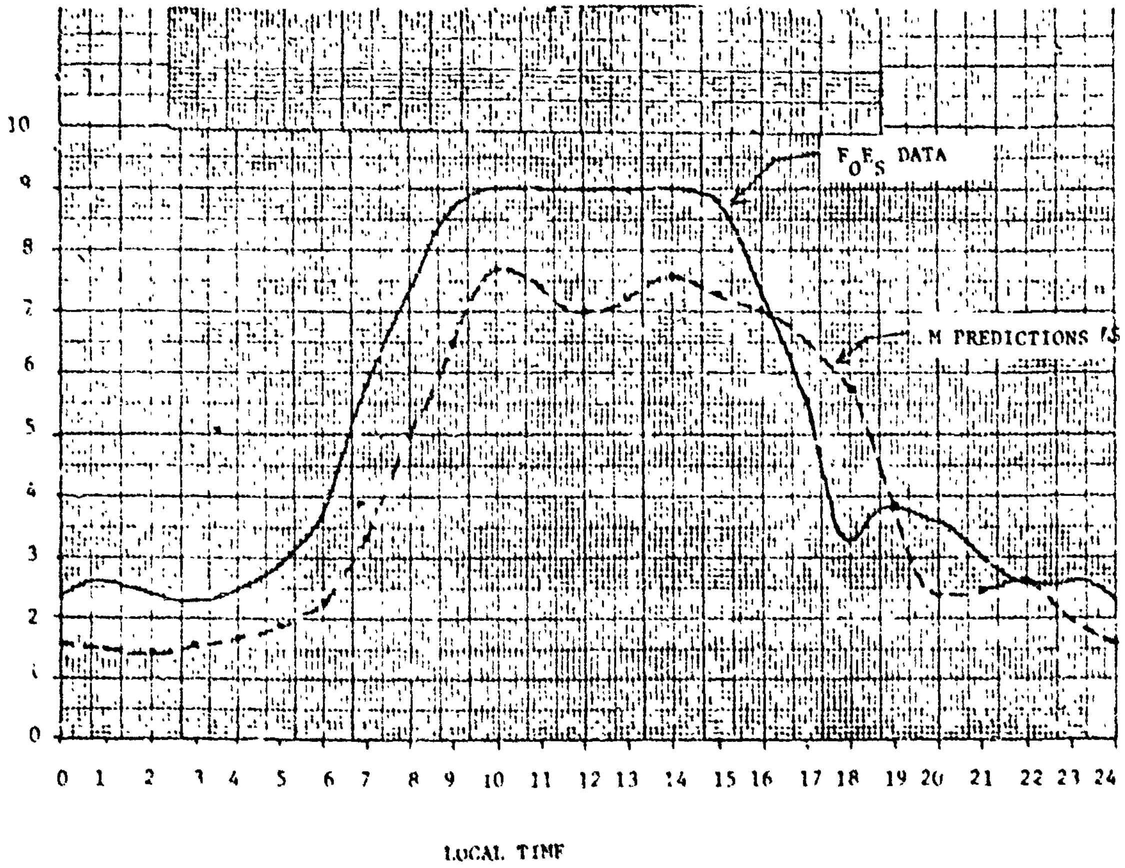


Fig.8 Huancayo, Peru 28SL 12S medians December 1964

DISCUSSION

Dr. I. Keroub, Israël

(1) Des travaux récents ont montré qu'il n'existe pas de corrélation évidente entre l'activité solaire et la statistique de FoEs aux latitudes moyennes. Auriez-vous des résultats expérimentaux informant ces conclusions?

(2) Par glissement d'une heure sur l'axe des temps, sur les figures 7 et 8 (Huancayo) de votre article, on obtient une meilleure corrélation. Ce décalage horaire systématique vous suggère-t-il une hypothèse, et si oui, laquelle?

Author's Reply

(1) The published literature is not clear on this. Opinions have been advanced to both negate and support the theory of dependence on solar activity. It is hoped there is no dependence, for this would simplify the problem; however, examination of data for mid and high solar activity years will have to be performed before the question can be resolved.

(2) I have not as yet considered examining a possible hypothesis to attempt resolution of apparent lack of correlation. First an examination for possible computer programming errors will be made. Second the data will be examined closely to ensure valid data interpretation. Should these two efforts fail, a suitable hypothesis will be developed.

SCATTER INJECTION/DUCTED MODE HF RADAR

by

Gary S Sales
Electromagnetic Sciences Division
Rome Air Development Center
Hanscom AFB MA 01731

1. INTRODUCTION

In order to provide for the reliable detection of missile induced ionospheric perturbations at distances of 8 to 13 Mm, an extended range HF radar has been conceived that uses Earth detached propagation modes, both ducted and chordal to minimize the losses over these great distances. It is well known that the utilization of these modes requires an injection and ejection mechanism since these modes are in general not available to a ground based transmitter. The successful launching of these Earth detached modes has depended historically on naturally occurring electron density gradients in the ionosphere, for example those that are present near the day/night terminator Any system depending on these naturally occurring gradients would have a very limited capability. The main purpose of this work was to study theoretically the effectiveness of artificially generated irregularities to facilitate the injection/ejection of HF signals into ducted modes.

2. THEORY

2.1 It was observed by several investigators during the operation of the powerful ionospheric heating system at Platteville, Co from 1971 to 1974 (Carroll, J C et al 1974, Utlant et al. 1974) that the most obvious effect of the absorbed radio energy was the generation of field aligned ionospheric irregularities. The production of artificial spread F was easily observed by the nearby ionosonde station at Erie, Co (Utlant et al 1974) The nominal dimensions of the irregularity region was given as 100 kilometers in diameter and 20 kilometers in height.

Given this volume of irregularities the decision was made to treat the HF scattering problem using the "specular" scattering theory of Booker (1956) and then ray tracing through a realistic model of the local ionosphere. This irregularity region was assumed to be imbedded in the local ionized medium, specified using the ITS-78 model (Barghausen, 1969) at an altitude which would be optimum for ionospheric ducting The particular choice for this altitude will be discussed later when the global model of ducted propagation is considered. At this point the computations were carried out for a set of altitudes, times of day, season and frequency. In order to determine how this scattering process affect the location of a ground station, near horizontal rays (0 and ± 5 degrees were launched at the nominal altitude of the irregularity region with various initial azimuths relative to magnetic north at Platteville, Co After computing the configuration of the scattering cone (Figure 1), the rays making up the cone at a given frequency, azimuth and altitude were then traced through the intervening ionosphere to the ground. Figure 2 shows the locus of the intersection of the cone with the ground after ray tracing. Altitudes of 170 and 200 kilometers and noon and midnight are shown for the summer period. Assuming field aligned irregularities, there will be symmetry for azimuths on either side of the magnetic field direction and therefor the calculating were carried out only for azimuths of 0 to 180 degrees in 45 degree steps. There is a clear progression towards the south (magnetic) as the azimuth decreases from 180 degrees towards 0 degrees. The ground intersection of the scattering cone varies very little with frequency when compared to the azimuthal variation.

The locus of the intersection of scattering cone with the ground indicates the optimum position for the reception or transmission of ionospherically ducting radio signals at the indicated frequency and azimuth. The azimuth of the ducted signal can be varied by moving either the receiver/transmitter to the ground location for the desired azimuth or moving the irregularity region and allowing the ground based system to remain fixed. The loci are computed on the basis of a single scatterer located at the center of the heated region Given the spatial extent of these irregularities, the ground intersection will assume a natural width that lessens the requirement for precise location The closeness of the frequency contours indicates that over the band 12-30 MHz satisfactory reception for a given azimuth can be accomplished from a single site.

For the global ducted propagation calculation a nominal distance of 10,000 kilometers was selected and extensive three dimensional ray tracing performed using a realistic ionospheric model which incorporates an E, F1 and F2-layer in daytime and only an F-layer at night with a realistic transition model between day and night. Most of these calculations used an elevated source at various altitudes in the ionosphere to simulate scatter injection using the irregularity region. At the injection point the rays were launched horizontally using several radio frequencies and at different times of day. Figure 3 shows an example of a single ray launched at an altitude of 120 km near local noon time. This ray is trapped between the E and F1-layers until sunset when the ray becomes a chordal mode on the bottom side of the nighttime F-layer until it reaches the sunrise transition when it becomes ducted between the F1 and F2-layer at an altitude of 180 km.

This particular ray is only shown as an example of the type of ducted modes that can occur and no consideration was given for this case to absorption effects along the path. Both absorption and spreading losses must be considered for a realistic assessment of these propagation modes. The absorption losses are very sensitive to the model chosen for the F1-layer. In general the published results, (Gurevich, 1973 and Matyugin, 1975) use a weak F1-layer resulting in a relatively deep ionization valley between F2 and E-layer. Under these conditions ducted propagation is easy in terms of the ability to trap HF signals for long distances. However, with more realistic models of the F1-layer the situation changes drastically. In general a valley in electron density is not a necessary condition for ducting to exist, but a minimum in the potential function defined by Gurevich (1971) as

$$\epsilon'(Z) = \left(1 - \frac{kN(Z)}{\omega^2}\right)\left(1 + \frac{2Z}{R_e}\right)$$

N(Z) is the electron density
ω is 2π x operating frequency
R_e is the radius of the Earth
k is a constant

In Figure 4a an isometric graph of electron density vs altitude and range is shown with a typical F1-layer present in the daytime region of the path. Although there is no minimum in electron density between the F1 and F2-layers, Figure 4b shows a small minimum in the potential function in this region, sufficient to trap HF radio waves.

When absorption losses are taken into account the deviative loss for the F1-E duct exceeds 60 dB for a path of 10,000 km making such a duct of little use in propagating to such distances. However, for the F1-F2 duct these losses are below 5 dB for the same path. In Figure 5, the losses have been calculated for elevated sources at the indicated altitudes. For local noon at the injection point, the losses for ground hop signals to a distance of 10,000 km varies between 100 and 20 dB, depending on the initial altitudes. This must be compared to the indicated absorption for the ducted modes, between 15 and 19 MHz, where the losses were less than 5 dB. For the particular model chosen all of the ducted modes originated when the source was located at 170 km altitude. The absorption calculated in Figure 5 are for non-deviative and deviative losses but do not include spreading losses.

A simple model was constructed for the analysis of azimuthal spreading. Using a spherical Earth model and a uniform duct of height H (see Figure 6) the cross sectional area A of the duct is

$$A = R_e \phi H \sin (D/R_e)$$

where R_e is the radius of the Earth,

ϕ is the angular width of the duct,
D is the distance from the injection point.

This analysis was then combined with a mirror type calculation assuming concentric spherical mirrors separated by both 30 and 300 km (Figure 7). Multiple rays were then launched, simulating for the H = 30 km guide a ducted mode where the launch angle varied between 85 and 89 degrees. For H = 300 km, a ground based antenna was modeled and launch angles, measured from vertical, of 70 to 85 degrees was used. The general character of these curves are the same showing antipodal and round the world focusing and R^{-2} dependence for $D < 10{,}000$ km. The major difference between the two cases is the number of modes at a given distance. Using the mirror model it is possible to count the number of modes available at any selected distance.

At a distance of 10 000 km and H = 30 the number of modes is 10 and increases to 18 at 20,000 km compared to 4 and 7 for the same two distances when H = 300 km. The small number of modes for the H = 300 km case is consistent with observations of F-region propagation. The large number of modes for ducted propagation (H = 30 km) indicates an expected complex structure to these ducted signals.

To summarize the theoretical work, the existence of ducted modes has been shown to be plausible on the basis of ray tracing and potential well calculations. In the daytime it is the region between F1 and F2 layers that is the most likely ducting region though the duct is small at best. At night these modes very easily detach themselves and become chordal until they reach the daytime hemisphere again. Under optimum conditions, the absorption losses can be kept very small, often under 5 dB for a 10,000 km path.

REFERENCES

Barghausen, A.F., et al.,	*Predicting Long-term Operational Parameters of High Frequency Skywave Telecommunication Systems*, ESSA Technical Report ERL 110-ITS, p.78, 1969.
Booker, H.G.,	*A Theory of Scattering by Nonisotropic Irregularities with Application to Radar Reflections from the Aurora*, J. Atmosph. Terr. Phys. Vol.8, p.204, 1956.
Carrol, J.C., et al	*The Platteville High Power Facility*, Radio Science Vol.9, p.889, 1974.
Gurevich, A.V.,	*Effect of Nonlinearity on the Generation of Circumterrestial Signals*, Geomagn i Aeronomiya, Vol.11, p.810, 1971.
Gurevich, A.V., Tsedelina, Ye.Ye.,	*Contribution to the Theory of Ultralong Propagation of Short Radio Waves*, Geomagn. i Aeronomiya, Vol.13, p.242, 1973.
Matyugin, S.H.,	*Influence of Ionosphere Parameters on Radio Wave Absorption in an Ionospheric wave Duct*, Radiofizika, Vol.18, p.1190, 1975
Utlant, W.F., et al.,	*A summary of Vertical Incidence Radio Observations of Ionospheric Modification*, Radio Science, Vol 9, p.895, 1974.

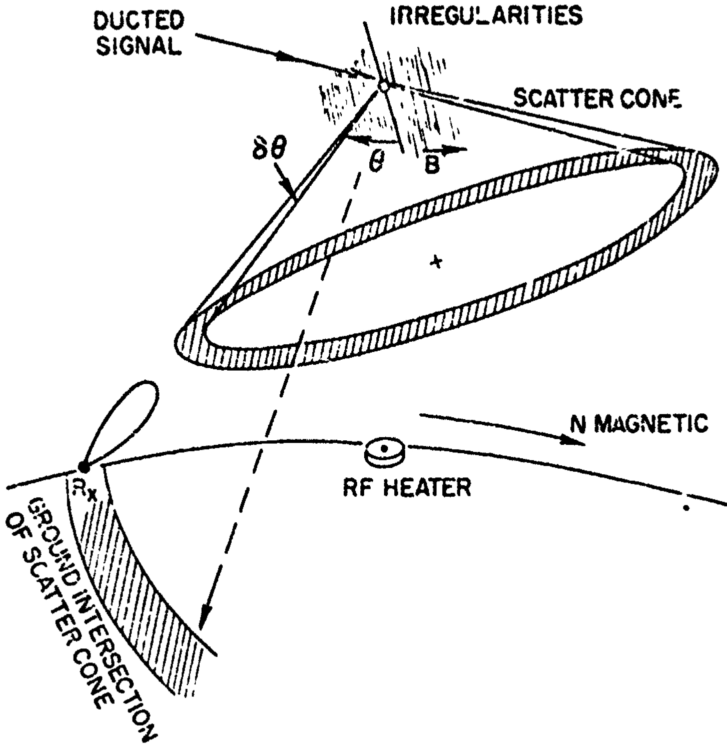

DUCTED SIGNAL
IRREGULARITIES
SCATTER CONE
δθ
θ
B
N MAGNETIC
RF HEATER
Rx
GROUND INTERSECTION OF SCATTER CONE

Fig.1 Intersection of scattering cone with the earth's surface. The shaded region on the ground is the region of interception of ducted signals after scattering

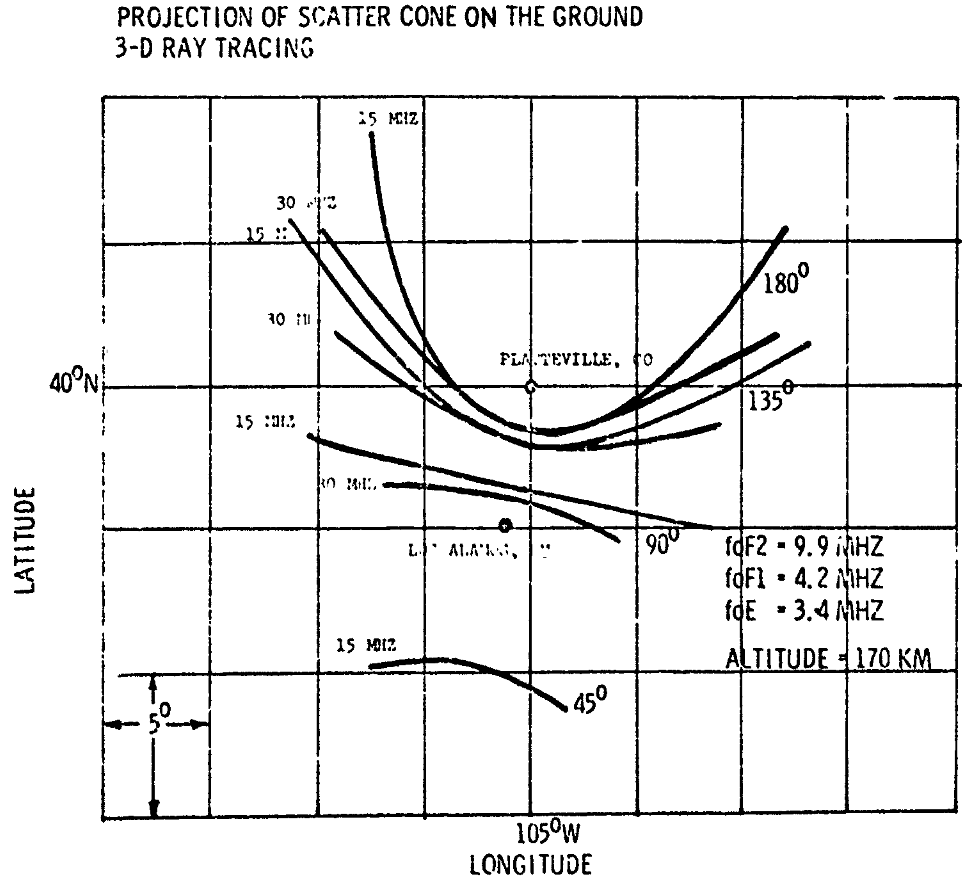

PROJECTION OF SCATTER CONE ON THE GROUND
3-D RAY TRACING

Fig.2 Projection on the ground of scatter cone after ionospheric refraction with heated Platteville region at 170 km altitude. Frequencies are 15 MHz and 30 MHz, and the numbers on the right of the curves are the azimuths of ducted propagation

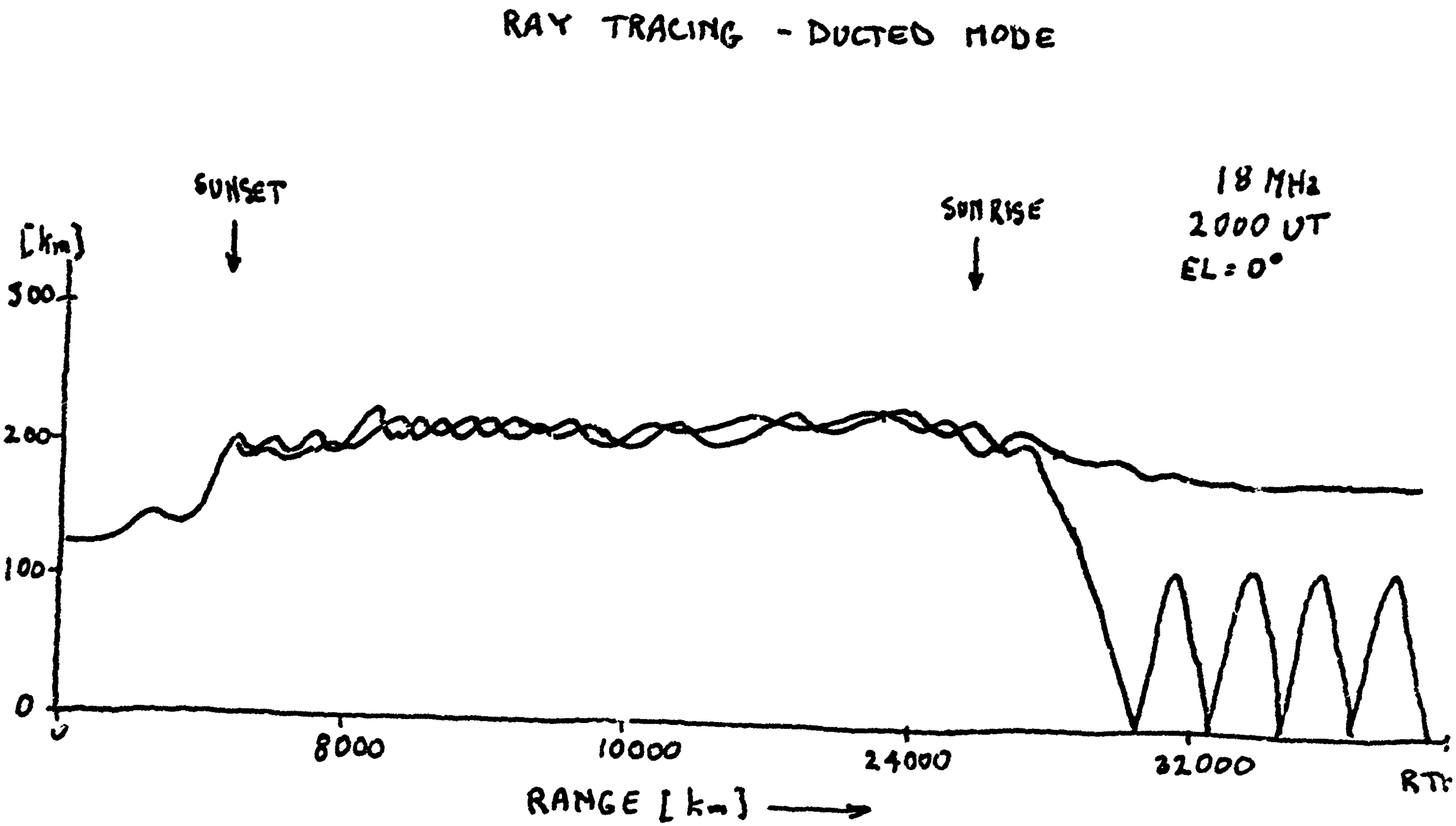


Fig.3 Simulation of long range ducted mode

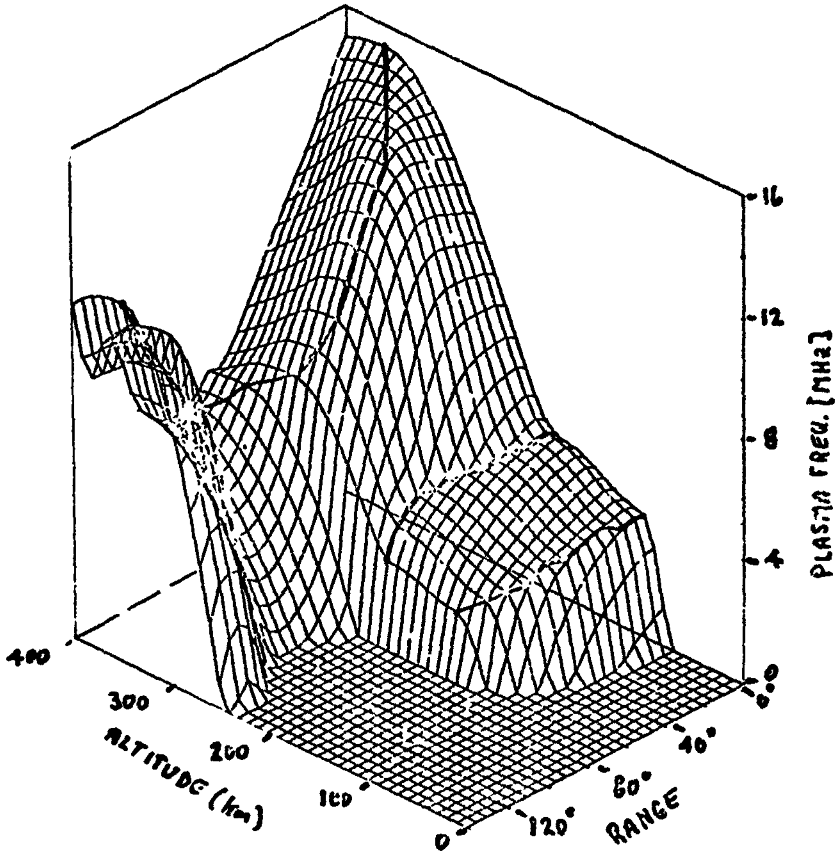


Fig.4a Three dimensional plot of plasma frequency vs range vs altitude

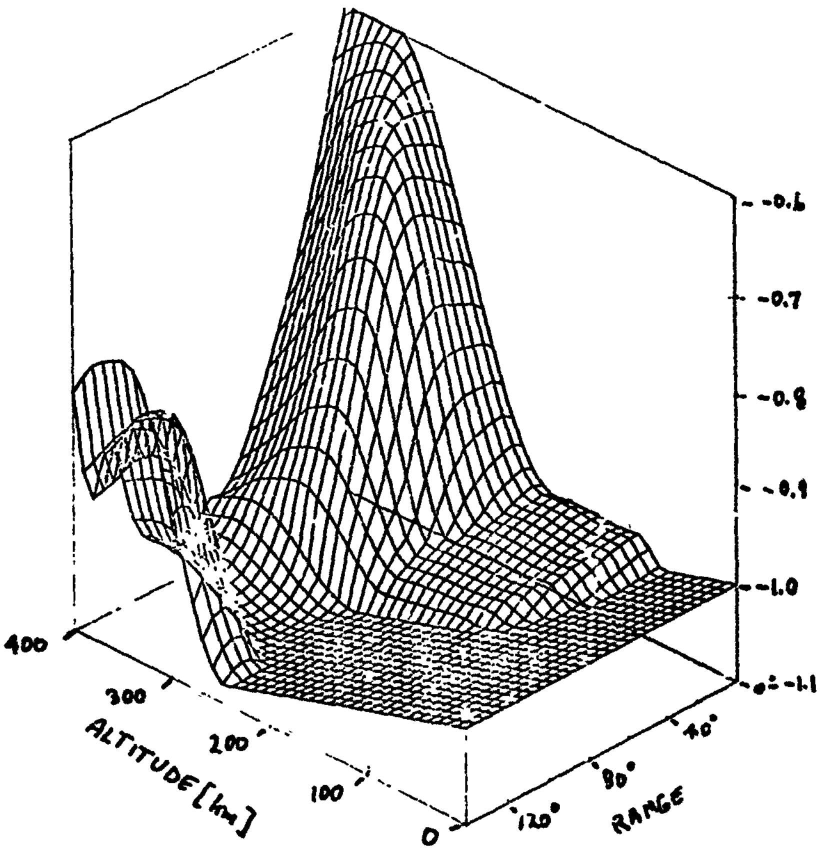


Fig.4b Three dimensional plot of potential vs range vs altitude

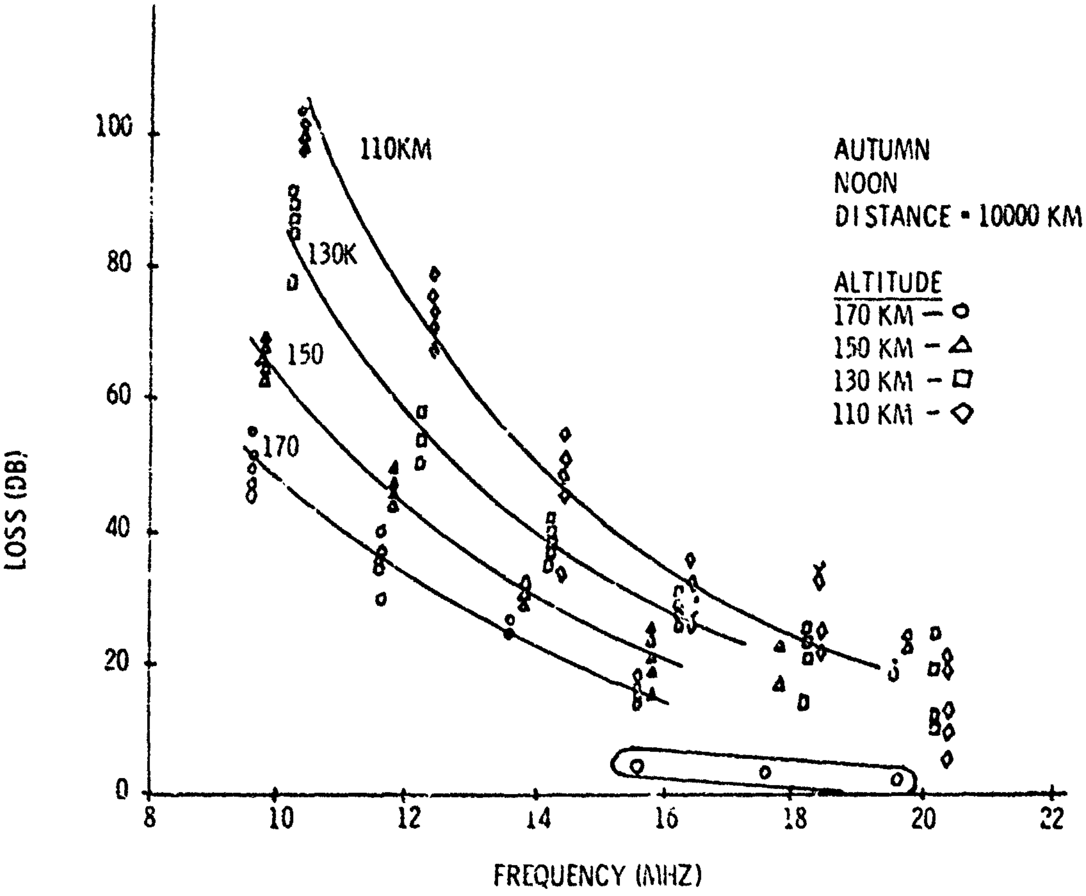


Fig.5 Propagation losses computed for a fixed range of 10,000 km and four heights of heated region. Note the points outlined between 15 MHz and 20 MHz, which correspond to ducted modes

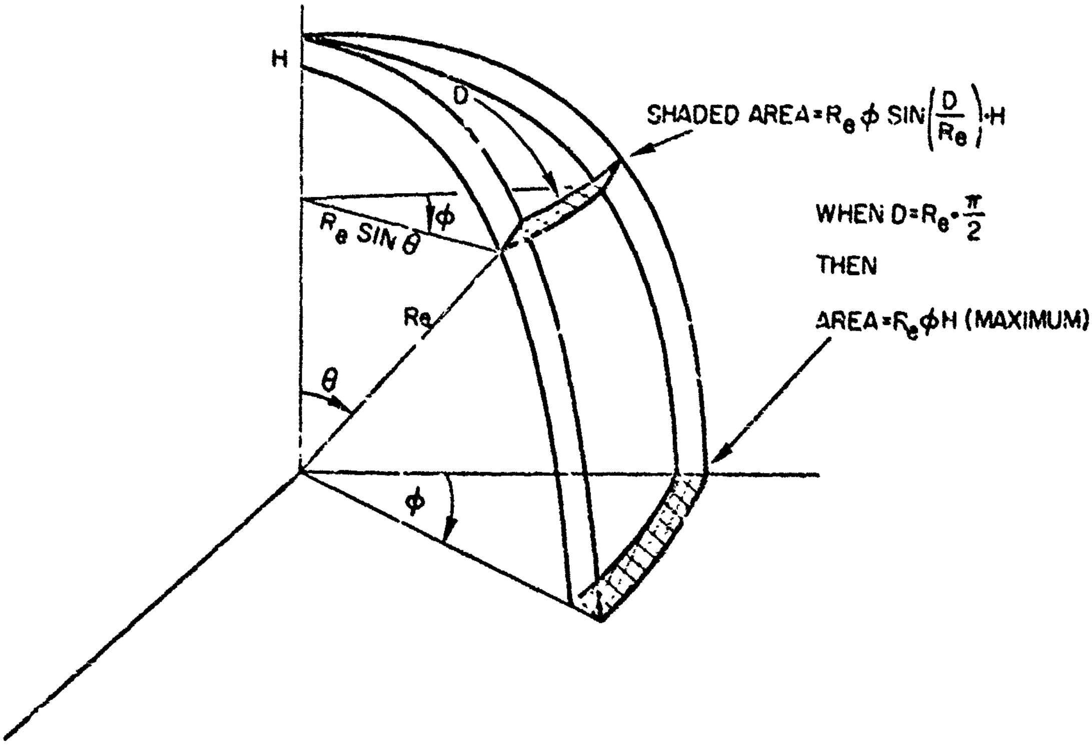
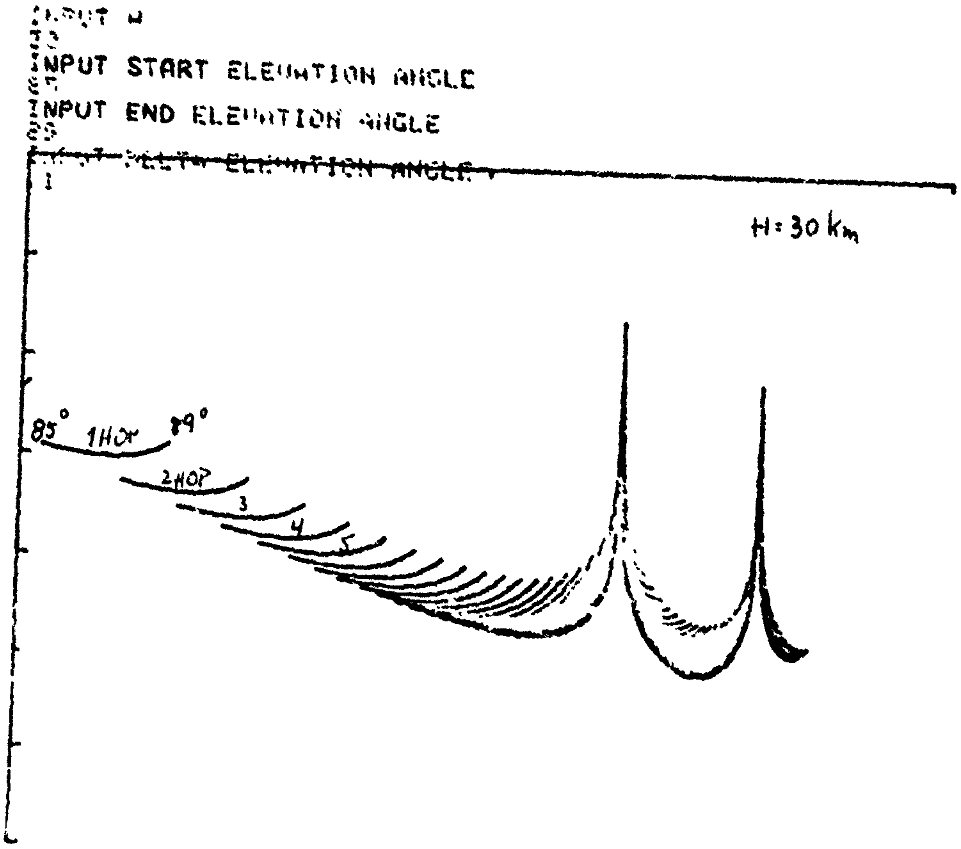
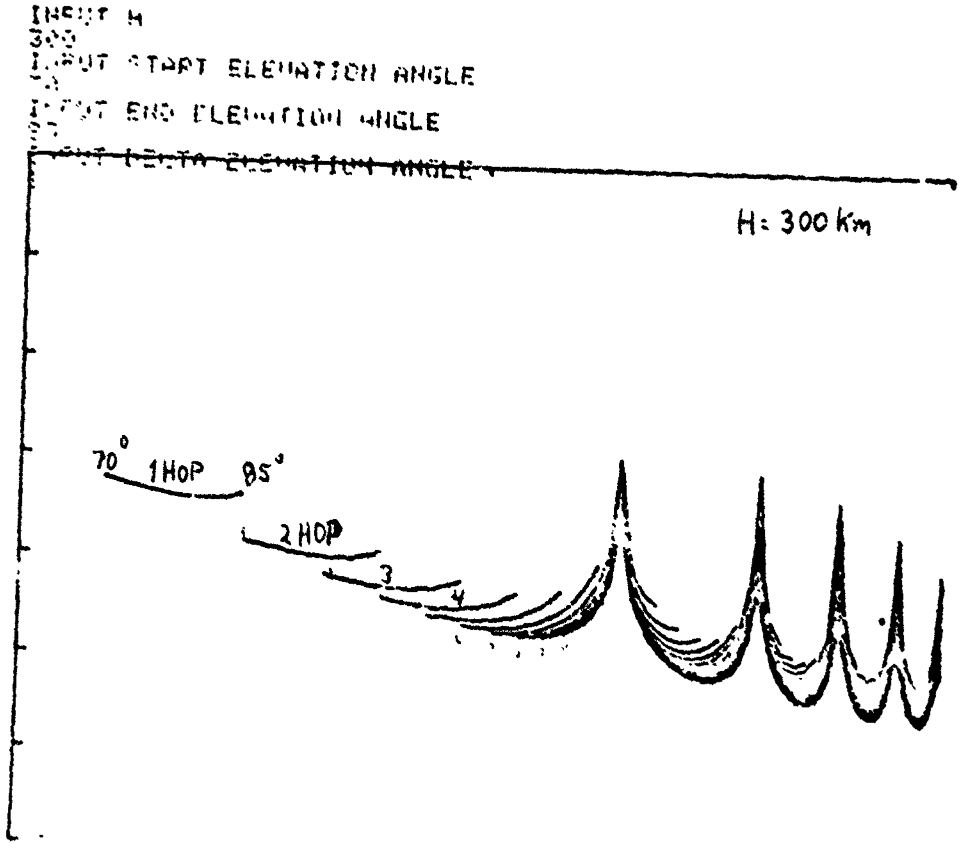

H
D
SHADED AREA = $R_e \phi \sin\left(\frac{D}{R_e}\right) \cdot H$
WHEN $D = R_e \cdot \frac{\pi}{2}$
THEN
AREA = $R_e \phi H$ (MAXIMUM)
$R_e \sin\theta$
ϕ
R_e
θ
ϕ

Fig.6 Illustrating azimuthal spreading in the duct and showing that maximum spreading occurs at a range of one quarter of the earth's circumference



```
INPUT H
30
INPUT START ELEVATION ANGLE
85
INPUT END ELEVATION ANGLE
89
INPUT DELTA ELEVATION ANGLE
1
```



```
INPUT H
300
INPUT START ELEVATION ANGLE
70
INPUT END ELEVATION ANGLE
85
INPUT DELTA ELEVATION ANGLE
```

Fig.7b Simulation of ducted mode structure. H = 300 km

DISCUSSION

H. Poeverlein, Ge

You had in mind communication between ground and an object in the duct. Did you also think of ground-to-ground communication by means of a natural way of feeding in or out at one end of the duct?

Author's Reply

We recognize the possibility of the use of naturally occurring ionospheric gradients as a means of ducted mode coupling, and in fact are using this technique presently to inject energy into ducted modes, which is subsequently rejected by means of a heater/modifier. The use of this technique in a practical communications system seems to be limited by the availability of suitable gradients.

E D.K. Shearman, UK

The proposal is put forward that this injection scheme be used for HF radar. In this case one would wish to use a narrow beam to determine the azimuth of the target, but the conical scattering property of the field-aligned irregularities will spread the radiated energy over a section depending on the ability to launch ducted modes. Will this not considerably degrade the azimuthal resolution of the radar?

Author's Reply

It is quite true that a radar of this kind would have poor azimuthal resolution, although the "beam" can be steered to some degree by varying heating altitude and steering the heater beam with respect to the zenith. Good azimuthal target estimates could conceivably be obtained by triangulation using two ducted mode radars, widely spaced.

Comment by N.C. Gerson, US

It should be noted that pairs of natural refractivity discontinuities in the ionosphere exists;these can be used to inject an HF ray, into, and eject it out of the ionospheric duct. These pairs are (a) the sunrise-sunset zone, (b) the two auroral zones and (c) the high electron density regions 15° (alpra) of the magnetic equator. Experiments have been made at HF in all three cases, attempting to inject and eject HF. In all three cases, low loss propagation, suggestive of a low loss channel, was found. For the antipodal type of propagation the predominant path was through the dark ionosphere, with paths spread through 90° around the sub-solar nadir point. It might be noted that the antipodal ($^1/_2$), round the world ($^2/_2$) and circumferential paths have already been investigated and in all cases relatively low path losses were encountered.

Author's Comment

These natural ionospheric discontinuities are certainly effective in ducted mode coupling but cannot be relied upon in the application envisaged, due to the spatial and temporal limitations involved. The antipodal and RTW focussing are indicated in Figure 7.

E.R. Schmerling, US

Have you computed the losses in the vicinity of the injection region resulting from the non-linear processes that produce enhanced collisions as a result of the heater transmitter? These should, presumably, be added to your path loss computations.

Author's Reply

Losses of this kind, who nature is not well understood and which are therefore difficult to estimate have not been taken into account explicitly. Rather, they will be included in effective scatter cross-section measurements, currently being made at HF, using heater-induced scatter

C.S. Goutelard, Fr

Deux paramètres importants interviennent dans les problèmes de détection: La localisation de l'engin dans l'espace dont ont parlé les précédents interlocuteurs et la mesure de la vitesse radiale qui se fait par l'étude du spectre doppler.

Dans les modes piégés lors des transitions jour nuit que nous observons depuis 15 ans, nous constatons, même sur de très grandes distances 10.000 à 15 000 kms, une bonne cohérance de l'effet doppler. Dans le cas de piégeage par chauffage de l'ionosphère, outre l'effet des modes multiples que vous signalez, le fait que les rayons traversent une zone perturbée devrait entrainer un étalement du spectre doppler. Au cours des mesures que vous avez faites ou des simulations avez-vous pu étudier ce phénomène et dans ce cas pouvez-vous nous communiquer vos conclusions?

Author's Reply

The Doppler broadening of the heater-scattered signal is of the order of a few Hertz, as reported in the literature a few years ago. This certainly limits the precision with which the radial component of low velocity targets can be measured, but would not be a major limitation for high velocity targets.

EFFETS IONOSPHERIQUES D'UNE ECLIPSE SOLAIRE DANS LES ILES DU CAP VERT

Par A.S. MENDES, F. BERTIN, P. VILA et J. PAPET-LEPINE
Service Météorologique National (PORTUGAL)
Centre National d'Etudes des Télécommunications (FRANCE)

SUMMARY

The installation of two ionosondes at SAL (Cape Verde, Islands) and at Ascension Island (conjugate point to SAL) as well as three polarimeters at SAL, DAKAR and BATHURST (Gambia) in addition to the permanent ionosonde in DAKAR has made it possible to detect various ionospheric effects of the eclipse

The passage of the entire eclipse at a supersonic speed seems to have produced a gravity wave front due to the cooling. The simultaneous foF2 and total electron content minima determine the initial time for the passage of a gravity-wave-like wake with a velocity around 250 ms^{-1} and whose fundamental period is 10 mn.

The ionospheric electric field perturbation has been studied. These are due to the ionisation loss in the E region dynamo in a cylindrical region of 500 km radius. As a result of the upper F region effect, a 80 km lifting takes place from 10 : 25 to 10 : 40 UT followed by a down-going transport from 10 : 45 to 11 : 00 UT without any ionisation loss at these levels.

A thermospheric contraction flux appearing 20 mn after the eclipse maximum as well as a forced thermal diffusion flux were also noted.

The above mentioned results have been essential for interpreting the more complex variations observed at the subtropical stations of OUAGADOUGOU and IBADAN as at the equatorial station of SARH.

I - INTRODUCTION

Il s'agit essentiellement d'observations ionosphériques, ayant pour but l'étude des perturbations de la région F tropicale, d'une part, et d'autre part, de l'étude des ondes de gravité de moyenne échelle engendrées par l'éclipse.

L'éclipse a eu lieu au solstice de Juin en période de faible activité solaire. Dans ces conditions, le vent neutre thermosphérique impose régulièrement au cours de la matinée une forte dissymétrie de l'ionisation de la région F en fonction de la latitude, entre le tropique Nord (faible maximum relatif) et le tropique Sud (maximum intense).

1 - Dans sa phase matinale, l'éclipse atteint sa totalité dans l'ionosphère du Cap Vert. C'est dans le méridien magnétique correspondant que le contraste de photo-ionisation d'éclipse le plus intense sera observable, entre cet archipel au tropique Nord (Ile de Sal, 23° Ouest, 17° Nord géographique, Archipel du Cap Vert), magnitude d'obscuration 1,07) et le tropique Sud (Ile Ascension, 14° Ouest, 8° Sud, magnitude 0,35). Ces deux points sont magnétiquement conjugués, compte tenu de la position de l'équateur magnétique à cette longitude.

Cette configuration exceptionnelle permet d'isoler l'effet de diffusion transéquatoriale d'ionisation dû au gradient thermique dans la région F2 intertropicale (chauffage des photo-électrons). Cet effet n'apparaît que sous forme transitoire en temps normal, bien qu'il intervienne probablement en permanence.

Pour cette expérience, un sondeur LNR-SP 35-16 a été installé à l'Ile Ascension et un sondeur COSSOR à l'Ile de Sal (Archipel du Cap Vert).

2 - Les perturbations à moyenne et petite échelle ont été étudiées à partir d'un système de stations ionosphériques comprenant deux polarimètres recevant les signaux à 136 MHz du satellite Intelsat II F3 situés à Dakar (Sénégal) et Bathurst (Gambie) et deux sondeurs ionosphériques verticaux situés à Dakar et Sal (Iles du Cap Vert). Plus près du sol, les effets de l'éclipse ont été mesurés par deux microbarographes placés l'un à Sal l'autre à Dakar.

3 - Outre Dakar, les stations ionosphériques africaines gérées par le CNET (Ouagadougou, Haute Volta ; Sahr, Tchad ; Djibouti) effectuent des observations systématiques renforcées, recommandées par les normes internationales pendant les éclipses.

La réalisation pratique de ces différentes expériences a été faite par le Groupe de Recherches Ionosphériques et le département Mesures Ionosphériques et Radioélectriques du CNET qui a, par ailleurs, obtenu le concours des autorités britanniques pour les installations de l'Ile Ascension et des autorités portugaises (Service des Transmissions et Service de la Météorologie) pour les installations dans les Iles du Cap Vert. Enfin, l'Université du Pays de Galles, partie prenante en ce qui concerne les ondes de gravité, a participé à ces expériences en fournissant les polarimètres et en s'occupant de la station de Bathurst.

II - ONDE DE GRAVITE DUE AU REFROIDISSEMENT

Il a été possible de mettre en évidence une onde de gravité de période 13 minutes (± 3 min) ayant une vitesse voisine de 750 m/s (± 30 m/s) et une longueur d'onde horizontale de 800 km. D'autre part, le dépouillement du microbarographe de Dakar montre également une onde de période 20 minutes environ, ce qui serait du même ordre de grandeur que celle mesurée au niveau ionosphérique. Dans ces conditions, il est possible de suggérer qu'on a affaire à une onde du type de Lamb compte tenu de la vitesse de phase mesurée. Si maintenant on tient compte de la géométrie du système éclipse -stations de sondages (4 stations : DAKAR, ABUKO, OUAGADOUGOU, TAMANRASSET), on montre que dans tous les cas l'angle entre la vitesse de propagation de l'onde et la vitesse de propagation de l'éclipse est de 60°. On en déduit que la vitesse de phase vaut la moitié de la vitesse de l'éclipse.

Le passage de la totalité de l'éclipse à vitesse supersonique produit donc un front d'onde de gravité qui montre la bonne cohérence obtenue entre les résultats expérimentaux et les résultats théoriques de CHIMONAS (1973). En particulier, si les résultats concernant la phase, la vitesse et la période sont parfaitement cohérents avec la théorie de Chimonas, il n'en est pas de même pour l'amplitude de l'onde (BERTIN, HUGHES et KERSLEY, 1977).

III - REPONSE DE LA REGION F2 TROPICALE A L'ECLIPSE SOLAIRE DU 30 JUIN 1973

L'évolution de foF2, lors de l'éclipse, ressemble de façon surprenante à celle des 28 et 29 Juin (jours de contrôle) ; la compensation par plusieurs mécanismes semble avoir troublé les différents processus physiques en évolution.

Une étude plus approfondie des données disponibles de N (h) en fonction de temps et leur distribution spatiale permet de distinguer plusieurs phases .

Le tableau I résume ces phases :

- jusqu'à 12.00 TU, on observe le jour de l'eclipse une augmentation de l'ionisation comparable à celle de la période de contrôle,
- De 12.00 à 12.15, une phase où l'ionisation s'accroit rapidement avec une chute sensible de hauteur aux 2 tropiques (alors que hmF2 reste constante les jours de contrôle),
- Enfin de 12.15 à 12.45, le tropique Nord émergeant de l'ombre de l'éclipse reçoit une quantité importante d'ionisation (qui ne peut provenir que des niveaux supérieurs de F2 puisque la couche F2 inférieure a été vidée et contractée par l'éclipse). La région F2 au sud à Ascension a été vidée de façon intensive, et c'est ce phénomène qui nous semble le plus significatif.

Les effets du pic F2 particulier à l'éclipse furent retardé de plus de 60 minutes le 30 Juin 1973 par 2 mécanismes, d'abord, le régulier "creux tropical d'été" qui vide l'ionisation de F2 au dessus du Cap Vert et de Dakar jusqu'à 10.30 TU ; deuxièmement, la perturbation magnétique globale qui empêche le processus de fontaine, laissant un dôme d'ionisation au dessus de l'équateur magnétique, minimisant ainsi les gradients habituels d'ionisation aux niveaux supérieurs de la couche F2 subtropicale : dans le tableau I, seule l'augmentation rapide de la densité du pic F2 de 12.00 à 12.15 peut être attribuée clairement à un sursaut du mecanisme de fontaine (Dunford, 1970) alors qu'elle atteint les 2 zones tropicales simultanément.

Il semble vraisemblable qu'à 12.15 TU l'ionisation a éte suffisamment entrainée par le processus de fontaine de l'équateur magnétique vers les zones tropicales N et S en provoquant des gradients d'ionisation subtropicale à des niveaux supérieurs. La configuration assymétrique de l'ionisation des zones tropicales N et S devient propice à un échauffement assymétrique conjugué par photoélectrons, spécialement ceux du Sud : par conséquent, la région N émerge juste de l'ombre de l'éclipse, commence à accumuler l'ionisation au niveau du pic F2 au moyen de diffusion thermique alors que le tropique Sud est intensément vidé (VILA, 1971).

C'est un flux transéquatorial d'ionisation de la zone S non éclipsée vers le tropique N précédemment éclipsé qui semble être la meilleure explication de ce changement rapide d'ionisation dans les deux tropiques. A cause de cette succession compliquée de perturbations qui l'a précédé et aucune orbite de satellite Isis n'entrecoupant la région de l'éclipse, notre interprétation a besoin d'être confirmée par une analyse plus complète.

Toutes les données disponibles concernant le contenu total d'électrons, le champ géomagnétique les EUV solaires et la spectrométrie de rayons X (y compris la couronne), la composition ionique, le flux photoélectronique et la température des différents composants permettront de faire des simulations réalistes de variation dans le temps.

IV - PERTURBATION DE CHAMP ELECTRIQUE

Cette perturbation est due à la perte d'ionisation dans la région E dynamo (100 km d'altitude) sur un espace cylindrique d'environ 500 km de rayon.

L'effet sur la région F supérieure est un transport ascendant sur 80 km de 10h25 à 10h40 TU, puis descendant de 10h40 à 11h00 TU, sans perte d'ionisation à ces niveaux.

Les deux effets de perturbation rapides qui précèdent n'avaient jamais été mis en évidence, et les données de SAL ont été indispensables à leur observation (VILA, BLANC et BERTIN, 1976).

V - FLUX DE CONTRACTION THERMOSPHERIQUE

Vingt minutes après le maximum de l'éclipse un accroissement progressif d'ionisation du haut vers le bas des couches F2 et F1 manifeste un transport d'ionisation dû à la contraction atmosphérique de l'ensemble encore refroidi par l'éclipse (VILA, BLANC et BERTIN, 1976).

Ce phénomène de vitesse lente ($\bar{v}_c = 50\ ms^{-1}$) se poursuit jusqu'à environ 12.30 TU. Il produit dans la région F supérieure au dessus de SAL un gradient d'ionisation interne vers le Sud le long des lignes de force du dessus magnétique (figure 1).

VI - FLUX DE DIFFUSION THERMIQUE FORCEE

Le gradient de densité électronique précédent subit le chauffage des photoélectrons émis dans l'ionosphère conjuguée non éclipsée. La thermalisation de ces photoélectrons créé un gradient de température électronique le long de la ligne de force magnétique, et orienté au sud, qui déplace le plasma vers le bas et vers le nord, à une vitesse moyenne de 250 ms^{-1} (figure 1). Par suite, l'ionisation afflue de 12.15 à 13.00 TU vers le maximum d'ionisation dont le taux d'accroissement moyen est 300 $cm^{-3}\ s^{-1}$.

Les résultats précédents ont été indispensables pour l'interprétation des variations, plus complexes, qui ont été observées au dessus des stations subtropicales de OUAGADOUGOU et IBADAN, et au dessus de la station Equatoriale de SARH.

VII - CONCLUSION

Les eclipses solaires restent des évènements d'un très grand intérêt pour l'étude de l'ionosphère dans toutes ces régions. En effet, elles font apparaître un grand nombre de phénomère photo chimique, dynamique et électrodynamique de façon très nette en fonction des phases spatiales et temporelles de l'éclipse. L'étude de ces phénomènes permettra des progrès importants en géophysique externe.

BIBLIOGRAPHIE -

BERTIN F., HIGHES K.A., KERSLEY L., 1977, "Atmospheric waves induced by the solar eclipse of 30 June 73" J.A.T.P. 39; 457-462.

CHIMONAS G., 1973, "Lamb waves generated by the 1970 solar eclipse" Planetary and space Science, 21, 1843-1854.

DUNFORD E., 1970, "Electric field and F region electron densities over Peru" J. Atmos. Terr. Phys. 32, 421-425.

VILA P., 1971, "Intertropical F2 ionization during June and July 1966", Radio Sci. 6, 7, 689-697.

VILA P., 1971, "New dynamical aspects of intertropical ionization, Radio Sci." 6, 11, 945-957.

VILA P., BLANC M., BERTIN F., 1976, "Réponse de l'ionosphère à l'éclipse solaire du 30 Juin 1973", document de travail CRPE/1008, CNET-CNRS.

LEGENDES DE LA FIGURE ET DU TABLEAU

Figure 1 - Schéma pr[illegible]e pour [illegible]es perturbations méridiennes transéquatoriales. V_c, vitesse de contraction [illegible] on thermosphérique). Tireté, contours isoioniques et gradient de densité électron[illegible] [illegible]curs fléché, flux de photoélectrons conjugués. Trait continu, isothermes électroniques et vitesse de diffusion thermique forcée.

Tableau 1 - Variation des paramètres de la couche F2 à la fin de l'éclipse solaire du 30 Juin 1973 en comparaison avec les jours de contrôle.

$\frac{dNm}{dt}$: valeur moyenne en densité du taux de variation du maximum d'ionisation de la région F2 (en $cm^{-3}\ s^{-1}$) donné dans les intervalles de temps indiqué sur le tableau ; les valeurs de hmF2 correspondantes (en km) sont soulignées.

Processus principal	TU	Tropique Nord				Tropique Sud	
		SAL		DAKAR		ASCENSION	
		Contrôle	30 Juin	Contrôle	30 Juin	Contrôle	30 Juin
La Gouttière du 30 Juin	1100	335	338	280	260	250	335
Contraction par l'éclipse		+ 67	+ 40	+ 45	+ 63	+ 74	+ 75
	1140	360	260	460	400	250	325
Effet de fontaine		+ 17	+ 60	+ 12	+ 17	- 120	+ 22
	1200	360	286	450	320	250	305
		- 15	+ 90	+ 100	+ 220	- 160	+ 300
	1215	360	270	445	310	250	285
Flux Transéquatorial		- 50	+ 116	+ 47	+ 112	0	- 170
	1245	365	290	420	305	290	270

TABLEAU 1

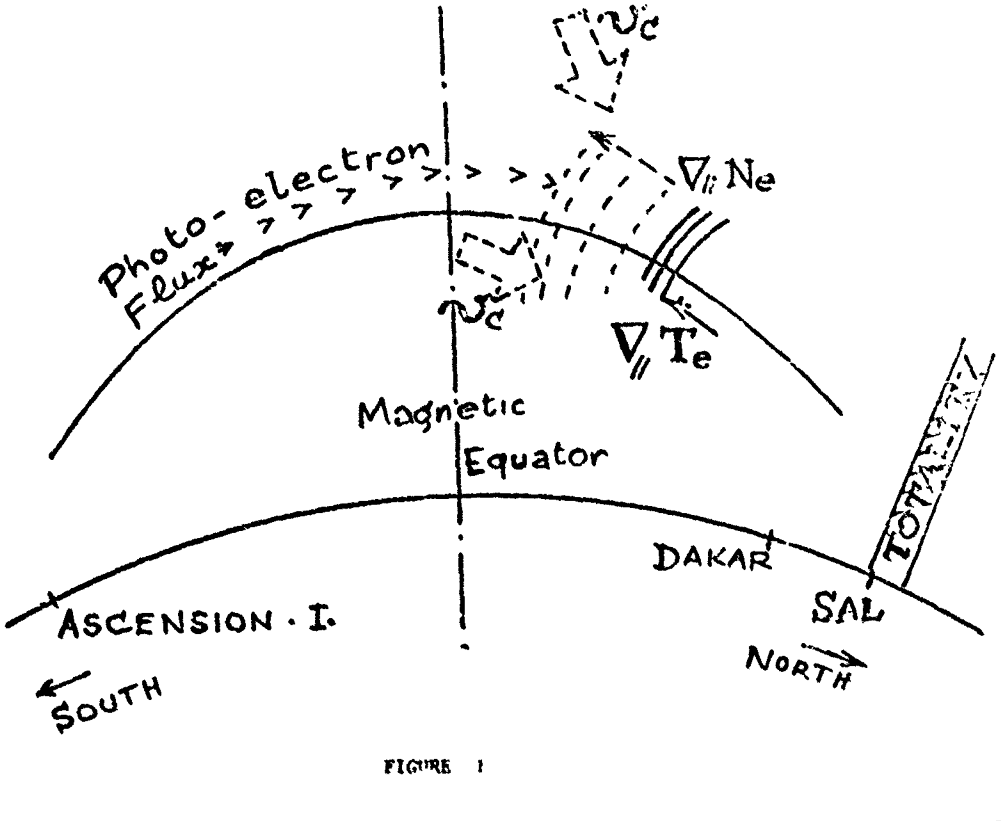


FIGURE 1

GROUND-WAVE AND SKY-WAVE SEA-STATE SENSING EXPERIMENTS IN THE U.K.

E. D. R. Shearman, W. A. Sandham,
Department of Electronic and Electrical Engineering,
University of Birmingham, P.O. Box 363, Birmingham 15, U.K.

E. N. Bramley and P. A. Bradley,
Science Research Council, Appleton Laboratory, Ditton Park, Slough, Berks., U.K.

SUMMARY

A study of the potential of both ground-wave MF/HF radar and sky-wave HF radar for sensing wave characteristics and surface current and their application in the U.K. began in 1975 Ground-wave experiments have included non-directional observations of 1.95 MHz sea echo (arising by bragg diffraction from 77 m sea waves) received over a 180° sector for a wide variety of sea-state and surface wind conditions. A marked dependence of echo amplitude on sea-state was noted for these 77 m waves as a function of the wind-dependent long-wavelength cut-off of the sea spectrum. The 0.01 Hz Doppler resolution employed, permitted identification of island and ship echoes, ships being tracked out to 210 km range. Using synthetic-aperture technique, wave directional features were established and correlated with wind data.

In 1978, a sky-wave radar project involving on-line digital processing was initiated, aimed at routine daily monitoring of the North Atlantic at ranges from 1,000-3,000 km over a 60° arc. A preliminary evaluation of the techniques was carried out in Summer 1978 during the JASIN (Joint Air-Sea Interaction) project in the North Atlantic. Comparisons are reported of radar-deduced surface wind directions and magnitudes with those measured in oceanographic vessels. An extension of the statistical theory used in the analysis of radar spectra is presented.

1. INTRODUCTION

Research in the U.K. on the techniques and potential applications of MF/HF radar for sea-state sensing began at the University of Birmingham in 1975, building on previous research on HF sky-wave radar for remote-sensing of the ionosphere.

The initial study of sky-wave radar sea-state sensing (SHEARMAN, BAGWELL and SANDHAM, 1977), led to two programmes of work which are described in the present paper. The first programme arose from a realisation that the basic mechanisms involved in radio-wave scattering from the sea could best be studied without the complicating influence of ionospheric propagation: such influence could be avoided by the use of a coastal radar station and ground-wave propagation, the technique used in the classic studies of D. D. Crombie (CROMBIE, 1955). It was also realised that for a country such as the U.K., with a major interest in the sea conditions over the surrounding continental shelf, ground-wave radar had great merits in itself as a remote sensor with a range of up to 300 km.

In the second part of a joint project with the Appleton Laboratory of the U.K. Science Research Council the sky-wave radar used in the earlier work is being equipped with on-line computing facilities for the survey on a routine basis of an area of the North Atlantic from 1,000-3,000 km in range and 60° wide in azimuth. This work commenced in 1978, and in the Summer of 1978 advantage was taken of the international oceanographic project JASIN (Joint Air-Sea Interaction) in the Rockall Bank area of the North Atlantic to compare HF sky-wave radar measurements with sea truth obtained from the oceanographic vessels.

This paper describes the first results of the ground-wave programme, the plans for the sky-wave programme and the preliminary results and analysis of the JASIN project.

2. GROUND-WAVE RADAR EXPERIMENTS

The location of the ground-wave radar site, which utilises a former LORAN A navigational aid transmitter, is shown in Fig. 1. The transmissions, at a frequency of 1.95 MHz, are radiated from an omnidirectional antenna, but ground-wave attenuation over land limits the effective azimuthal coverage to about 210°. Extra ground losses outside this arc are shown in the figure. The transmitter generates 50 μs, 150 kW pulses at a repetition rate of 305/9 pulses per second and both carrier frequency and repetition rate are referred to a highly stable crystal standard. A preliminary report of the experiments has been given elsewhere (SANDHAM, SHEARMAN and BAGWELL, 1978).

If the received echoes are sampled at a chosen time-delay after the transmitted pulse, the echoes contributing will be those arising from an annulus whose radius is defined by the time-delay and whose thickness (the range-resolution) is defined by the pulse duration, 50 μs in this case corresponding to a range resolution of 7.5 km. The ground areas will give rise to a zero Doppler-shifted echo, but the sea will have a characteristic spectrum. From the chaotic wind-driven wave pattern on the sea-surface, the electromagnetic scattering mechanism selects those sea-waves with a length of half a radio wavelength travelling radially outward from, or inward to, the radar. These waves give a resonant first-order back-scatter and the scatter from waves of other lengths and directions is by second-order mechanisms giving much lower amplitude.

Since the selected radially-travelling sea waves have a characteristic velocity, they give rise to discrete Doppler shifts, positive and negative respectively for the approaching and receding waves as discussed in an accompanying paper (BROCHE, 1979). For a radio wavelength λ_0, the resulting first order 'Bragg resonant' Doppler shift is,

$$f_B^{(1)} = \pm (g/\pi\lambda_0)^{\frac{1}{2}}$$

where g is the acceleration due to gravity.

For the 1.95 MHz radar, the Doppler shift $f_B^{(1)}$ is ± 0.142 Hz and the length of the sea waves selected is 77 m.

Two forms of processing are employed; the first is for on-line analysis of a single, selected range-gate and uses a commercial digital-store/sweeping filter spectrum analyser with an XY plotter output; the second is for off-line analysis and uses analogue tape recording, digitisation on replay and multi range-gate spectrum analysis in a main-frame computer. The typical coherent dwell time used is 150 s, which with a Hanning (raised cosine) weighting function yields a spectral resolution of 0.01 Hz. In each the input to the spectrum analyser from the receiver is centred at a frequency of 8 Hz to avoid spectral folding.

Figure 2 shows on-line spectra from a range of 75 km on two days, 'A' for a day when the sea was relatively calm and the average wind speed was less than 5 knots. The 77 m sea waves were therefore very low in magnitude, with the approach and recede Doppler lines being 10 dB and 25 dB respectively below the ground echo at the same range. In spectrum 'B' however, the wind had been blowing at 20 knots long enough for 77 m waves to become fully developed and so the Bragg lines were respectively at + 7 dB and 0 dB relative to the ground echo.

Figure 3 shows, in 3 dimensional format, a complete set of spectra observed at 15 km range increments from the same site. A threshold has been set at - 25 dB to display a noise-free floor. Features of interest are the variation of ground echo amplitude with range, with a sharp peak at the range of Lundy Island in the Bristol Channel (see Fig. 1), the variation of sea-wave echo amplitude with range and the presence of ship echoes. The discrete echo at 30 km range and negative Doppler shift of 0.08 Hz is a ship echo of the kind from which tracks have been extracted. The Doppler spread echo at 45 km range is of the nature expected from a ship turning within the 2½ minute coherent dwell time.

Figure 4 shows a radial-range time track extracted from an identified tanker of 18,000 tonnes gross, outward bound from Milford Haven. The small rectangles identify by their side-lengths the range-gate width and the coherent dwell time, respectively, defining the accuracy of location. The arrows indicate the predicted radial velocity of the target from the Doppler shift. A good match was obtained from the tanker track out to a range of 120 km and further identification was made of an approaching tanker as indicated on the diagram.

The maximum range achieved to date is 210 km, with a mean power of only 250 W and omnidirectional transmit and receive antennas. The discrete form of ship echoes when resolved by very fine resolution Doppler filtering as employed here, suggests that direction-finding technique could be used for directional determination. The experiments suggest that ship-tracking beyond microwave radar range may be possible with a much simpler ground-wave radar installation than might be envisaged if a large beam-forming array were to be used.

Such direction-finding techniques cannot however be used to map sea clutter as a function of direction since the scatterers are continuously distributed. An array capable of achieving resolution is needed, but a simple realisation of such an array was achieved by the synthetic aperture technique of Teague et al. (TEAGUE, TYLER and STEWART, 1975). The echoes from the 1.95 MHz transmitter were received in a vehicle, using a receiver with a very stable local oscillator, and recorded as the vehicle was driven at constant speed along a straight road. The direction-dependent Doppler shift in conjunction with the discrete line structure of the echo spectrum permitted mapping of the echo amplitude vs direction as shown in Fig. 5. For a homogeneous wind field, such observations yield the directional sea-wave energy spectrum. This assumption could not be made for the offshore wind prevailing on this occasion, but there was good correlation of the direction of maximum return and the direction of mean wind over the previous 12 hours.

The disadvantage of this technique is that it cannot be used to measure simultaneously the spectrum of the echoes and their direction, since it relies upon an assumption of discrete line structure in the spectrum. To overcome this defect, observations have been made by an interrupted-motion synthetic aperture technique, in which the vehicle is stopped at regularly spaced positions to make a stationary coherent spectrum analysis, the phase progression of the returns in each spectral resolution cell with distance being used to determine direction. Analysis of these results, which could make possible directional mapping of surface current and wave-directional spectra without assumptions of homogeneity, is in progress.

3. SKY-WAVE RADAR EXPERIMENTS

As a joint project between the University of Birmingham and the U.K. Science Research Council Appleton Laboratory, an extended period of synoptic observations of the sea in the North Atlantic is planned and the equipment is now being instrumented. The coverage of the radar, which is located in south-west England, has an aximuthal extent of ± 30° centred on 261° E of N, a beamwidth at 15 MHz of 4° and a range coverage of about 1,000 to 3,000 km with a 75 km range resolution. Early experiments in remote surface wind direction measurement with the same radar have been reported (SHEARMAN, BAGWELL and SANDHAM, 1977). The main new development is the installation of a PDP 11/34 computer and peripherals for on-line processing of the sea-echo. Multiple range-gate spectral analysis, incoherent averaging of spectra, extraction of surface wind parameters from the spectra and graphical mapping of the deduced data will be carried out in the computer. Figure 6 shows the computer-mapped wind vectors produced in a test of the programmes to be used. The input was in the form of simulated radar echo spectra, the computer using the ratio between Bragg line amplitude to deduce wind-direction, resolving the inherent left-right ambiguity in this process by a continuity condition for wind vectors in adjacent radar resolution cells.

3.1 Comparison of sky-wave radar measurements with sea-truth

A vital step in the introduction of radar oceanographic techniques is clearly the testing of the radar-deduced ocean and wind measurements against measurements by buoys and oceanographic vessels. A unique opportunity for such comparison for a sky-wave system arose in the summer of 1978 in the Joint Air-Sea Interaction project (JASIN), in which 14 oceanographic vessels of many nations undertook a coordinated experiment to study air-sea interaction in the Rockall Bank area of the North Atlantic.

To provide coverage of the JASIN area (see Fig. 7) a special northwards-directed rhombic antenna was erected for transmission and a narrower steerable beam used for reception to provide three azimuthal receiving beam positions within the transmitted beam as shown in Fig. 7. This together with radar range resolution gave a total of 18 resolution cells in the JASIN area. Each cell covered an area of 75 km in range by about 125 km in cross range. The azimuthal resolution was obtained by the use of a receiving aerial array having a beamwidth of 7°.

Sea back-scatter data were recorded on 17 dates during the JASIN period, and some preliminary results obtained on three days are briefly described here. Radar data obtained during the periods 1000-1030, 1300-1330 and 1600-1630 UT on 24 and 25 August have been used to deduce the dominant wind directions in the JASIN sea areas. The results have been compared with wind data obtained by a number of ships in the area at the time.

The radar was operated on a frequency of 15.066 MHz, and ionospheric conditions indicated that propagation was via the one-hop Es mode at all times. Separate Doppler spectra were obtained for the 18 sea areas, shown in Fig. 7 centred on ground ranges 870, 945, 1020, 1095, 1170 and 1245 km and along azimuths 318°, 325°, 332° E of N from the transmitter. Figure 8 shows at (a) examples of three individual power spectra for slant ranges 1125, 1200 and 1275 km in direction 318° on 24.8.78 at 0959 UT, illustrating the noise-like variability of the spectral estimate obtained in a single dwell. At (b) is shown the result of averaging the power spectra from 10 successive dwells; the reduction in variance of the spectra can be clearly seen by comparison of the spectra from different ranges, since changes in the ocean characteristics, of which the spectra are a manifestation, would be expected to change slowly with location. For each 10-fold average spectrum, the ratio of the spectral powers of the two first-order Bragg lines (approach and recede) was taken. This ratio was used to deduce the direction of the dominant wind with respect to the radar boresight, using two alternative relations which have been advocated by other workers : (a) LONG and TRIZNA, 1973, and (b) STEWART and BARNUM, 1975. In each case a 90% confidence interval for the Bragg line ratio was estimated by taking the measured value plus and minus 3.3 dB, (see Appendix). This range was then converted to a corresponding range of wind directions, using the relations given in (a) and (b) above. The results for the 18 sea areas were pooled in three groups of six so as to give a range of wind directions for each of the three areas, A, B, C shown in Fig. 7. For each area a corresponding range of directions was estimated from the wind data recorded by ships (up to 9 in number) in the JASIN area at the time. These data were two-minute averages taken at hourly intervals, but since the ship positions did not in general correspond closely to the radar sea areas, a considerable amount of interpolation and extrapolation of the ship data was necessary.

A comparison of the radar and ship-measured wind directions is shown in Fig. 9. Since the radar results are subject to an ambiguity (no discrimination between directions at equal angles on either side of the boresight) the directions corresponding more closely to the ship - wind directions have been chosen in each case. The results show that on 24 August, when wind speeds, as measured by ships in the area, were about 10 m s^{-1}, there was reasonably good agreement between the ship and radar-measured directions. The two radar relationships (R1 and R2) were also in general agreement. Poorer results were obtained on 25 August when the wind speed was only 5 m s^{-1}, indicating that under these conditions the presently available relations between wind direction and wave spectra are less reliable.

Wind speeds during the first two observation periods of 23 August were deduced from the radar spectra, using an empirical formula (STEWART and BARNUM, 1975) as later modified (MARESCA and BARNUM, 1977). This formula relates the mean wind speed in a radar resolution cell to the - 10 dB width of the stronger first-order Bragg line. The radar-deduced wind speeds at 1000 and 1300 UT have been compared with the presently available ship measurements, made at 0900 and 1200 UT. Both data sets gave speeds in the range 5-13 m s^{-1}, with good agreement for the morning data between radar and ship measurements at nearby locations. The afternoon data did not give such good agreement. However, it is noted that the ship measurements indicated a more irregular pattern of measured wind at 1200, so that detailed correlation of the geographical patterns at 1200 and 1300 would not be expected. As more detailed data become available permitting simultaneous comparisons a closer correlation should be possible.

Stable continuous-wave transmissions from the Scripps Institute vessel, Atlantis II on a frequency close to that used by the sky-wave radar were monitored at the radar site at the same times as sea-state soundings were carried out. The objectives of this subsidiary experiment were to try to monitor Doppler effects imposed by the ionosphere on signals travelling over approximately the same raypaths as for the radar. The received spectra have been processed and show Doppler spreads and displacements due to ionospheric effects and to ship motion. The results are under study to separate the two effects and so characterise the ionospheric path.

4. CONCLUSIONS

The present state of progress in the work can be summarised as follows :

Satisfactory signal processing techniques have been evolved for on-line and off-line spectral analysis of sea echo both by ground-wave and sky-wave modes. Coherent dwell times of 150 seconds for ground wave and 50 seconds for sky-wave with ten-fold averaging have yielded data of good quality. These techniques are being included in an on-line processing system for an improved sky-wave radar for synoptic studies of the North Atlantic.

In ground-wave observations, ship-tracking up to a range of 210 km achieved with omnidirectional antennas suggests a useful new supplementation of conventional microwave radar for marine traffic monitoring. Direction finding techniques should be applicable in the discrete echo conditions observed when fine resolution Doppler filtering is employed.

Synthetic aperture technique has proved useful for determining the directional spectrum of sea waves at MF, but if used with continuous motion of the receiving vehicle, it is not possible to measure the shape of the sea spectra or any current-induced Doppler shifts. A new interrupted-motion synthetic aperture technique shows promise of overcoming this defect.

An extension to the Barrick and Snider analysis of the spectral statistics of sea echo has been made and used in the determination of confidence intervals for the Bragg line ratio.

In the JASIN project, reasonably good agreement was found between the wind directions deduced from the radar spectra and those measured by ships in the area, when the wind speed was 10 m s^{-1}. Poorer agreement was found for wind speeds of 5 m s^{-1}.

ACKNOWLEDGEMENTS

The work described has been essentially a team enterprise, and would not have been possible without the dedicated work of D. J. Bagwell, S. Theodoridis and G. D. Burrows of the University of Birmingham, F. D. G. Bennett, C. R. Carter and M. A. Trower of the Appleton Laboratory, Sq. Ldr. J. Symons of the Royal Signals and Radar Establishment, B. W. Fursman of R.A.F. East Blockhouse, and D. V. Tibble of the Ministry of Defence.

REFERENCES

BARRICK, D. E. and SNIDER, J. B., 1977, "The statistics of HF sea-echo Doppler spectra", IEEE Trans. Ant. Prop., AP-25, 19.

BROCHE, P., 1979, "Estimation du spectre directionnel des vagues par radar decametrique coherent", Proc. Conf. "Special topics in HF propagation", Lisbon, AGARD Conference Proceedings No. 263.

CROMBIE, D. D., 1955, "Doppler spectrum of an echo at 13.56 Mc/s", Nature, 175, 681.

LONG, A. E. and TRIZNA, D. B , 1973, "Mapping of North Atlantic Winds by HF radar sea backscatter interpretation", IEEE Trans. Ant. Prop., AP-21, 680.

MARESCA, J. W. and BARNUM, J. R., 1977, "Measurement of oceanic wind speed from HF sea scatter by skywave radar", IEEE Trans. Ant. Prop., AP-25, 132.

SANDHAM, W. A., SHEARMAN, E. D. R. and BAGWELL, D. J., 1978, "Remote sensing of sea-state and surface winds in the Irish and Celtic seas using HF radar", Proc. Conf. "Antennas and Propagation", London, IEE Conference Publication No. 169, Part 2, pp. 1-5.

SHEARMAN, E. D. R., BAGWELL, D. J. and SANDHAM, W. A., 1977, "Progress in remote sensing of sea-state and oceanic winds by HF radar", Proc. Conf. "Radar-77", London, IEE Conference Publication No. 155, pp. 41-45.

STEWART, R. H. and BARNUM, J. R., 1975, "Radio measurements of oceanic winds at long ranges : An evaluation", Radio Science, 10, 853.

TEAGUE, C. C., TYLER, G. L. and STEWART, R. H., 1975, "The radar cross-section of the sea at 1.95 MHz : Comparison of in-situ and radar determinations", Radio Science, 10, 847.

APPENDIX : CONFIDENCE INTERVALS ASSOCIATED WITH BRAGG LINE RATIOS

The measured ratio of the spectral powers at the two first-order Bragg line frequencies is subject to statistical variations which can be estimated from the assumed Gaussian distribution of the echo signal voltage (BARRICK and SNIDER, 1977). On this basis, the probability distribution of the estimates of the power P_n at a given frequency obtained by averaging the n estimates of the power at that frequency from n spectra, is

$$p(P_n) = \frac{n^n P_n^{n-1} \exp(-nP_n/\overline{P})}{\Gamma(n)\,\overline{P}^n},$$

where $\overline{P}$ is the true mean value of P_n.

If q_n is the ratio in decibels of the values of P_n measured at the two Bragg line frequencies, it may be easily shown, by considering the distribution of the ratio of two uncorrelated quantities, that the distribution of q_n is

$$p(q_n) = \frac{A\Gamma(2n) \exp\left[An(q_o + q_n)\right]}{\left\{\Gamma(n)\right\}^2 \left[\exp(Aq_o) + \exp(Aq_n)\right]^{2n}},$$

(a case of the statistical F distribution), where $A = (\ln 10)/10 = 0.230$, and $q_o = 10 \log Q_o$, where Q_o is the ratio of the true mean powers at the two frequencies. The distribution of $r_n = q_n - q_o$ is then

$$p(r_n) = \frac{A\Gamma(2n) \exp(Anr_n)}{\left\{\Gamma(n)\right\}^2 \left[1 + \exp(Ar_n)\right]^{2n}}$$

which is symmetrical about zero so that the mean $r_n = 0$. The confidence intervals to be associated with any measured value of q_n can be readily obtained from the cumulative distribution of r_n, i.e.

$$\int_{-\infty}^{r} p(r_n)\, dr_n,$$

which is the probability that q_n is less than $(q_o + r)$. The integral can be expressed in closed form as a polynomial in $[1 + \exp(Ar)]^{-1}$, or equivalent evaluations made from tabulations of the F distribution.

The confidence intervals for q_n so obtained are symmetrical about the observed value, and the half widths in decibels, for 50% and 90% probabilities are as follows, for various values of n :

n	50%	90%
1	4.8	12.8
3	2.5	6.4
5	1.9	4.7
10	1.3	3.3

The value for the 90% confidence limits for 10 spectra averaged incoherently (n = 10) is thus 3.3 dB, the value quoted in section 3.1.

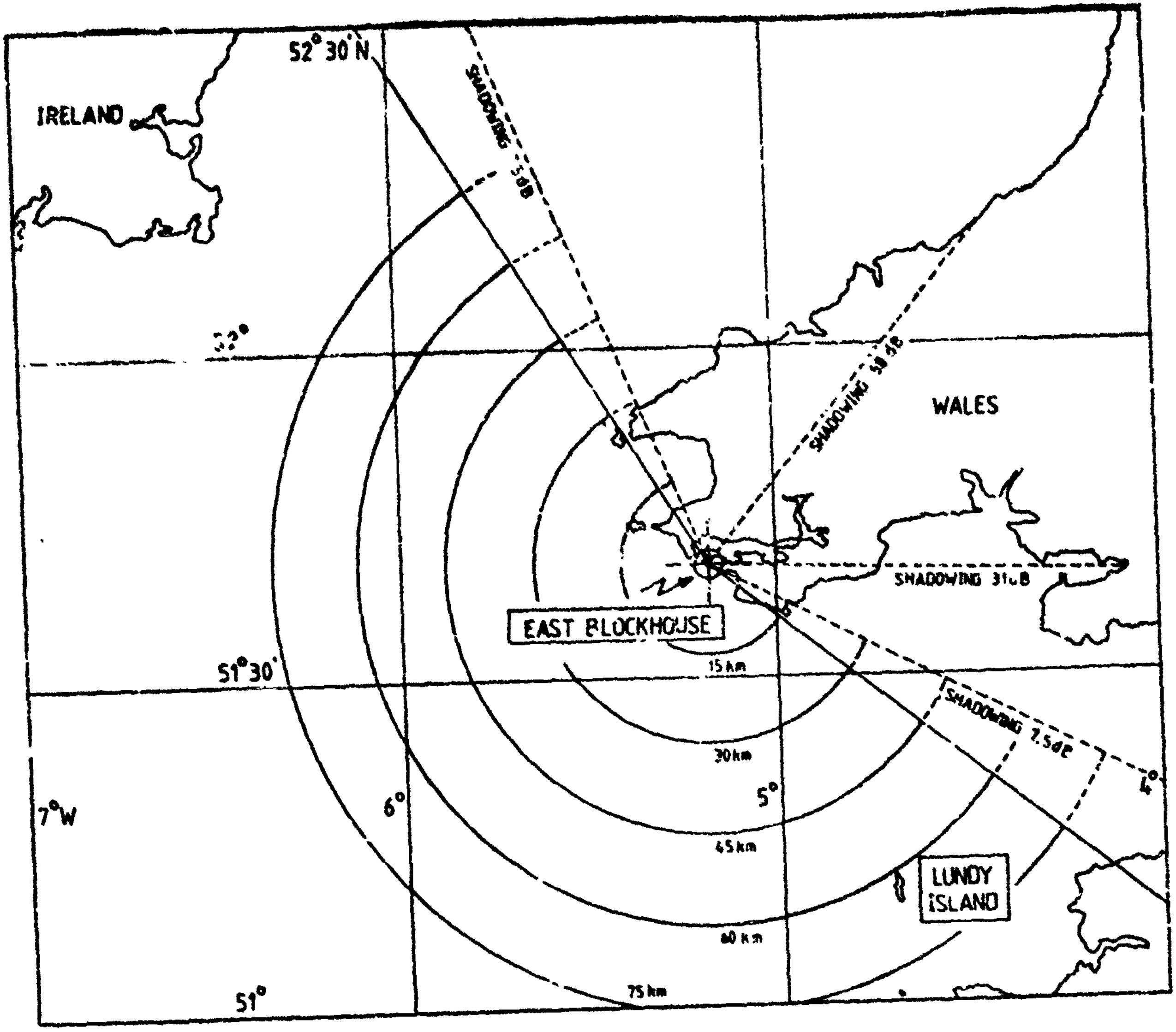

52° 30' N
IRELAND
SHADOWING 3dB
52°
SHADOWING 38 dB
WALES
SHADOWING 31dB
EAST BLOCKHOUSE
15 km
51°30'
SHADOWING 7.5dB
30 km
7°W
6°
5°
4°
45 km
LUNDY ISLAND
60 km
75 km
51°

Fig.1 Coverage area of Irish Sea by 1.95 MHz ground-wave radar at East Blockhouse, Milford Haven

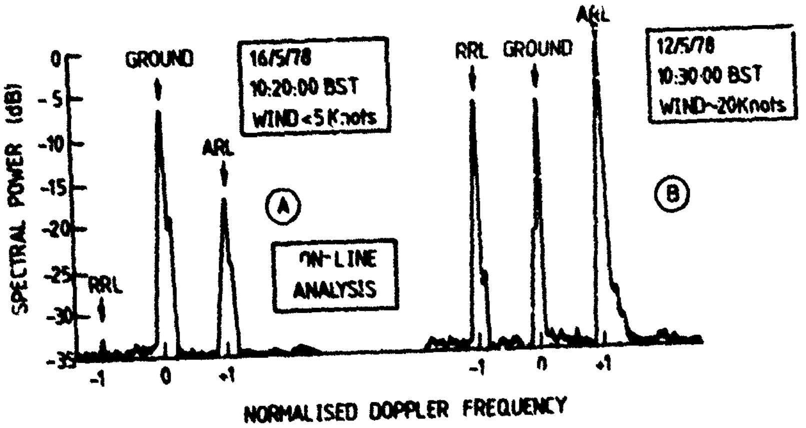


Fig.2 Doppler spectra at 1.95 MHz and 75 km range for two different sea-states. Spectrum B was taken on a day when the wind was sufficiently high and was blowing long enough for 77 m ocean waves to develop fully

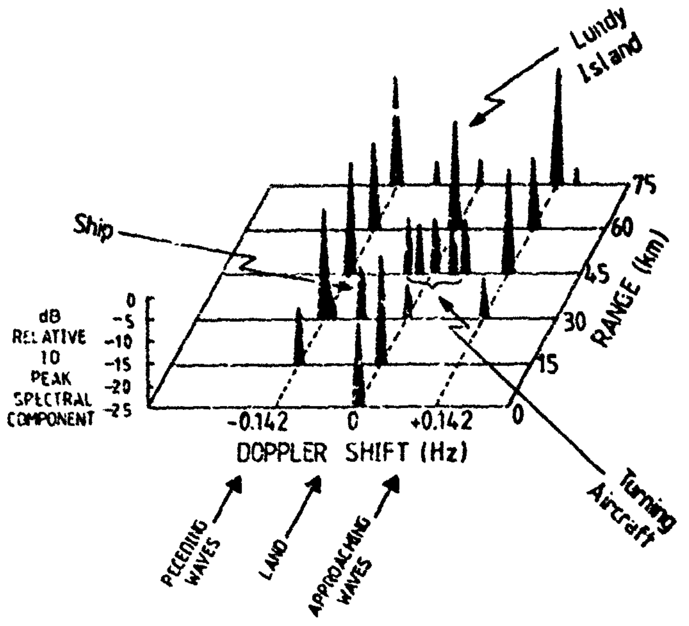


Fig.3 Echo spectra at 1.95 MHz vs range, showing sea-state variations, shipping and an island

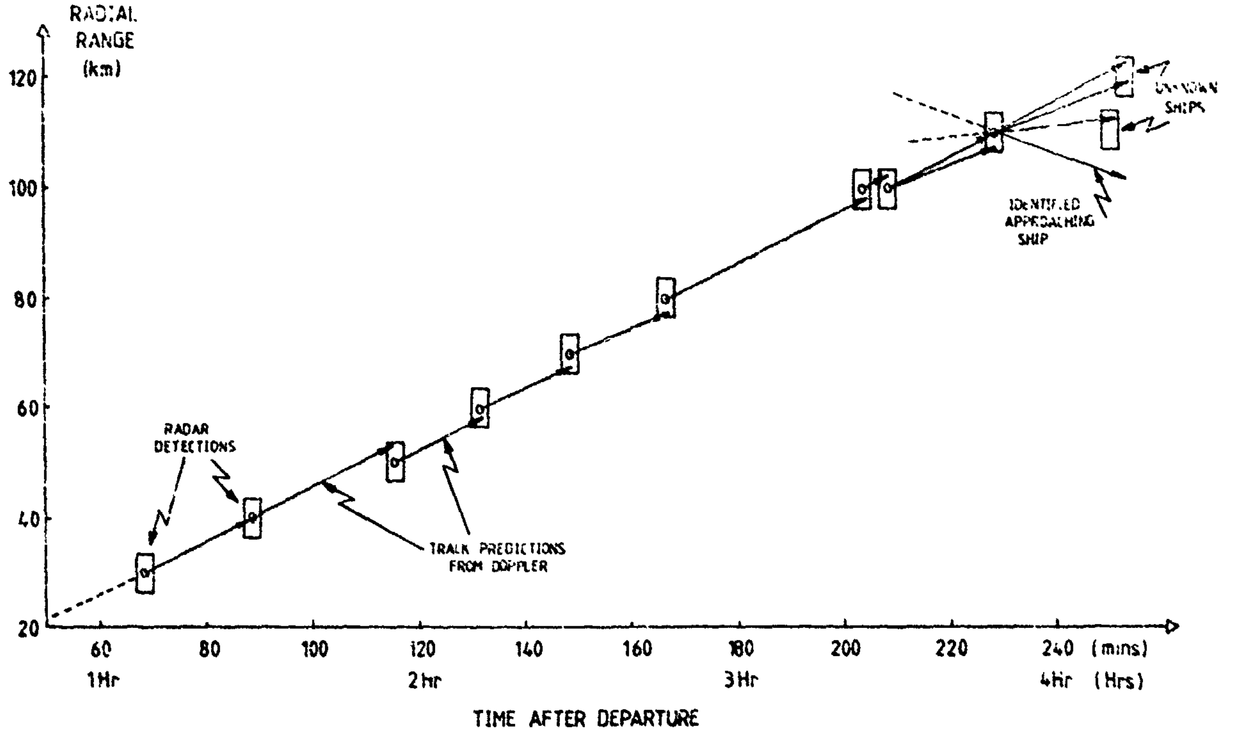


Fig.4 Range/time track of tanker leaving Milford Haven. Rectangles show plots (150 s coherent dwell, 7.5 km range resolution) and arrows show predicted track from Doppler shift

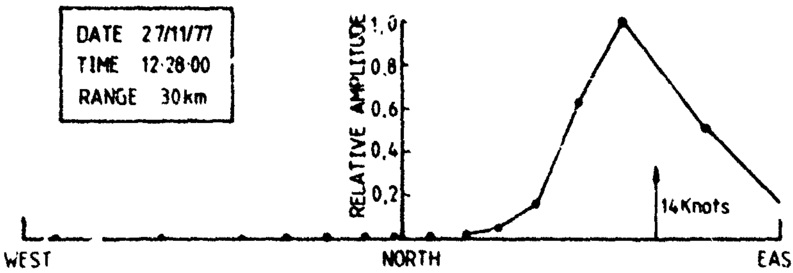


Fig.5 Amplitude/direction plot of receding 77 m waves derived from 1.95 MHz radio back-scatter. The wind average over the preceding 12 hours is indicated. The azimuths which are shown are those from which the waves are travelling, for comparison with the usual convention for wind direction

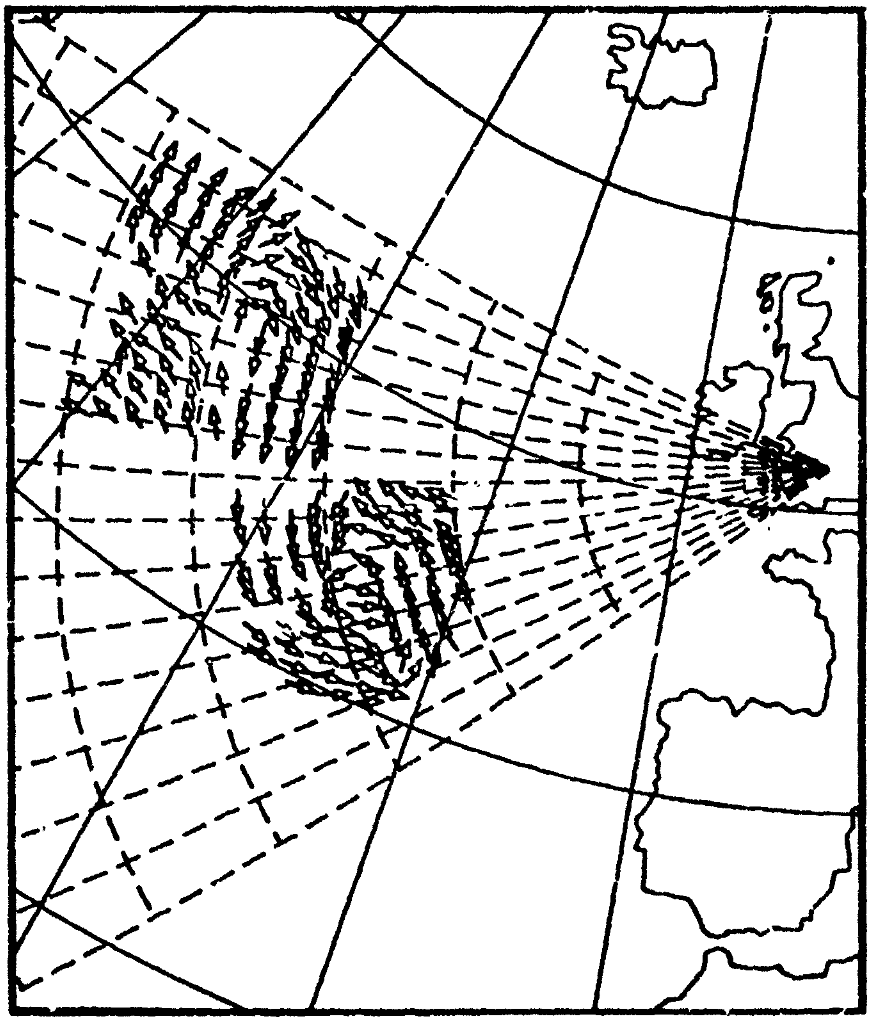


Fig.6 Computer plot of coverage of North Atlantic by sky-wave radar showing wind vectors computed automatically from simulated HF radar Doppler spectra input

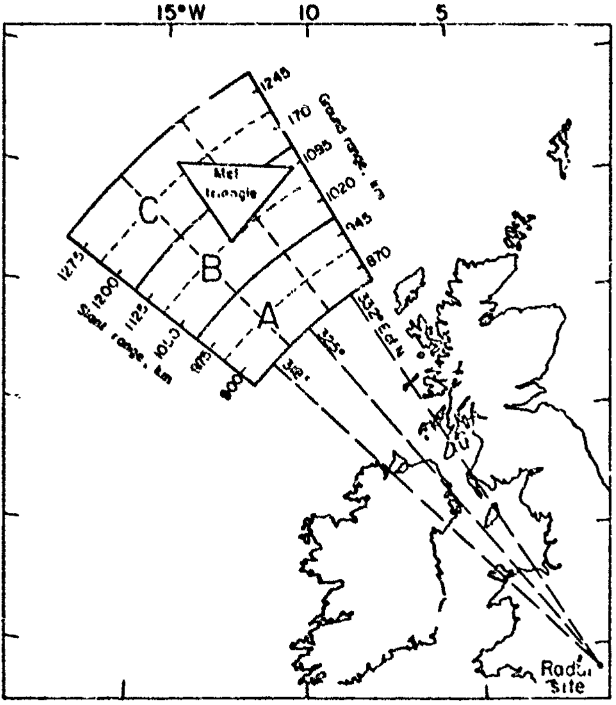
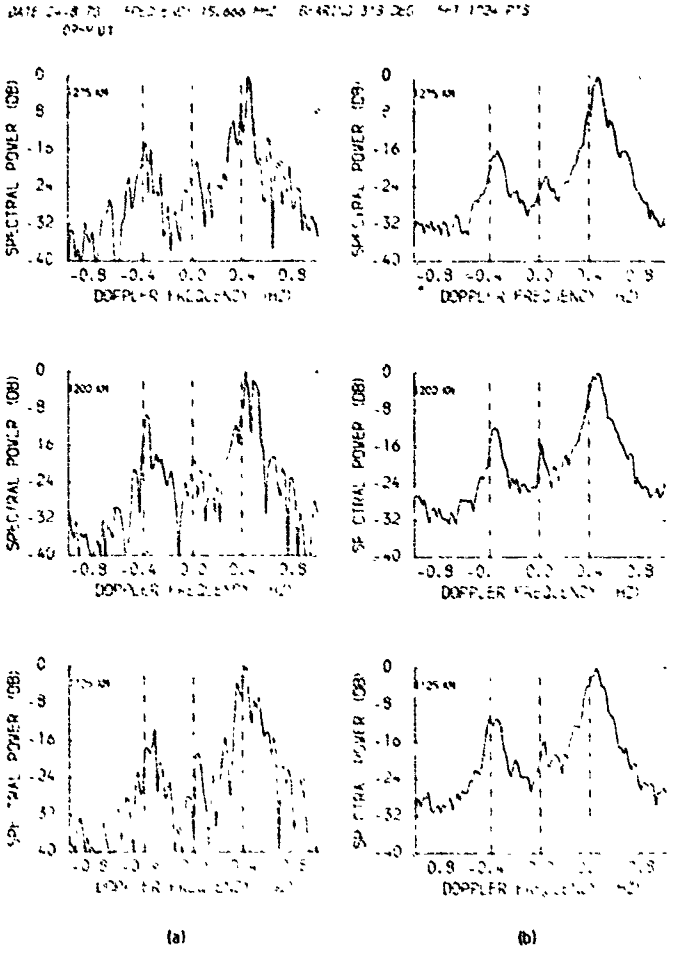

15°W
10
5
62°N
60
58
56
54
52
Met triangle
C
B
A
1245
1170
1095
1020
945
870
Ground range, km
1275
1200
1125
1050
975
900
Slant range, km
332°
325°
318°
Radar Site

Fig.7 Resolution cells of sky-wave radar used during JASIN project, Summer 1978



(a)

(b)

Fig.8 Sample spectra obtained during JASIN project from sea at slant ranges 1125, 1200 and 1275 km and bearing 318°, by sporadic-E propagation

(a) Shows single spectra obtained in one 50 second coherent dwell

(b) Shows improvement in quality of spectra resulting from incoherent averaging of 10 spectra

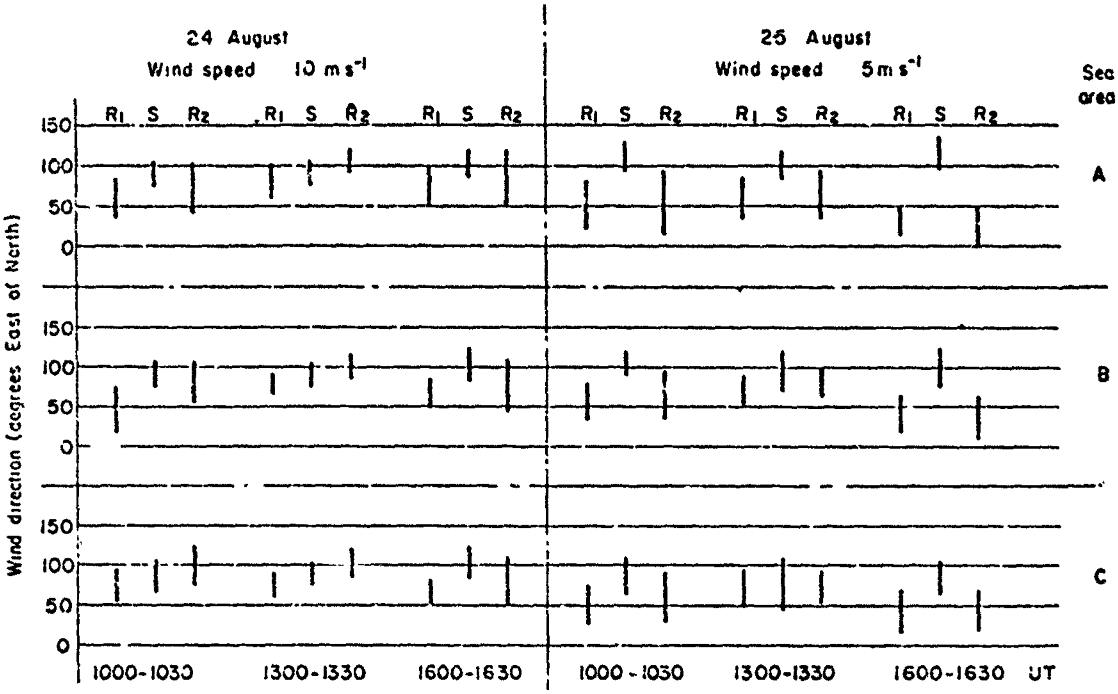


Fig.9 Comparison of wind directions measured by ships with those deduced from remote sky-wave Doppler radar data

S : ship measurements

R_1: Long and Trizna radar/wind relationship

R_2: Stewart and Barnum radar/wind relationship

DISCUSSION

J.R. Wait, US

Can the ionospheric contribution to the total Doppler shift be deduced by a combined analysis of the single-hop and double-hop data?

Author's Reply

The two-hop sea-scatter that I discussed was from twice the range of the main one-hop illuminated, in this case I believe we may assume specular reflection at the first sea bounce point and attribute any additional Doppler shift observed on the second sea-scatter path to the second reflection. However, the two ionospheric reflection points will be considerably separated and no correlation in motion may be assumed. If we observe the same ocean path by one and two hop modes and endeavour to eliminate the Doppler shift due to ionospheric motion, we shall have two difficulties. Firstly the echoes will be likely to overlap in time and Doppler shift and therefore be difficult to separate. Secondly, the ionospheric reflection points will still be significantly separated spatially for a 2,000km range. The reflection points would be 500km apart and it would be very dangerous to assume well correlated motion. My answer to the question would therefore have to be that such a technique is likely to be practical only rarely.

C.S. Goutelard, Fr

Le Professeur Shearman a insisté dans son exposé sur le fait qu'une grande directivité n'etait pas nécessaire pour la localisation des mobiles. Il me semble que, si dans le cas qu'il a exposé cela est vrai, il n'est pas possible de généraliser notamment dans le cas des radars OTH, surtour sur les longues distances. En effet dans ce cas les inhomogénéités de la zone ionosphérique où se crée la réfraction étale le spectre doppler des echos, comme l'a expliqué le Prof. Shearman il y a déjà une quinzaine d'années. Cet étalement du spectre augmente avec la largeur du lobe d'antenne tout comme, d'ailleurs, les trajets multiples. Ces deux points introduisent des limitations dans la détection des cibles que l'on peut repousser par des techniques d'integration cohérentes ou incohérentes, mais ici apparait également des limitations. Dans ces problèmes il apparait donc indispensable d'avoir de bonne directivites d'antennes.

Author's Reply

My remarks on the use of omnidirectional or widebeam antennas and direction finding techniques for ship tracking were restricted to the ground-wave mode. In this there are no medium-induced Doppler shifts or multiple paths.

For sky-wave propagation, differential Doppler shifts between different directions and different modes of propagation certainly complicate the situation of using Doppler filtering to separate targets from clutter. I agree with the comments for this mode.

The beam width used in our special observation of the JASIN oceanographic area was 7°, while the steerable-beam phased area to be used to survey the North Atlantic uses a 4° beam. Useful results are obtained with these beam widths, but they are still considerably larger than the 0.5° beam width of the WARF array.

Comment by Dr. Washburn, US

I would just like to comment that the data collected at WARF for sea state monitoring usually consists of Doppler spectra at various ranges, azimuths and times of day. Noncoherent (post detection) averaging is necessary to produce a smoothed, stable spectral estimate. This averaging is performed over range, azimuth and time. Furthermore, some spectra are discarded when they show evidence of ionospheric spectral smearing such as produced by multipath or by ionospheric coherence limitation. Dr Maresca and Dr Tom Georges at NOAA are exploring specific techniques for improving data quality.

SUMMARY OF SESSION V

REMOTE SENSING AND OCEAN SURVEILLANCE

The initial paper by **Professor Broche** presented the results of several experiments performed recently in the western Mediterranean Sea during which a ground based HF Doppler radar was operated at two frequencies (near 6 and 12MHz), in conjunction with a buoy which supplied in-situ measurements of the main parameters of the sea surface.

The second paper by **Washburn, Sweeny, Barnum** and **Zavoli** concentrated on describing the capabilities of the Wide Aperture Research Facility (WARF) operated by SRI in California. Various radar techniques and propagation management tools used on line and in real time were discussed. This was followed by a paper by **Moresca** and **Carlson** which presented HF Skywave estimates of the track, surface wind and wave of hurricane Anita. The track of the hurricane and intensity of the waves were deduced from measurements made at a range greater than 3000Kms utilizing the SRI WARF radar. The ability to steer the radar to observe selected storm regions makes HF Skywave an excellent remote sensing device to monitor intense storms.

The fourth paper by **Barrick** and **Lipa** described the remote sensing of sea state and surface currents with HF radars. The paper emphasized the ability of HF radars to measure surface characteristics to an accuracy of a few percent. The concept has application to environmental monitoring, erosion analyses, off shore oil and mining operations etc.

SEA-STATE DIRECTIONAL SPECTRA OBSERVED BY H.F. DOPPLER RADAR

ESTIMATION DU SPECTRE DIRECTIONNEL DES VAGUES PAR RADAR DECAMETRIQUE COHERENT

P.BROCHE

L.S.E.E.T., Université de TOULON, 83130 LA GARDE (France)

SUMMARY

Basic elements of the description of the sea surface, and fundamental phenomena implied in the backscattering of the electromagnetic waves by this surface are reminded.

Results of experiments performed with a ground-wave H.F. Doppler radar are presented and compared with a sea-truth supplied by conventional oceanographic means. Two points are stressed :

- the estimation of the wind direction, by measuring the ratio of the amplitudes of the two Bragg lines in the doppler spectrum of the backscattered H.F. wave.

- the estimation of the dominant frequency in the sea-state spectrum and of the significant height of the waves by measuring the cut-off frequencies which are present on both sides of the Bragg lines.

RESUME

On rappelle les éléments principaux relatifs à la description de l'état de surface de la mer, ainsi que les phénomènes fondamentaux impliqués dans la rétrodiffusion par cette surface des ondes électromagnétiques.

Les résultats de campagnes expérimentales menées avec un radar décamétrique cohérent fonctionnant en onde de sol sont présentés et comparés avec les données d'une vérité-mer assurée par des moyens océanographiques habituels.

L'accent est mis sur l'estimation :

- de la direction du vent, par la mesure du rapport des amplitudes des raies de Bragg dans le spectre de puissance de l'écho radar.

- de la fréquence dominante du spectre des vagues et de leur hauteur significative, par la mesure des fréquences de coupure qui apparaissent de part et d'autre des raies de Bragg.

1. INTRODUCTION

Les études menées dans la dernière décade à la suite des premiers travaux de CROMBIE (1955) ont montré que les radars cohérents fonctionnant dans la gamme décamétrique (H.F.) constituent un instrument adapté à l'observation à distance de la surface de la mer.

On rappelle tout d'abord ci-dessous quelques éléments relatifs à la description de l'état de cette surface, ainsi que les mécanismes fondamentaux par lesquels celui-ci influe sur les propriétés de l'onde électromagnétique rétrodiffusée (écho radar).

On présente ensuite les résultats acquis dans le cadre de plusieurs campagnes d'expériences durant lesquelles une comparaison suivie a pu être faite des données fournies par le radar avec celles d'une vérité-mer assurée par des moyens conventionnels (bouées) sur le site exploré par le radar. L'accent est mis sur l'estimation de deux paramètres :

- la direction du vent
- la fréquence dominante du spectre des vagues et l'énergie correspondante (hauteur significative).

On discute la qualité de cette estimation, ainsi que les conditions qui sont de nature à l'améliorer.

2. L'ETAT DE LA MER ET LA RETRODIFFUSION DES ONDES H.F.

2.1 Description de la surface de la mer

On admet généralement que les variations $\xi(x,y,t)$ de la cote du point de coordonnées (x,y) de la surface de la mer résultent de la superposition des effets d'une infinité d'ondes libres, sinusoïdales, indépendantes. La pulsation ω et le vecteur d'onde $\vec{k}$ de chacune d'elles sont reliés par l'équation de dispersion des ondes de gravité :

$$\omega^2 = gk$$

Cette superposition est décrite par le spectre d'énergie des vagues, $S(\vec{k})$, tel que :

$$\overline{\xi^2} = \iint S(\vec{k})\, d\vec{k}$$

$S(\vec{k})$ dépend à la fois du nombre d'onde k (donc de la pulsation ω), et de la direction du vecteur d'onde $\vec{k}$ repérée par son angle θ par rapport à une origine quelconque (on choisit souvent la direction du vent). On peut donc aussi écrire :

$$\overline{\xi^2} = \iint S(\omega,\theta)\, d\omega\, d\theta$$

et on pose :

$$S(\omega,\theta) = S(\omega)\, f(\omega,\theta)$$

avec :

$$\int f(\omega,\theta)\, d\theta = 1 \qquad \text{soit} \quad S(\omega) = \int S(\omega,\theta)\, d\theta$$

(a) $S(\omega)$ est le spectre "omnidirectionnel" de la hauteur des vagues. C'est lui que permettent de calculer les données d'une bouée mouillée en un point fixe qui mesure les variations temporelles de ξ. Ce spectre est relativement bien connu, du moins pour les vagues en phase de croissance ou en équilibre avec le vent : on peut le décrire convenablement par le modèle de Pierson-Moskowitz :

$$S(\omega) = \frac{\beta g^2}{\omega^5} \exp\left(- \frac{5}{4} \left(\frac{\omega_m}{\omega}\right)^4 \right)$$

dont l'allure est représentée Fig. 1 (g est l'accélération de la pesanteur). Il dépend des deux paramètres β et ω_m (pulsation du maximum d'amplitude spectrale), qui sont eux-mêmes fonctions du vent et de ses conditions d'action sur la surface de la mer (distance et durée). Il existe à ce propos un certain nombre de formules semi-empiriques, citées par RAMAMONJIARISOA (1975), et en particulier le modèle JONSWAP (HASSELMANN et al., 1973).

$S(\omega)$ donne également accès aux paramètres statistiques habituels, comme la hauteur significative H 1/3, qui est la hauteur moyenne du tiers des vagues les plus élevées, et qui est donnée par :

$$H_{\frac{1}{3}} = \frac{4g}{\omega_m^2} \sqrt{\beta/5}$$

(b) La fonction $f(\omega,\theta)$ décrit la répartition angulaire de l'énergie des composantes de pulsation ω. Sa connaissance est beaucoup plus fragmentaire que celle de $S(\omega)$ et les quelques modèles existant sont purement empiriques. Parmi eux, le modèle de Longuett-Higgins, Cartwright et Smith, cité par TYLER et al. (1974) est couramment utilisé et il est donné par :

$$f(\omega,\theta) = a + (1-a)\cos^s \frac{\theta}{2}$$

θ est repéré par rapport à la direction du vent avec lequel la mer est supposée en équilibre (mer complètement développée). a représente la proportion (faible, de l'ordre de 10^{-2}) de l'énergie qui se propage dans la direction opposée à celle du vent. Le paramètre d'anisotropie s varie avec la fréquence des vagues étudiées et avec la vitesse du vent : sa valeur est grande ($s \gtrsim 10$) pour les vagues dominantes (fréquence ω_m), qui ont une vitesse voisine de celle du vent, et elle diminue lorsque la fréquence augmente. La pertinence de ce modèle sera discutée ci-dessous (§ 4.1).

La télédétection de l'état de surface de la mer peut donc avoir comme objectif la détermination des quelques paramètres caractéristiques des modèles acceptés à priori comme convenables, (f_m, β ou H1/3, θ), ou, au contraire, l'estimation des amplitudes spectrales $S(\omega,\theta)$ pour plusieurs valeurs de ω et de θ.

2.2. Rétrodiffusion des ondes électromagnétiques

L'examen du spectre de puissance de l'onde électromagnétique de longueur d'onde λ_R, de fréquence f_R rétrodiffusée par la surface de la mer (Fig. 2) montre que l'essentiel de la puissance est contenu dans deux raies bien individualisées, dont les décalages Doppler par rapport à la fréquence émise sont opposés et correspondent à la célérité des ondes de gravité de longueur d'onde $\lambda_g = \lambda_{R/2}$: il s'agit donc d'un mécanisme de diffusion cohérente, analogue au phénomène de Bragg dans les cristaux, qui sélectionne dans le spectre directionnel des vagues les composantes adaptées. Schématiquement, si $\vec{k}_i$ et $\vec{k}_s$ sont les vecteurs d'onde des ondes émise et reçue (déterminés par la fréquence émise et l'orientation du lobe des antennes), $\vec{k}_g$ le vecteur d'onde d'une onde de gravité, et si ω_i, ω_s, ω_g sont les pulsations correspondantes, on a :

$$\vec{k}_s = \vec{k}_i \pm \vec{k}_g$$

$$\omega_s = \omega_i \pm \omega_g$$

comme $\vec{k}_s = - \vec{k}_i$ (rétrodiffusion), on en déduit :

$$\vec{k}_g = \pm 2 \vec{k}_i \qquad \text{ou} \qquad \lambda_g = \frac{\lambda_R}{2}$$

et le décalage Doppler est donné par :

$$\omega_s - \omega_i = \pm \omega_g = \pm \sqrt{2gk_i}$$

soit, en fréquence : $\Delta f(\text{Hz}) = \pm f_B \sim \pm 0{,}1 \sqrt{f_R(\text{MHz})}$

La gamme 0,1 - 0,4 Hz pour f_B, dans laquelle se trouvent habituellement les fréquences dominantes des spectres des vagues, correspond ainsi à la gamme 1-16 MHz.

Les énergies de ces deux raies, dites raies de Bragg sont proportionnelles aux composantes du spectre de l'état de surface de la mer de vecteurs d'ondes $\pm 2\vec{k}_i$: la raie négative est proportionnelle à $S(2\vec{k}_i)$, la raie positive à $S(-2\vec{k}_i)$.

Signalons qu'on observe souvent un décalage d'ensemble de la position des deux raies de Bragg qui est interprété comme l'effet d'un courant superficiel dont il est ainsi possible de mesurer la composante radiale (STEWART and JOY, 1974 BARRICK et al. 1977).

Il existe de plus, entre et de part et d'autre des raies de Bragg, un continuum dit "écho du second ordre", d'un niveau sensiblement inférieur et dont on attribue la présence à des interactions non linéaires :

- entre l'onde radio et la surface de la mer (diffusion multiple)

- entre les ondes océaniques elles-mêmes, donnant naissance à des déformations de vecteur $\vec{k}$ adapté mais qui n'obéissent pas à l'équation de dispersion des vagues libres.

Schématiquement, si $\vec{k}_1$, $\vec{k}_2$ sont les vecteurs d'ondes de deux composantes de l'état de surface de la mer et ω_1, ω_2 les pulsations correspondantes, on a :

$$\vec{k}_s = \vec{k}_i \pm \vec{k}_1 \pm \vec{k}_2$$

$$\omega_s = \omega_i \pm \omega_1 \pm \omega_2$$

et donc :

$$\begin{aligned} \pm \vec{k}_1 \pm \vec{k}_2 &= - 2\vec{k}_i \\ \omega_s - \omega_i &= \pm \omega_1 \pm \omega_2 \end{aligned} \tag{1}$$

l'énergie présente à la fréquence ω_s résulte de l'ensemble des combinaisons $\vec{k}_1$, $\vec{k}_2$ qui satisfont les conditions ci-dessus. L'écho du second ordre, pour une fréquence radio donnée, dépend donc de l'ensemble du spectre directionnel de l'état de surface de la mer.

3. CONDITIONS EXPERIMENTALES

Deux types de fonctionnement peuvent être envisagés pour le radar :

- en onde de ciel, l'onde électromagnétique subit, entre le radar et la surface de la mer une ou plusieurs réflexions sur l'ionosphère, dont les mouvements introduisent un effet Doppler qu'il est parfois difficile de séparer de celui dû à la surface de la mer. La portée est grande, les fréquences sont nécessairement assez élevées, et ne permettent pas l'observation par les effets du premier ordre de la partie la plus énergétique du spectre des vagues.

- en onde de surface, l'onde électromagnétique est guidée, au-delà de la zone de visibilité directe, par la discontinuité de conductivité entre la mer et l'atmosphère. La portée est faible (50-100 km). La fréquence peut être assez basse mais la dimension qu'il est nécessaire de donner aux antennes pour leur assurer une directivité suffisante rend délicat de descendre au-dessous de quelques MHz.

Le radar utilisé par le L.S.E.E.T. pour les expériences dont les résultats sont décrits ci-après fonctionne en onde de sol à partir d'une station à la côte située aux environs de Toulon (Méditerranée occidentale).

L'émission consiste en impulsions de cadence et de durée variables (habituellement 100 Hz et 100 µs). Deux fréquences différentes peuvent être utilisées simultanément par entrelacement des impulsions correspondantes, dans la gamme 2-30 MHz. La plupart des expériences ont été réalisées à 6 MHz et 12 MHz (fréquences de Bragg : 0,25 et 0,35 Hz), pour lesquelles les portions de mer étudiées ont une étendue radiale de 15 Km et une largeur, liée aux caractéristiques des antennes, de 10 et 5 Km respectivement à la distance de 40 Km de la station.

A cette distance, on a disposé d'une bouée qui a servi de support à des moyens classiques de mesure des paramètres météorologiques et océanographiques. Cette vérité-mer comprenait la mesure du vent (vitesse U et direction Θ) et la mesure de la hauteur des vagues (hauteur significative H 1/3, spectre omnidirectionnel $S(\omega)$).

Bien que l'on ait procédé à l'enregistrement de la totalité du signal de rétrodiffusion afin de permettre tout autre traitement ultérieur, les spectres radar utilisés résultent en général de la sommation incohérente de 8 spectres consécutifs, chacun étant calculé en temps réel par FFT avec une résolution de 0,01 Hz. Ils correspondent donc à environ 14 minutes de signal.

4. RESULTATS

L'essentiel des travaux réalisés à ce jour vise à la mise au point de relations empiriques entre les propriétés des spectres du premier ordre (amplitude et largeur des raies de Bragg) et la vitesse du vent (LONG and TRIZNA, 1973 ; STEWART and BARNUM, 1975 ; de MAISTRE et al. 1977 ; BARNUM et al., 1977 ; MARESCA and BARNUM, 1977). D'autre part, les études entreprises sur les spectres du second ordre (BARRICK et al. 1974) débouchent sur des méthodes d'inversion destinées à en extraire l'ensemble du spectre directionnel des vagues (BARRICK, 1977 ; LIPA, 1977).

Les résultats présentés ci-dessous s'inscrivent dans cette double perspective, ils concernent l'estimation de la direction du vent, de la fréquence dominante et la hauteur significative des vagues.

4.1. Estimation de la direction du vent

(a) Le principe de cette estimation est bien connu : les amplitudes A_- et A_+ des raies de Bragg sont proportionnelles à $S(\pm 2\vec{k}_i)$ et leur rapport est :

$$(2) \qquad R = \frac{A_+}{A_-} = \frac{S(-2\vec{k}_i)}{S(2\vec{k}_i)} = \frac{f(\omega_B, \pi-\Theta)}{f(\omega_B, \Theta)}$$

Θ est l'angle de la direction du vent avec celle du faisceau radar ($\vec{k}_i$). Le rapport R présente l'avantage, du point de vue expérimental, de ne pas dépendre de la valeur absolue de la puissance reçue. Son utilisation n'impose donc pas la connaissance de grandeurs (gains des antennes, atténuation le long de la propagation ...) qui sont difficiles à évaluer.

Un modèle pour $f(\omega,\Theta)$ étant choisi, la relation (2) permet d'estimer Θ à partir de la mesure de R. La qualité de cette estimation est liée à celle du modèle utilisé, mais on peut déjà prévoir qu'elle sera limitée parce que la répartition angulaire réelle dépend de conditions difficiles à connaître ou même à prendre en compte (distance et durée de l'action du vent). Il est donc inutile de disposer d'un modèle trop précis et il est préférable que, même approximatif, il couvre des conditions de validité étendues.

(b) Le modèle LCS décrit en (2.1.b) est utilisé, sous une forme simplifiée dans laquelle a=0, par STEWART and BARNUM (1975). En fait, il est possible d'utiliser la forme complète, a étant déterminé à partir de la plus grande valeur R_m observée pour R : $a = 1/R_m$.

D'autres auteurs emploient un modèle de la forme (LONG and TRIZNA, 1973 ; MARESCA et al., 1977) :

$$f(\omega,\Theta) = (1 + b\cos^2\Theta) \qquad |\Theta| < \frac{\pi}{2}$$

$$f(\omega,\Theta) = (1 + b\cos^2\Theta)^{-1} \qquad \frac{\pi}{2} < |\Theta| < \pi$$

de sorte que :

$$R = (1 + b\cos^2\Theta)^{-2} \qquad |\Theta| < \frac{\pi}{2}$$

$$R = (1 + b\cos^2\Theta)^{2} \qquad \frac{\pi}{2} < |\Theta| < \pi$$

b peut également être déterminé à partir de $R_m = (1 + b)^2$.

On a reporté Fig. 3 les variations de R avec Θ pour ces deux modèles, pour plusieurs valeurs de s, et pour R_m = 20 db.

Il est important de remarquer que si $s > 2$, ce qui est probablement le cas si l'on ne travaille pas avec des fréquences très élevées, les courbes peuvent être considérées comme peu éloignées, au regard de la précision dans la mesure in-situ de $\Theta(\pm 20°)$ et dans la mesure de R (au mieux ± 3 db, compte-tenu de la variabilité observée).

La pertinence de tels modèles est certainement limitée aux cas où les vagues étudiées sont en équilibre avec le vent, ce qui nécessite un temps d'autant plus court que leur fréquence est plus élevée et que le vent est plus fort. Il y a donc lieu de choisir une fréquence de Bragg assez élevée (mais pas trop pour ne pas tomber dans la gamme où s étant faible ($s < 2$), la sensibilité du modèle à la valeur effective de ce paramètre est grande) et d'écarter les mesures correspondant à des vents trop faibles.

(c) Ces considérations sont confirmées par les résultats expérimentaux de la Fig. 4, qui représentent les variations avec l'angle Θ mesuré in-situ, du rapport R observé à la fréquence de 6 MHz (Fig. 4a, $f_B = 0{,}25$ Hz) et à la fréquence de 12 MHz (Fig. 4b $f_B = 0{,}35$ Hz) :

- on n'observe pas de valeur de R supérieure à 20 db, ce qui justifie le choix $a = 0{,}01$ ou $b = 9$ effectué pour tracer les courbes de la Fig. 3.

- la dispersion des points est plus grande à 6 MHz qu'à 12 MHz : le temps nécessaire à l'équilibre mer-vent étant plus grand, il est normal que la corrélation ($R - \Theta$) soit plus faible.

- à 12 MHz, la corrélation (R,Θ) est sensiblement améliorée si l'on ne tient pas compte des points qui correspondent à $H\ 1/3 < 1$ m, pour lesquels la vitesse du vent est insuffisante pour saturer la composante de fréquence 0,35 Hz dans un temps bref à l'échelle de la variabilité des phénomènes météorologiques (quelques heures). On élimine ainsi la plupart des points pour lesquels les composantes étudiées (raies de Bragg) ne sont pas en équilibre avec le vent, et les points restants se répartissent autour d'une courbe moyenne voisine de celle correspondant à $s = 3$, ou même de la droite :

$$R(\mathrm{db}) = 20 \cdot \frac{\Theta - 90}{90}$$

avec une précision compatible avec l'indétermination signalée en (b).

Un traitement analogue appliqué aux données recueillies à 6 MHz conduirait à élever le seuil de 1 m pour H 1/3, et à éliminer un trop grand nombre de points, réduisant d'autant l'intérêt de la méthode.

Les mesures présentées ne permettent donc pas de choisir entre les modèles proposés, ni de préciser s -elles ont d'ailleurs été acquises dans des conditions météorologiques variées et il serait normal que ce paramètre présente une certaine dispersion- , mais celles correspondant à 12 MHz permettent, à partir de leur comportement moyen, d'estimer convenablement la direction du vent : cela est illustrée Fig. 5, où on a porté, en fonction de Θ, la valeur Θ^* estimée à partir de $s = 3$. Θ^* ne diffère pas, dans la majorité des cas, de Θ, de plus de ± 20° .

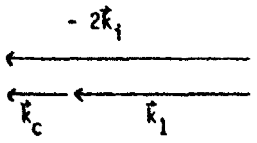
4.2. Estimation du spectre omnidirectionnel des vagues

Très souvent, le spectre du second ordre expérimental (Fig. 2) montre clairement de part et d'autre des raies de Bragg, une discontinuité de son amplitude qui est liée à l'existence d'une fréquence de coupure voisine de $\omega_m/2\pi$ dans le spectre omnidirectionnel des vagues (cf 2.1 a). Ceci s'explique aisément à partir des considérations développées § 2.2 : pour la partie du spectre du second ordre telle que $\omega_s - \omega_i > \omega_B$, par exemple, on a :

$$-2\vec{k}_i = \vec{k}_1 + \vec{k}_2$$

$$\omega_s - \omega_i = \omega_1 + \omega_2$$

Le spectre des vagues ne contenant pas de pulsation inférieure à $\omega_c = 2\pi f_c$ (et donc de nombre d'onde inférieur à $k_c = \omega_c^2/g$), la plus petite valeur de $(\omega_s - \omega_i)$ attachée à une contribution non nulle au spectre du second ordre correspond au schéma ci-dessous (si $\omega_c < \frac{\omega_B}{\sqrt{2}}$) :



pour lequel :

$$\omega_s - \omega_i = \omega_c + \omega_1$$

avec :

$$\omega_1 = \sqrt{gk_1} = \sqrt{g(2k_i - k_c)} = \sqrt{\omega_B^2 - \omega_c^2}$$

Un raisonnement identique est possible pour les trois autres combinaisons des signes dans les relations (1) (& 2.2), conduisant aux quatre pulsations de coupure dans le spectre du second ordre :

$$\omega_{s+} = \sqrt{\omega_B^2 - \omega_c^2} + \omega_c = -\omega_{s-}$$

$$\omega_{d+} = \sqrt{\omega_B^2 + \omega_c^2} - \omega_c = -\omega_{d-}$$

On peut remarquer que dans le cas d'une houle dont le spectre directionnel est très étroit, autour d'une direction donnée Θ_H, ce raisonnement implique une dissymétrie dans la position des raies positives et négatives dépendant de Θ_H que l'on peut ainsi estimer (P. FORGET, 1979).

On a déterminé, à partir de la position de ces coupures, et plus particulièrement de ω_{d+} et ω_{d-} (coupures situées entre des raies de Bragg, pour lesquelles :

$$\omega_c = (\omega_B^2 - \omega_d^2) \; / \; 2\,|\omega_d|$$

la fréquence la plus basse présente dans le spectre des vagues. Compte-tenu de la dynamique des spectres radar, on a assimilé f_c à la fréquence dominante f_m. Ceci a été réalisé pour une fréquence radio de 12 MHz sur tous les spectres qui le permettaient, et le résultat est illustré Fig. 6, où on a porté la fréquence ainsi estimée f_m* (l'incertitude est de l'ordre de ± 0,03 Hz) en fonction de la fréquence f_m mesurée par la bouée sur le site de l'expérience. L'accord est convenable, excepté pour les fréquences > 0,25 Hz qui correspondent en fait à des mers de très faible amplitude.

Si la mer est complètement développée, le paramètre β du spectre de Pierson-Moskowitz est voisin de $\beta = 0{,}008$, et la hauteur significative des vagues est reliée à f_m par :

$$H\,1/3 = 0{,}04/f_m^2 \qquad (3)$$

(pour les cas où la houle est absente).

La Fig. 7 montre la relation qui existe entre l'estimation ainsi effectuée de H 1/3 et la valeur mesurée sur la bouée. Les désaccords qui existent sont liés à la fois à l'imprécision sur f_m (cf Fig. 6), mais aussi au fait que les conditions de stationnarité indispensables à l'application de la relation (3) ne sont pas remplies. De ce point de vue, l'accord ne serait pas meilleur si l'on remplaçait $H^*_{1/3}$ par une estimation $\check{H}_{1/3}$ à partir de la valeur mesurée de f_m.

5. CONCLUSION

Les résultats présentés ci-dessus, et surtout la comparaison de chacun des paramètres estimés à l'aide des spectres radar avec la vérité-mer suggèrent les conclusions suivantes :

(a) Seul un modèle moyen, très approximatif, de la répartition angulaire de l'énergie des vagues peut couvrir les conditions météorologiques très variées dans lesquelles les données ont été recueillies. Il est alors possible d'en déduire une estimation convenable de la direction du vent, pourvu que la fréquence et la vitesse du vent soient assez élevées pour que les composantes de la surface de la mer correspondant aux raies de Bragg soient rapidement saturées par le vent à l'échelle de temps des variations de celui-ci.

(b) La fréquence dominante du spectre omnidirectionnel des vagues peut être déduite avec une bonne précision de la mesure des fréquences de coupure présentes dans le spectre Doppler de part et d'autre des raies de Bragg. Cette précision est de l'ordre de quelques centièmes de Hertz, sauf pour les valeurs élevées que l'on rencontre pour les mers peu agitées (H 1/3 << 1 m.).

Dans l'hypothèse d'une mer complètement développée, la hauteur significative se déduit de la fréquence dominante, mais la qualité réduite de cette estimation confirme qu'en général, un deuxième paramètre est indispensable pour décrire de façon satisfaisante le spectre omnidirectionnel. Du point de vue de la télédétection, ce paramètre est, par exemple, à rechercher dans le niveau relatif des échos du second ordre et du premier ordre (BARRICK, 1977 , FORGET, 1979).

REMERCIEMENTS

Les résultats présentés ci-dessus ont été acquis avec le support ou la collaboration du C.N.R.S., du C.N.E.X.O., de l'INSTITUT FRANCAIS DU PETROLE, de la COMPAGNIE FRANCAISE DES PETROLES et de la SOCIETE NATIONALE ELF AQUITAINE.

REFERENCES

BARNUM J.R., MARESCA J.W. and SEREBRENY S.M. (1977), High resolution Mapping of oceanic wind fields with skywave radar. IEEE Trans. AP 25, p. 128.

BARRICK D.E., HEADRICK J.M., BOGLE R.W. and CROMBIE D.D. (1974). Sea backscatter at H.F. : Interpretation and utilization of the echo. Proc. IEEE, 62, p. 673.

BARRICK D.E. (1977). Extraction of wave parameters from measured H.F. radar sea-echo Doppler Spectra. Radio-Science, 12, n° 3, p. 415.

BARRICK D.E., EVANS M.W. and WEBER B.L. (1977). Ocean surface currents mapped by radar. Science, 198, p. 138.

CROMBIE D.D. (1955). Doppler Spectrum of sea echo at 13,56 Mc/s. Nature, 175, p. 681.

de MAISTRE J.C., BROCHE P. and CROCHET M. (1977). Offshore wind measurements by H.F. Doppler ground wave radar. Boundary Layer Meteorology, 13, p. 3 .

FORGET P. (1979). Observation des vagues par radar décamétrique en ondes de sol : modélisation et résultats expérimentaux. Thèse de 3ème cycle, Université de Paris VI.

HASSELMANN K., BARNETT T.P., BOUWS E., CARLSON H., CARTWRIGHT D.E., ENKE K , EWING J.A., GIENAPP H., HASSELMANN D.E., KRUSEMAN P., MEERBURG A., MULLER P., OLBERS D.J., RICHTER K., SEEL W., WALDEN H. (1973). Measurements of wind-wave growth and swell decay during the joint North sea wave project (Jonswap). Erg. Zur Deutsch. Hydr. Zeitsch. Reihe A (8°), Nf 12. D.H.I. Hamburg.

LIPA B. (1977). Derivation of directional ocean wave spectra by integral inversion of second order radar echoes. Radio Science, 12, n° 3, p. 425.

LONG A.E. and TRIZNA D.B. (1973). Mapping of North Atlantic winds by HF radar sea backscatter interpretation. IEEE Trans. AP 21, p. 680.

MARESCA J.W. and J.R. BARNUM (1977). Measurement of oceanic wind speed from H.F. Sea scatter by skywave radar. IEEE Trans. AP 25, p. 132.

RAMAMONJIARISOA A. (1975). Note sur la prévision par des méthodes semi-empiriques des spectres d'énergie des vagues de vent. Annales Hydrographiques, 5ème série, Vol. 4, fasc. 1.

STEWART R.H. and JOY J.W. (1974). H.F. radio measurements of surface currents Deep Sea Research, 21, p. 1039.

STEWART R.H. and BARNUM J.R. (1975). Radio measurements of oceanic winds at long ranges : an evaluation. Radio Science, 10, p. 653.

TYLER G.L., TEAGUE C.C., STEWART R.H., PETERSON A.M., MUNK W.H., and JOLY J.W. (1974). Wave directional spectra from synthetic aperture observations of radio scatter. Deep Sea Research, 21, p. 989.

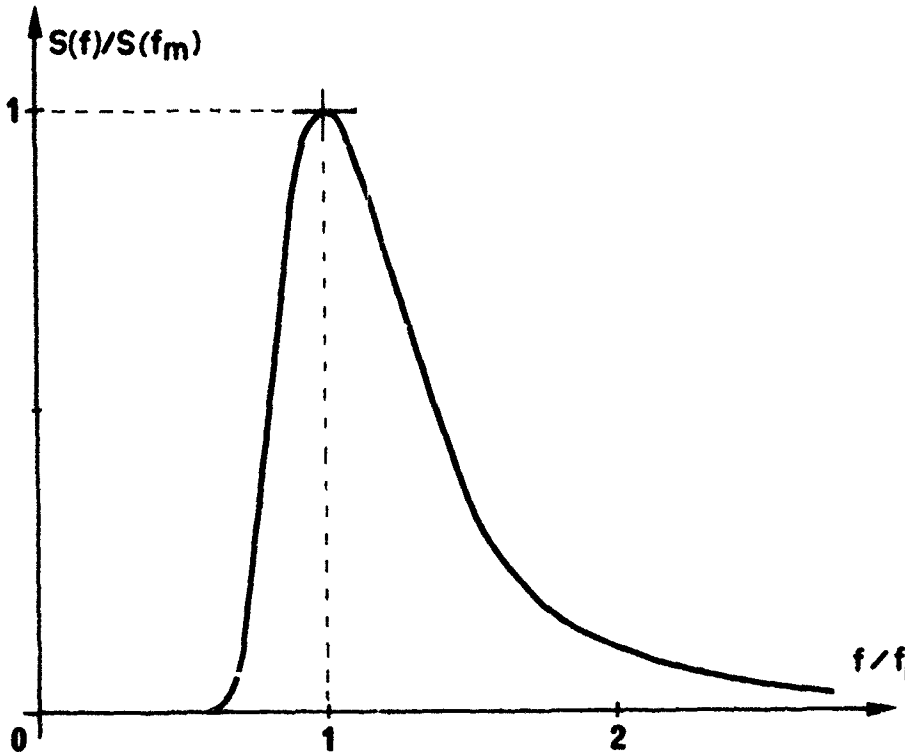


Fig. 1 - Spectre de Pierson-Moskowitz (f_m est inversement proportionnel à la vitesse du vent U ; pour U = 10 m/s, on a f_m = 0,137 Hz).

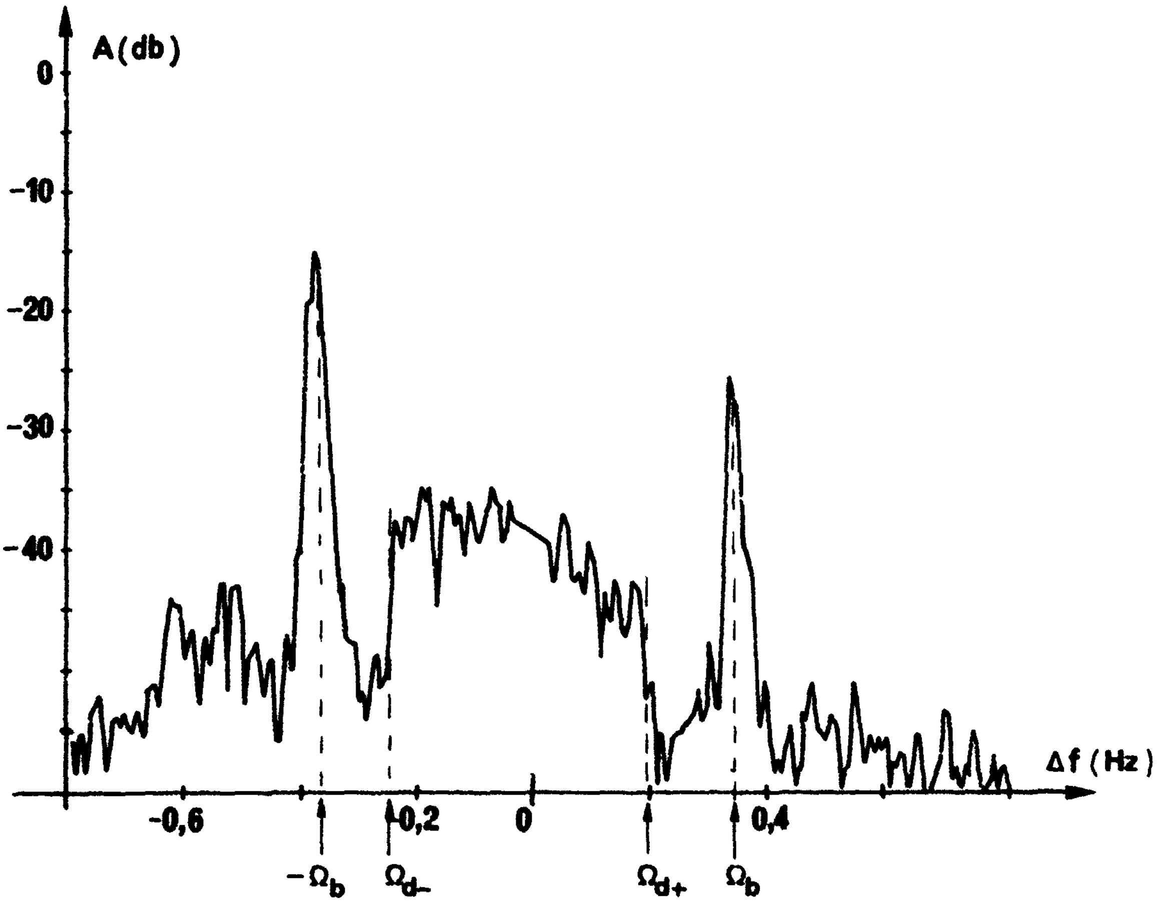


Fig. 2 - Exemple de spectre radar. La fréquence radio est 12,4 MHz. $\pm\Omega_b$ représentent les raies de Bragg, $\Omega_{d\pm}$ les fréquences de coupures citées dans le texte (§ 4.2).

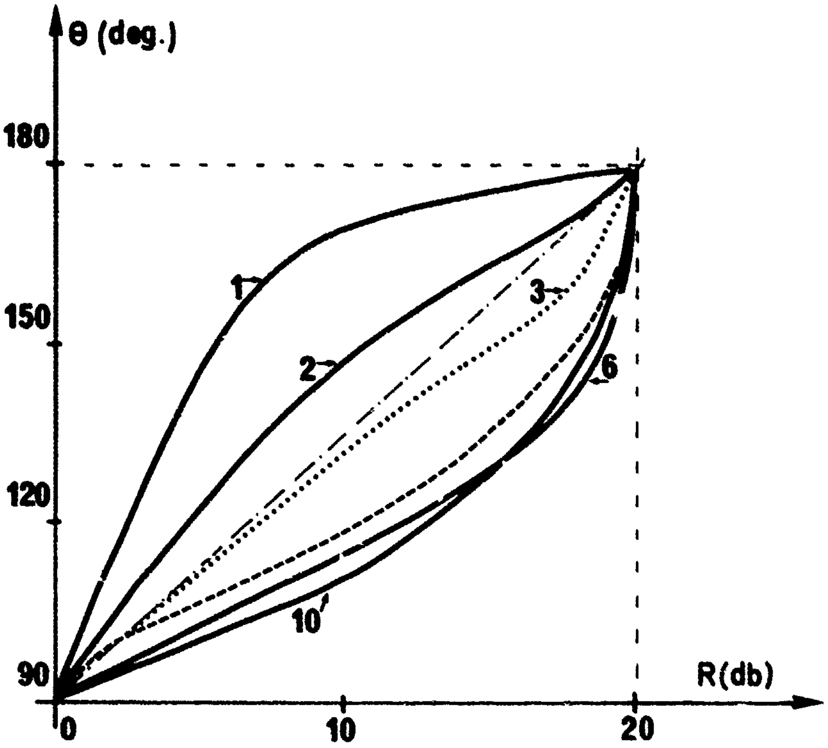


Fig. 3 - Variations du rapport $R = A_+/A_-$ en fonction de l'angle θ entre la direction du vent et l'axe du faisceau radar. Le modèle LCS est représenté par les courbes en trait plein et la courbe en pointillé (la valeur de s est indiquée par une flèche). Le modèle $(1 + b\cos^2\theta)$ est représenté par la courbe en tirets.

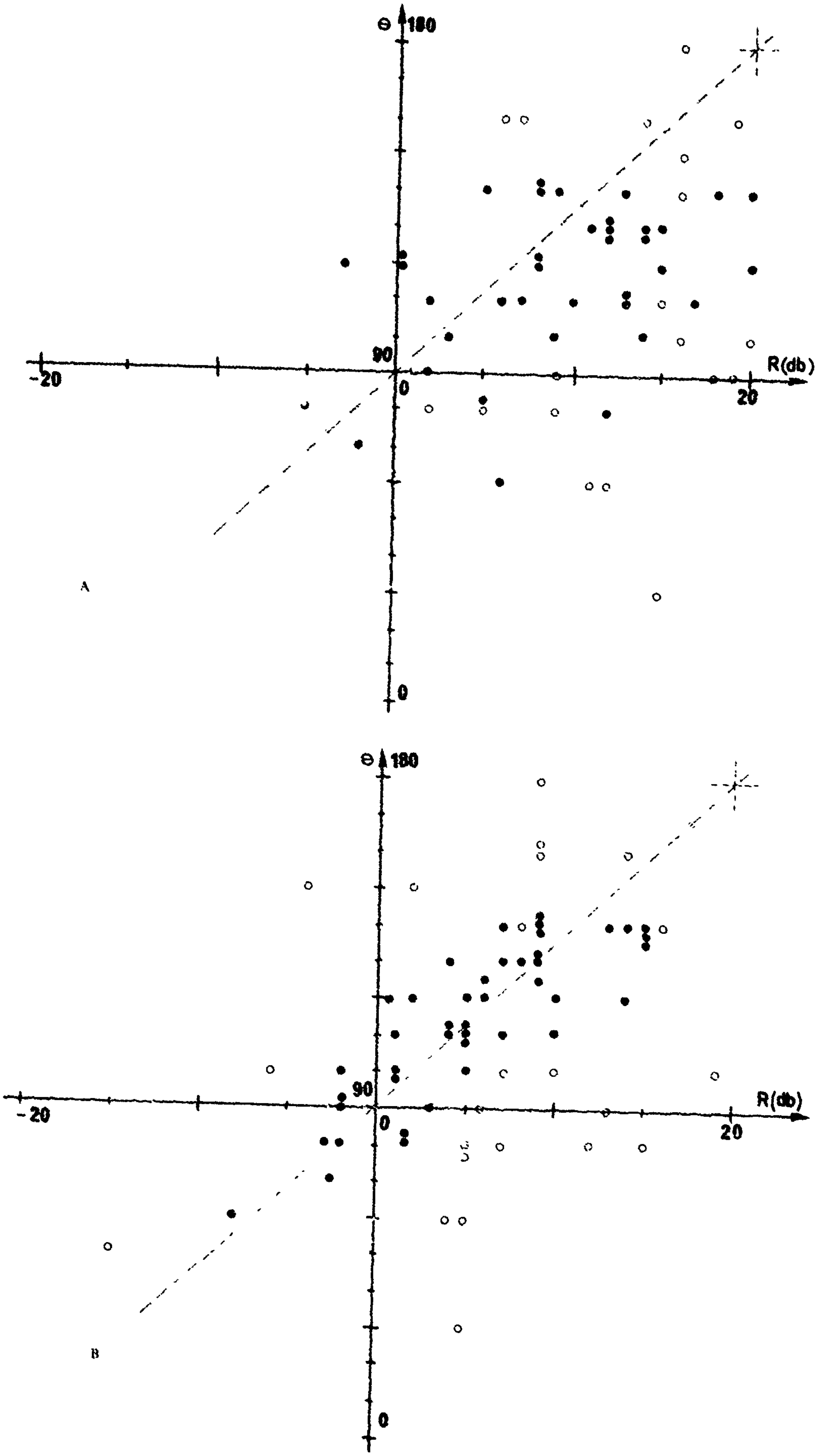


Fig. 4 - Résultats expérimentaux : R en fonction de θ pour une fréquence radio de 6 MHz (4.a) et pour une fréquence de 12 MHz (4.b). Les cercles correspondent à $H_{1/3}<1$ m

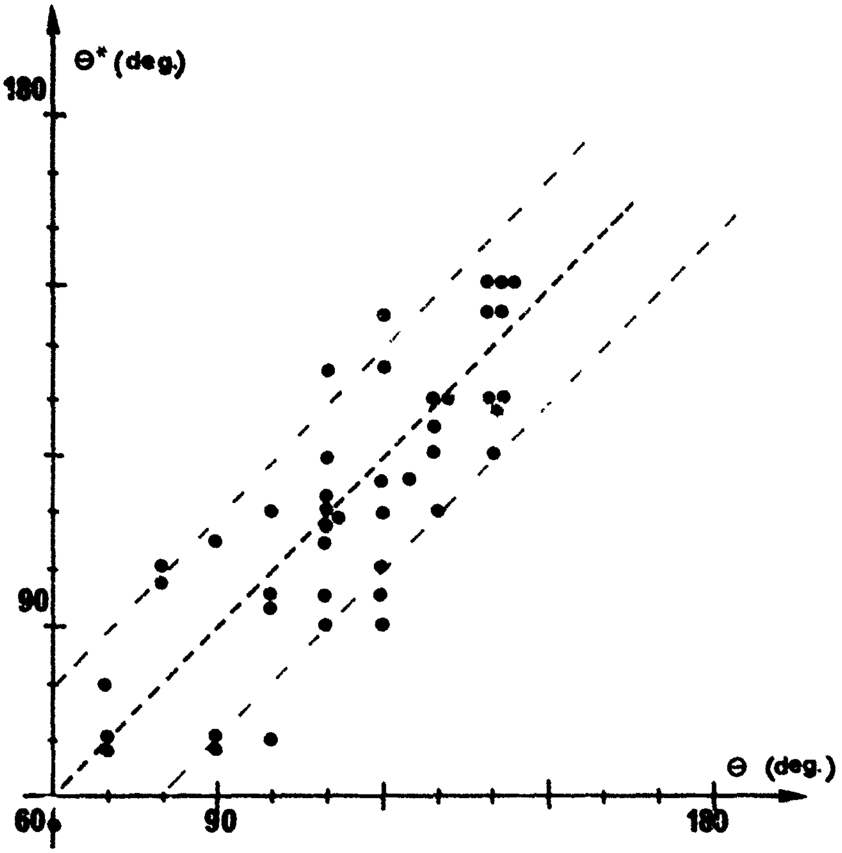


Fig. 5 - Valeur de l'angle entre la direction du vent et celle du faisceau radar θ^*, estimée d'après la valeur de R, en fonction de la valeur θ, mesurée par un anémomètre.

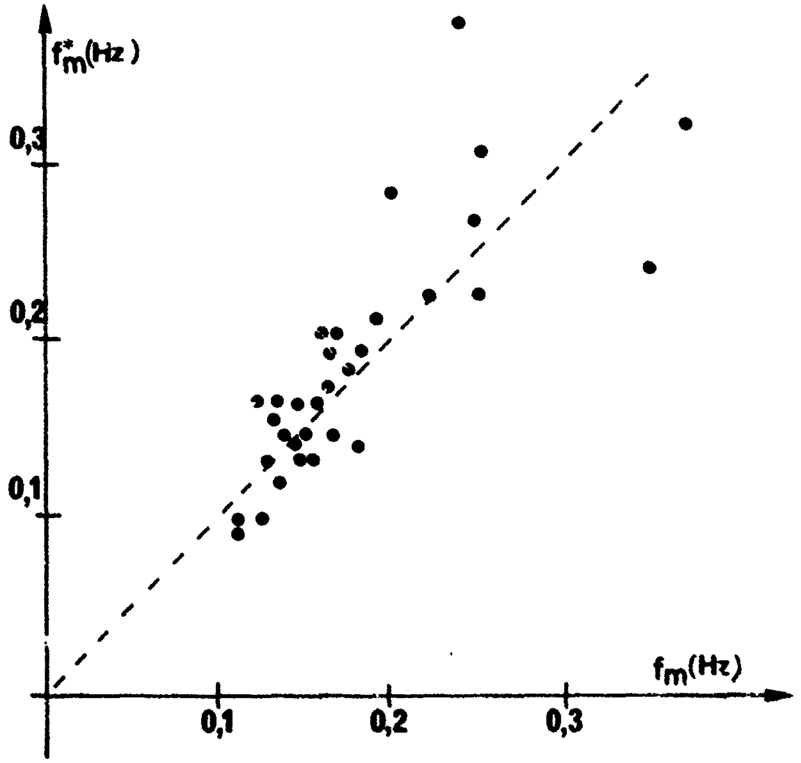


Fig. 6 - Valeur de la fréquence dominante des vagues f_m^* estimée d'après les spectres du second ordre en fonction de la valeur f_m mesurée par une bouée.

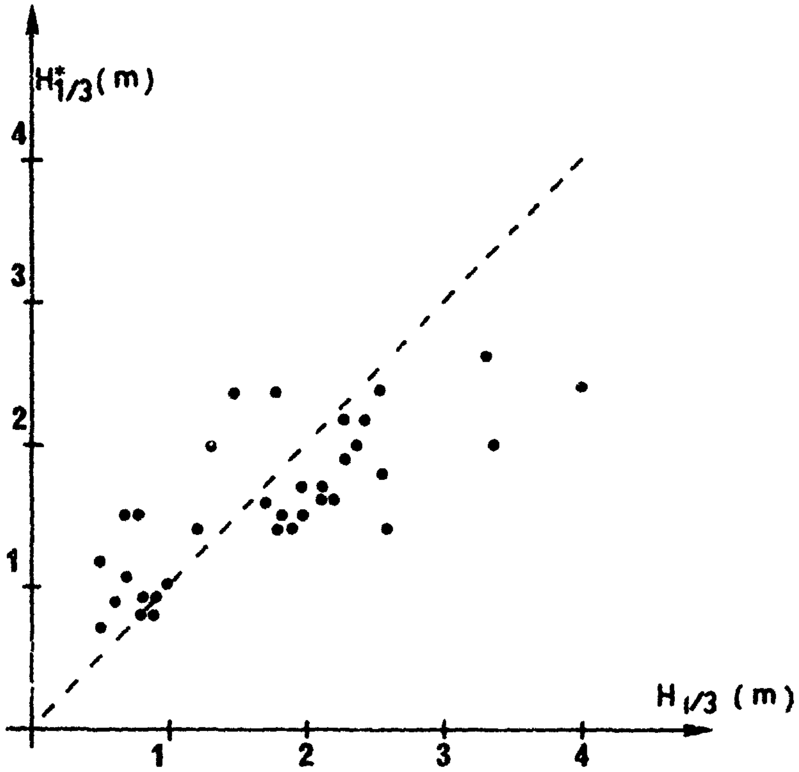


Fig. 7 - Valeur de la hauteur significative $H1/3^*$ estimée en fonction de la valeur $H1/3$ mesurée par une bouée.

DEVELOPMENT OF HF SKYWAVE RADAR FOR REMOTE SENSING APPLICATIONS

Taylor W. Washburn, Lawrence E. Sweeney, Jr.,
James R. Barnum, and Walter B. Zavoli
SRI International
Menlo Park, California 94025

SUMMARY

HF skywave radar systems rely upon ionospherically propagated (skywave) radio waves for remote sensing at very long ranges and over wide areas. The Wide Aperture Research Facility (WARF) in California provides an experimental testbed for HF radar design and evaluation. Uses of the WARF skywave radar include the tracking of aircraft and ships, sea state monitoring, and ionospheric research. In this paper we discuss the capabilities of this radar, as well as various radar techniques and propagation management tools used--generally on-line and in real time--to maximize radar performance.

A central feature of WARF is its 2.5-km-long receiving aperture formed by 256 pairs of vertically polarized monopoles spaced equidistant over the 2.5-km distance and connected to a common processing point through cables and a switching network. This electronically steerable receiving array (±32° about its east-west boresight) has selectable reception from either direction, and operates over 6 to 30 MHz, producing a 0.5° azimuthal beam at midband. Combined with a bistatically located transmitter utilizing steerable broadbeam (6°) illumination and a swept-frequency continuous wave (SFCW) signal format, the HF radar provides very high resolution in azimuth and range.

Backscattered echoes from targets, land clutter, sea clutter, and noise are separated from each other in range and in radial velocity by time-delay and Doppler processing of the received SFCW signals. Various signal-processing techniques are employed for minimization of other-user interference. These techniques include signal blanking, waveform randomization, and excision of undesirable portions from the processed time- and frequency-domain data. Spatial rejection of out-of-beam energy has been explored through application of adaptive beamforming techniques.

Propagation management and performance monitoring are two auxiliary functions essential to the successful operation of an HF radar. Properly executed, these functions maximize the returned target signal relative to noise and clutter, minimize multipath, avoid RFI to and from other HF users, and provide accurate conversion to geographic coordinates. At WARF these requirements are met through use of a separate backscatter ionogram sounder that produces clutter-to-noise ratio maps over a wide area, a spectrum surveillance system that searches for quiet portions of the HF spectrum, real-time measurements on the received radar signal (clutter, noise, and transmitter-off noise), and a steep incidence (V-I) sounder. Examples of remote sensing measurements are included to demonstrate HF radar system performance and to illustrate the integration of essential auxiliary functions to assure maximum overall performance.

1. INTRODUCTION

The Wide Aperture Research Facility (WARF) in California has explored the benefits of high range and azimuth resolution--up to 1.5 km in range and 0.5° in azimuth--for over-the-horizon (OTH) radar. Applications include tracking of aircraft and ships and monitoring of land and sea--in particular ocean wind and wave spectrum properties. OTH radar which is also termed HF skywave radar, relies on ionospherically propagated radio waves to detect targets at very long range and over large areas. Radar ranges from 700 km to in excess of 3000 km are typical for one-hop ionospheric propagation via conventional modes. Large, fixed antennas in common use at HF achieve 60° sector coverage implying total area coverage of about 3×10^7 km². Radars operating at these relatively low frequencies must deal with problems of high ambient noise, large antenna beamwidths, and limited bandwidth due to spectrum crowding and ionospheric dispersion. In spite of these difficulties the long range/wide area coverage coupled with the capability for essentially continuous surveillance makes OTH radar an attractive device for remote sensing. A review of OTH radar technology may be found in Headrick and Skolnik (1974).

In this paper we first present the general capabilities of WARF followed by a description of several performance maximizing techniques in use there including:

- Signal processing of OTH radar signals (Section 3).
- Interference avoidance and minimization through frequency/time filtering and adaptive beamforming (Section 4), and
- Clutter-to-noise monitoring for prediction of radar sensitivity (Section 5).

2. WARF OVERVIEW

A unique feature of the WARF is a 2.5-km long receiving aperture formed by 256 paris of vertical monopoles spaced equidistance and connected to a common processing point through delay cables and a reed relay switching network. Array steering is available in 1/4° increments over ±32° from the east-west boresight direction. Achievable coverage area is depicted in the map of Figure 1. The array has east-west selectable reception since each array element is a pair of 5.5 m monopoles spaced 4.7 m along boresight and combined with appropriate delay cable. On average the front to back ratio is 13 dB. Operating frequencies vary from 6 to 30 MHz and, although the elements are not impedance matched, the array is externally noise limited at frequencies below 25 MHz. This is due in part to the use of 1 dB noise figure preamplifiers inserted at the point where one-eighth the array elements are combined and steered. Azimuthal beamwidth is 0.5° at 15 MHz; amplitude weighting in the beamformer produces average sidelobe levels of -25 dB. The vertical pattern is controlled by the twin-monopole pattern whose -3 dB points at 15 MHz are 9° and 45°; low-angle gain is limited by the use of a small ground screen (22 m wide with 0.6 m mesh). Although highly variable with frequency and elevation angle, array directive gains exceed 30 dB between 8 and 28 MHz and 8° and 35° elevation.

The WARF transmitter is located slightly bistatically (185 km to the southeast) to permit use of a swept-frequency continuous wave (SFCW) signal format. An eighteen-element array of vertical log periodic elements (pointed west) or folded triangular monopole elements (pointed east) spanning a 205 m aperture achieves a 6° azimuthal beamwidth at 15 MHz. Transmit gain exceeds 20 dB at the peak of the vertical pattern. The transmit array is delay-cable steerable over ±30° in 4° steps; average sidelobe levels are -20 dB or better. Ground screen of 180 m width helps maintain gain at low elevation angles.

For radar applications two servo-tuned 10 kW (avg.) power amplifiers are employed. When losses are included, a realistic average value of power product ($P_TG_TG_R$) is 93 dBW over the range of frequencies and elevation angles in use.

Ionospheric backscatter sounding is normally carried out with a 10 kW transmitter feeding a single log periodic antenna and by receiving on the full receiving array, 1/8 of this array, or a horizontal, rotatable log periodic antenna. Steep-incidence ionospheric soundings between the two WARF sites are routinely interleaved with backscatter soundings. Thus both types of soundings are generated simultaneously with radar operations. A summary of the WARF specifications is presented in Table 1.

Table 1

WARF OTHR SYSTEM PARAMETERS

	Main System		Backscatter Sounder	Vertical-Incidence Sounder
● **Transmitting Site**				
Power	20 kW CW		10 kW CW	10 W
Antennas	East 18-Element FTM Array, 9-26 MHz		Rotatable Horizontal LPA	Delta
	West 18-element LPA Array, 6-30 MHz Both 205 m Long		6-30 MHz	2-13 MHz
	6° Azimuthal Beamwidths at 15 MHz			
Directive Gain	~20 dBi (at 15 MHz)		~13 dBi	~5 dBi
Coverage	±32° from E/W in 4° Steps			
● **Receiving Site**				
Antenna	2.5-km Array of 256 Whip Pairs 0.5° Azimuthal Beamwidth at 15 MHz			Delta
Directive Gain	~30 dBi			~5 dBi
Coverage	±32° from E/W in 0.25° Steps Electrically Steerable to East or West			
Operating-Frequency Selection	Spectrum Surveillance System			
● **Combined System**				
Power Product	$P_TG_TG_R$ = 93 dBW at 15 MHz			
Waveform	SFCW			
	Ships	Aircraft		
Range Resolution	3 km	18 km		
Integration Time	12.8 sec	2.1 sec		
Dwell Period	25.6 sec	2.1 sec		
Cycle Time	30 sec	7.5 sec		
No. Receive Beams	5 (× 0.25°)	3 (× 0.50°)		
No. Range Cells	21	12		
Nominal Area Coverage Rate	4,100 km²/min	72,000 km²/min		
Automatic Detection and Tracking	Yes	Yes		

Critical factors in successful operation of an OTH radar include maintenance of high signal-to-noise ratio (SNR) and signal-to-clutter ratio (SCR), conversion of radar coordinates to ground coordinates, identification of multipath, and avoidance and/or suppression of radio-frequency interference. Adequate SNR is achieved by controlling the quantities in the standard radar equation, wherein the notable aspects at HF are the high levels of external noise and the loss term which includes propagation effects (polarization mismatch, absorption, ionospheric focusing, ionospheric disturbances and irregularities).

Since OTH radars are essentially "look-down" radars, SCR is crucially important for detection of point targets. SCR is increased by reducing the pulse width and narrowing the antenna beam; however, even with 10 kHz bandwidth and a 1/2° beam the clutter may exceed 10^6 m^2. Further clutter rejection is achieved with Doppler processing; filter widths typically range from 1 Hz to less than 0.01 Hz depending on the application and on the nature of the propagation path. Maximizing SCR is not of interest in the case of remote sensing of the ocean since the clutter itself is the target. Here the primary requirement is for long integration time (50 to 100 s or more) to reveal the detailed structure in the clutter Doppler spectrum while avoiding the appearance of Doppler-shifted spectrum replicas produced by ionospheric multipath.

3. WAVEFORM AND SIGNAL PROCESSING

A sweep-frequency continuous wave (SFCW) signal is employed at WARF either in the form of an extended linear frequency ramp or as a repetitive, narrow frequency ramp. SFCW is a useful waveform in that it permits 100% duty cycle transmissions, is amenable to operating in a high interference environment and may be readily processed off-line with variable parameter values such as range and Doppler resolution. The extended linear frequency ramp generates steep-incidence or backscatter ionograms by processing the received signals for amplitude vs. time delay as a function of radio frequency. The essence of this processing consists of mixing or deramping the received signal with a local version of the sweep generator, amplifying and down converting to audio frequencies and then carrying out spectrum analysis on the result. In the WARF system increasing audio frequency corresponds to increasing time delay (radar range).

The repetitive waveform is processed for echo magnitude vs. time delay and Doppler frequency at a single (narrowband) radio frequency. Spectrum analysis on the signals received during a single sweep provides magnitude vs. radar range; maintaining sweep-to-sweep phase coherence and carrying out spectrum analysis over many sweeps provides Doppler information for each range cell.

A simplified block diagram of the SFCW radar system is presented in Figure 2. Intersite synchronization is maintained within 10 μs using the ground wave signal propagating between the two sites. The receiver is gain and phase stable (within 0.5 dB and .5°) and is operated over 3-30 MHz (Phillips, A. C. and C. A. Cole, 1974). Due to large variations in received signal strength 100 dB of automatic gain control is provided. Noise figure is <3 dB. For radar target processing, the audio signal receiver output (sometimes termed offset video) is digitized for spectrum analysis in a minicomputer.

Figure 3 sketches the instantaneous frequency vs. time of the transmitted waveform where a bandwidth W is swept every T_r s. The receiver output signal is sampled at f_s Hz properly synchronized so that N samples are captured for M sweeps. An additional constraint on f_s is that the Nyquist criterion is satisfied, i.e. $f_s > W_r/2$, where W_r is the final receiver bandwidth. The total coherent processing time T is MT_r, so that the waveform time-bandwidth product is TW. These basic waveform parameters are further defined in Table 2 along with derived quantities of achieved coverage and resolution for range and Doppler. Typical values of radar parameters used for aircraft and ship tracking at WARF have been listed in Table 1. Without compensation for target movement, maximum SNR is achieved if the targets accelerate slowly enough and do not traverse range cells within one coherent integration period. Further, certain approximations must be satisfied to avoid signal distortion (Barrick, D. E., 1973).

Table 2

RANGE/DOPPLER PROCESSING OF REPETITIVE SFCW SIGNALS

Basic Parameters

W = Swept bandwidth

T_r = Waveform repetition period ($= 1/f_r$)

N = Samples/sweep

M = Sweeps coherently processed

T = Coherent integration period ($= MT_r = MN/f_s$)

f_s = Sample frequency ($= Nf_r$)

W_r = Receiver bandwidth ($f_s > 2W_r$)

Derived Quantities

Time delay coverage ($= T_r W_r/W$)

Time delay resolution ($= 1/W$)

Unambiguous Doppler coverage ($= \pm f_r/2$)

Doppler resolution ($= f_r/M$)

Prior to spectrum analysis of NM samples of the repetitive sweep signal, amplitude weighting is applied individually over both the N samples within each sweep and over the M sweeps. Hanning weighting is most commonly used at WARF so that the sidelobes in time delay and in Doppler are below -50 dB five resolution cells away from the delay-Doppler cell of interest. Figure 4 plots the magnitude of the spectrum computed for a radar example where $W = 8.33$ kHz, $T_r = 16.6$ ms, $N = 32$, and $M = 128$. Thus integration time T is ≈ 2 s and the time bandwidth product is ≈ 1.6×10^4. The spectrum has strong clutter lines spaced at the waveform repetition frequency, f_r. Successive f_r lines correspond to successive range cells and the frequencies between correspond to Doppler frequencies. Assuming certain low-noise, low-loss equipmental requirements are met, the signals at frequencies a few Hz removed from the clutter lines usually represent external noise. The major exception to this condition is produced by radar scatter from aurora-induced ionospheric irregularities. Intense auroral clutter with Doppler spread of tens of Hertz is often observed.

For [illegible] of interpretation one may align successive range cells atop each other to produce a range-Doppler ma[illegible] his display can be used for detecting aircraft or ships since range progression with time is easily discerned. Actual target tracking employs automatic detection algorithms for improved sensitivity and to reduce operator workload. Figure 5(a) is a typical aircraft detection example plotting linear magnitude vs. Doppler for each range line; the magnitudes are purposely clipped 20 dB above the average noise level to maintain display sensitivity. An aircraft echo is visible at -25 Hz. As shown in Figure 5(b), the display is improved through application of post whitening separately for each Doppler cell in order to remove spread-range signals such as clutter or RFI. This process eliminates distractions to improve the visibility of discrete-range targets.

A companion data format reveals Doppler width and relative strength characteristics of clutter, noise, and targets by collapsing range information through an RMS average over the range cells at each Doppler frequency. The resulting range averaged power spectrum is plotted in Figure 5(c) on a logarithmic scale, facilitating comparisons of strong clutter and weak noise signals.

Range-Doppler maps and power spectra are the two most common methods for viewing repetitive-sweep backscatter data at WARF The waveform parameters of Figure 5 are typical for detecting high speed targets such as aircraft, meteors, or missiles at the expense of loss of detail in the clutter echoes. Clutter fine structure or targets near or superimposed on the clutter are best detected through increased bandwidth W and longer integration time T; WARF commonly employs time-bandwidth products of 10^4-10^7. Figure 6 is an example of the sea-echo Doppler spectrum (Maresca, J. W., Jr. and C. T. Carlson, 1978) for W = 50 kHz, T_r = 200 ms, N = 16, M = 256. The resulting 51.2 s integration period reveals a great deal of detail in the radar backscattered sea-echo. The strong Bragg-line echoes may be analyzed to infer surface wind direction, wind speed, and surface currents. The weaker second-order structure may be employed to infer sea state and the presence of swell (Barrick, D. E., et al., 1974).

Since ocean surface scatter is a random process, noncoherent averaging in time and in space improves the reliability of the computed spectral features.

4. RADIO-FREQUENCY INTERFERENCE

Operating frequencies for OTH radars vary diurnally by a factor of five or more in order to maintain adequate illumination of the desired coverage area. Since the HF spectrum is already crowded, it is imperative to minimize RF interference both to and by the radar. Considerable effort has been undertaken at WARF to avoid and minimize the effects of RFI. The techniques and procedures studied are briefly described in the following.

4.1. Clear-Channel Search Receiver

During radar operations a frequency scanning receiver, fed by either the highly directive radar antenna or by a less directive antenna, is monitored to locate a clear channel within the frequency band that will propagate to the desired radar coverage area. This band generally ranges from 1 to 5 MHz in width. The desired channel width of 10 to 100 kHz is usually available when the desired center frequency is above about 14 MHz; however at low frequencies (6-10 MHz), and especially at night, the usable frequency band is compressed so that clear channels of 100 kHz extent are usually nonexistent. At these times the radar sweep generator may be adjusted to use whatever channel width is available at the expense of increased clutter and lower time delay resolution. Alternatively, one or more of the interference rejection methods described below are utilized.

4.2. Time Domain Weighting of the Transmitted Signal

The transmitted frequency spectrum depends in a complex manner on swept bandwidth, repetition rate, frequency- and phase-flyback characteristics among others. The amount of signal radiated outside the actual swept bandwidth depends on the details of the method of waveform generation. For example, at WARF it has been found that, if the sweep generator resets to its start frequency while maintaining continuous phase, the transmitted signal frequency sidebands are significantly lower than if phase discontinuities are present. However, simple time domain weighting over the waveform repetition interval is an effective means of minimizing frequency-sidelobe interference to other HF users. A reasonable degree of such weighting results in about 1 dB loss in average transmitted power.

4.3. Time Domain Blanking

Interfering signals discrete in frequency or in time can be discriminated against at the receiver by a variety of excision techniques. One technique in current use at WARF is to discriminate against interfering signals that occupy only a small portion of the swept bandwidth by blanking out a portion of the received waveform (Barnum, J. R., et al., 1979). The blanking template is "learned" by receiving for a period of time out of synchronization with the transmitter sweep and then applying appropriate smoothing to the template to minimize sidelobes when the received signal is spectrum analyzed. An improved intersite communications capability is currently being installed at WARF, which will permit application of the template to the transmitted signal as well, reducing interference to other users.

A similar excision technique has been applied in the signal processor (i.e., following A/D conversion) to remove impulsive noise bursts from atmospherics. In one continuous 24-hour experiment at WARF, 40% of the radar dwell periods were noticably contaminated by atmospherics. An automatic time domain signal excision algorithm was applied before spectrum analysis to improve the median signal-to-noise ratio at the output of the radar processor by 6 dB (McKinney, D. and W. B. Zavoli, 1974).

4.4 Phase Decorrelator for Sweep Generator

Since coherent processing is carried out on a sweep-to-sweep basis, the sweep generators must be coherent with each other. If an interfering signal is phase stable its carrier may create a coherent (discrete-Doppler/distributed-range) interference "line" on the range-Doppler map. The sweep generator start phase is stepped in quadratic fashion from sweep to sweep, identically for both transmit and receive generators, spreading the energy from the coherent interferer over all Doppler frequencies. The net effect is that the Doppler blind speed created by the interferer is removed at the expense of moderately increasing the average noise level. For aircraft detection the Doppler processing gain is about 20 dB; randomization of generator start phase reduces the peak interference level 20 dB in the radar output by spreading it over all Doppler cells. The acceptable increase in average noise level determines how strong an interferer this technique may be used against. Pseudo-random start phase stepping has also been tested, but quadratic start phase stepping performs as well and is simpler to implement.

4.5. Post Detection Bias Removal For Each Doppler Cell

At the output of range-Doppler analysis, interfering signals with time coherence over the period of a frequency sweep appear as constant-Doppler, spread-range signals. These RFI signals add a relatively unfluctuating bias component to targets and to the Rayleigh-distributed clutter and noise. The presence of RFI is detected by computing the ratio of mean to variance over all range cells separately for each Doppler cell. Ratios exceeding a calculable threshold imply a bias. For each of these Doppler cells an algorithm is applied to subtract the mean bias computed at that Doppler from the amplitude at each of the range cells at that Doppler. Thus the bias of the range-correlated interferer is removed while, on the average, only minimally affecting discrete echoes from targets. Hence the name bias removal interference suppression algorithm (BRISA). Figure 7 depicts an example of a 2 s radar dwell period processed with and without BRISA. A large interferer at -20 Hz Doppler with about 23 dB SNR is virtually completely removed through application of BRISA, without affecting detection sensitivity at other Dopplers. Application of this algorithm in radar operations at WARF show up to 30 dB of interference suppression, the amount depending on the coherence properties of the interferer and on the receiver passband characteristics.

4.6. Adaptive Antenna Beamforming

In addition to the RFI suppression techniques described above which involve time and frequency domain filtering, antenna beamforming techniques have been explored at WARF utilizing adaptive algorithms to reject unwanted out-of-beam energy (Griffiths, L. J., 1976; Washburn, T. W. and L. E. Sweeney, Jr., 1976). For this work the 2.5 km receiving aperture was split into eight 32-element subarrays of 320 m spacing whose outputs fed an eight-channel, gain and phase matched receiving system. Both deterministic (25 dB Dolph taper) and adaptive beamformed outputs were generated digitally from the A/D converted receiver outputs. Figure 8 is a block diagram of the digital array processor, where each of eight processing channels is passed through a tapped delay line filter. The summed result is the beamformer output which is then processed with the usual radar signal processing operations. Beamformer weights have been computed by several techniques including (1) the Griffiths P-vector algorithm (Griffiths, L. J., 1969), (2) a linearly constrained least squares procedure (Frost, O. L. III, 1972), (3) the optimal technique involving the inverse of the signal covariance matrix which assures minimum mean-square error for the case of stationary statistics, and recently (4) a generalized sidelobe cancelling technique (Griffiths, L. J., and C. W. Jim, 1978).

Table 3 presents equations for the "optimal" beamforming coefficients and the P-vector algorithm. Conventional coefficients are fixed in time and computed a priori. The optimal coefficients are data-dependent but piecewise fixed in that they use correlation coefficients computed from the actual signal environment over some time interval. The recursive P-vector coefficients are a continually updated approximation to the optimal coefficients but they have the additional property of following nonstationarity in the statistics of the signal environment.

Table 3

BEAMFORMING EQUATIONS

Conventional

$$y(n) = \underset{\sim}{A}^T \underset{\sim}{X}(n)$$

$\underset{\sim}{A}$ is a vector of Dolph Coefficients

$\underset{\sim}{X}(n)$ is a vector of signals received at elements X_i $(i = 1,2,\ldots)$ as a function of time n.

Time Domain Adaptation (Griffiths)

$$y(n) = \underset{\sim}{W}^T(n)\underset{\sim}{X}(n)$$

$$\underset{\sim}{W}(n+1) = \underset{\sim}{W}(n) + \mu\left[\underset{\sim}{P}_{Xd} - y(n)\underset{\sim}{X}(n)\right]$$

Covariance Matrix Inversion (MMSE, Stationary Statistics)

$$\underset{\sim}{W}_{opt} = \underset{\sim}{R}_{XX}^{-1}\underset{\sim}{P}_{Xd}$$

$$\underset{\sim}{R}_{XX} = \underset{\sim}{X}(n)\underset{\sim}{X}^T(n)$$

(Each element of $\underset{\sim}{R}_{XX}$ is a correlation coefficient computed as a time average over M time samples.)

Realistic signal environments were created by operating a radar repeater at 1300 km range in New Mexico as the in-beam desired signal and by using either a second radar repeater or a separate radio source transmitting within the radar operating bandwidth as the out-of-beam unwanted signal. This unwanted signal was located at essentially the same range as the wanted signal but with 2° azimuthal separation. In the former case the out-of-beam repeater simulates a point source of clutter and in the latter simulates an other-user interferer (both CW and AM signals have been employed in testing). Figure 9 is a sample comparison of deterministic and adaptive beamforming for the case of two repeaters generating the desired and unwanted signals. The repeater gains were adjusted to yield essentially equal signal strength when pointing at the desired signal with the 25 dB Dolph taper beamforming. The unwanted repeater signal was decreased an additional 15 dB with adaptive beamforming. Here the Griffiths P-vector algorithm was employed using 8 elements and 4 taps per element. The adaptive beamformer also cancels some of the ground clutter as evidenced by a 10 dB reduction in the clutter to noise ratio.

Up to 20 dB additional rejection of point-source clutter has been measured for an adaptive beamformer operating with a realistic signal environment, although 10-15 dB is more typical. Similar rejection has been observed for the case of other-transmitter RFI. Gain and phase errors in subarray beamforming hardware limit the conventional sidelobe levels to about -25 dB; it is noteworthy that the same errors exist for the adaptively beamformed signals.

Comparison of time domain recursion and covariance matrix inversion as methods for computing beamformer weighting coefficients shows that comparable rejection can be achieved, but the covariance matrix performance varies with the time location and quantity of data used for computing the required correlation coefficients. Further it has been observed that time-domain algorithms can impart time modulation effects in the range-Doppler analysis normally following beamforming. However, such effects are minimized through proper control of the loop gain (μ in Table 3). A consideration for implementation of an adaptive beamforming capability is that the matrix inversion approach requires more processor speed and capacity.

Recent work by Griffiths has shown promising performance with a generalized sidelobe cancelling structure (Griffiths, L. J. and C. W. Jim, 1978) allowing the implementation of a variety of time domain recursive algorithms (Frost's algorithm in particular). The canceller generates an error beam through spatial preprocessing of the signals of the signals received at the elements to remove the desired signal components. An adaptive algorithm computes a least squares best estimate of the remainder which is then subtracted from the output of a conventional beam. This beamformer is currently being implemented for real-time use at WARF with the aid of a high-speed hardware array processor.

5. PROPAGATION DIAGNOSTICS

The complexity and highly variable nature of ionospheric radio propagation requires careful attention in matching radar operating parameters to the dynamics of the ionosphere. An overview of the interaction of propagation characteristics with radar operation is found in Headrick and Skolnik (1974). At WARF we have found real-time monitoring essential to making proper range gate and operating frequency selection for the purposes of (1) maximum illumination of the desired coverage area, (2) minimization of multipath, and (3) conversion of ionospheric coordinates to ground coordinates.

The monitoring procedure consists principally of steep-incidence ionograms which are interpreted with the well-known transmission curves (Davies, K., 1965) to extrapolate to oblique paths. This technique assumes horizontal stratification of the ionosphere and has been found to work well at the temperate latitudes of WARF. Tilts caused by diurnal effects and medium-to-large-scale ionospheric disturbances are sensed with oblique backscatter ionograms, since the minimum delay focusing contours have been found sensitive to such tilts (Croft, T. A., 1972; Georges, T. M. and J. J. Stephenson, 1968). Steep-incidence and oblique backscatter ionograms are typically alternated every 5 min at WARF with equipment separate from the radar equipment so that radar operations are undisturbed.

Conventional ionograms do not yield measurements of various loss mechanisms--absorption, F-layer obscuration, defocusing and disturbance effects; however, the clutter from the sea in particular may serve a valuable radar calibration role. For a saturated sea the total radar cross section is relatively constant with radio frequency and elevation angle for large enough wind speeds and for a given wind direction. Whereas the total cross section may vary significantly with the Phillip's equilibrium constant, (α), the directional distribution of the ocean waves, and the radar-to-wind direction; it can be relatively constant over large areas for many hours. It is currently conjectured that given the ocean wave parameters--potentially measurable by the radar itself--the total cross section may be inferred within 3 dB (Maresca, J. W., Jr., 1979). A first order estimate for the average total radar cross section including propagation effects such as polarization losses is -23 dB. The clutter cell size is known ($= c \times R \times \theta \times \tau \times \sec(\beta)/2$) so the ratio of total clutter to noise (CNR) measured with the radar may be used to infer the SNR of a point target of known cross section co-located with the clutter cell, i.e. undergoing the same path losses and utilizing the same values of antenna gain. CNR monitoring has been very useful in performance assurance for aircraft and ship tracking; when CNR falls below a specified threshold new operating parameters are considered.

CNR measurements are more difficult to interpret over land because cross sections may vary as much as 30-40 dB from place to place and because the relatively high Brewster angle for land scatter produces large cross section variation with elevation angles below about 15°. It is also expected that CNR monitoring would be substantially more difficult to interpret in the high latitude ionosphere.

It should be emphasized that simple correspondence between CNR and SNR relies on the fact that the same system characteristics are used in deriving both quantities. This is of course the case when both quantities are derived from radar data.

Recently at WARF a backscatter sounder has been modified to produce maps of CNR in range and frequency (essentially backscatter ionograms calibrated in CNR). If sounder parameters can be related to radar parameters, these measurements may be converted to estimates of achievable radar sensitivity, i.e. maps of minimum detectable cross section (MDCS). The azimuthal dimension may be added through time sequential soundings at various steer angle positions. From radar equation arguments

$$\mathrm{CNR} = \frac{P_T G_T G_R \lambda^2 \sigma_o c\theta}{(4\pi)^2 R^3 LFkT_o 2Wb}$$

Most quantities are easily converted between radar and sounder with the possible exception of the

The sounder modification employed at WARF to measure clutter and noise separately is to use a second, matched receiver properly offset in frequency to sweep the HF spectrum a fraction of a second ahead of the receiver tuned to receive backscatter. Both receiver outputs are digitized. The noise channel is analyzed to locate interference-free channels whose content provides a measure of the noise power. The clutter channel is continually spectrum analyzed to provide a range profile of clutter vs. radio frequency; normalization by the noise produces CNR.

Figure 10 is the superposition of a conventional backscatter ionogram with an array of quantized CNR values. Quantization in tens of dB is employed only for the purpose of this comparison. A more finely quantized version of this CNR map can be inspected to predict what combinations of frequency and time delay would provide good radar sensitivity.

6. CONCLUSIONS

We have described some techniques in use at WARF which address problems particularly important in the HF radio propagation environment. These include: (1) the use and processing of the SFCW signal format for separating small moving targets from the clutter and for observing fine scale details within the clutter return itself, (2) a number of techniques to minimize interference problems in the crowded HF spectrum, and (3) a method for inferring radar sensitivity from CNR measurements.

Three major aspects of successful OTH radar operation have not been addressed in this paper. First we have omitted discussion of data processing following range-Doppler analysis although large effort has been devoted at WARF to automated processing and display methods for detecting and tracking aircraft and ships. Secondly, techniques for conversion of radar coordinates to ground coordinates has not been discussed since it is a subject worth treatment by itself. Finally multipath deghosting is important for target tracking as well as for inferring ocean surface characteristics from the clutter structure. Some success with removing multipath has been obtained with point targets. Presently the preferred approach with distributed targets such as sea clutter is to avoid multipath and to recognize and discard data with multipath contamination (Maresca, J. W., Jr., 1979).

ACKNOWLEDGMENTS

WARF was designed and constructed under the Defense Advanced Research Projects Agency sponsorship beginning in 1966; the work described here was supported by the USAF System Command, Electronic Systems Division and by the Office of Naval Research. Development of WARF has been a team effort of all members of the Remote Measurements Laboratory at SRI. For the work reported here we would particularly like to acknowledge A. C. Phillips for design of critical RF hardware, D. E. Westover for creating virtually all software used at WARF and J. E. King and B. J. Richards for administrative support.

REFERENCES

BARNUM, J. R., W. F. Marshall and A. C. Phillips, 1979, SRI Technical Report in preparation.

BARRICK, D. E., 1973, "FM/CW Radar Signals and Digital Processing," NOAA Technical Report ERL 283-WPL 26, U.S. Dept. of Commerce, Washington, D.C.

BARRICK, D. E., et al., 1974, "Sea Backscatter at HF: Interpretation and Utilization of the Echo," Proceedings of the IEEE, Vol. 62, No. 6, pp. 673-680.

CROFT, T. A., 1972, "Sky-Wave Backscatter: A Means for Observing Our Environment at Great Distances," Reviews of Geophysics and Space Physics, Vol. 10, No 1, pp. 73-155.

DAVIES, K., 1965, Ionospheric Radio Propagation, National Bureau of Standards, Washington, D.C.

FROST, O. L., III, 1972, "An Algorithm for Linearly Constrained Adaptive Array Processing," Proceedings of the IEEE, Vol. 60, pp. 926-935.

GEORGES, T. M. and J. J. Stephenson, 1968, "HF Radar Signatures of Traveling Ionospheric Irregularities, 3D Ray-Tracing Simulation," Scatter Propagation of Radio Waves, edited by E. Thrane, AGARD Conference Proceedings, Voo. 37, Part 1, Paper 55.

GRIFFITHS, L. J., 1969, "A Sample Adaptive Algorithm for Real-Time Processing in Antenna Arrays," Proceedings of the IEEE, Vol. 57, pp. 1696-1704.

GRIFFITHS, L. J., 1976, "Time-Domain Adaptive Beamforming of HF Backscatter Radar Signals," IEEE Trans. on Antennas and Propagation, Vol. AP-24, No. 5, pp. 707-720.

GRIFFITHS, L. J. and C. W. Jim, 1978, "A Generalized Sidelobe Cancelling Structure for Adaptive Arrays," Technical Report No. SPL 78-2, SRI International, Menlo Park, California.

HEADRICK, J. M. and M. I. Skolnik, 1974, "Over-the-Horizon Radar in the HF Band," Proceedings of the IEEE, Vol. 62, No. 6 pp. 664-673.

MARESCA, J. W., Jr. and C. T. Carlson, 1978, "Tracking and Monitoring Hurricanes by HF Skywave Radar over the Gulf of Mexico," Final Report, SRI International, Menlo Park, California.

MARESCA, J. W., Jr., 1979, private communication.

McKINNEY, D. and W. B. Zavoli, 1974, "Suppression of Impulsive Noise in OTH-B Radars," Technical Report 25, SRI International, Menlo Park, California.

PHILLIPS, A. C. and C. A. Cole, 1974, "An Eight-Channel Coherent Receiver to Support Studies of Digital Adaptive Beamforming at HF," Technical Report 24, SRI International, Menlo Park, California.

WASHBURN, T. W. and L. E. Sweeney, Jr., 1976, "An On-Line Adaptive Beamforming Capability for HF Backscatter Radar," IEEE Trans. on Antennas and Propagation, Vol. AP-24, No. 5, pp. 721-732.

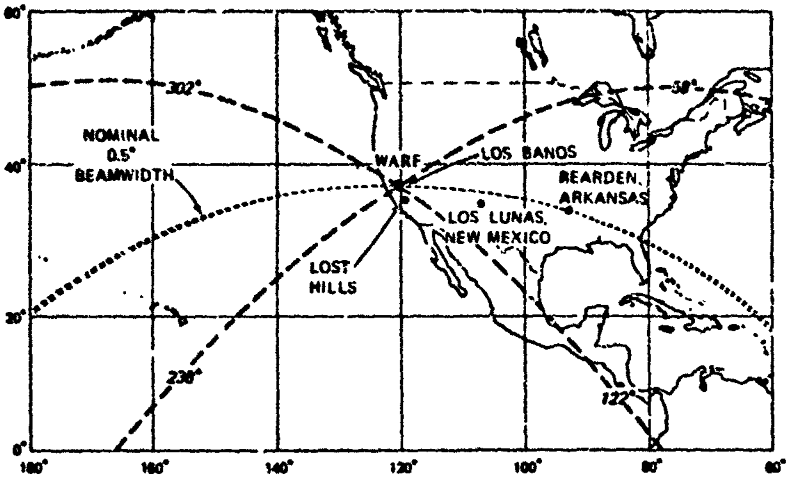

NOMINAL
0.5°
BEAMWIDTH
302°
58°
WARF
LOS BANOS
BEARDEN, ARKANSAS
LOS LUNAS, NEW MEXICO
LOST HILLS
238°
122°
60°
40°
20°
0°
180°
160°
140°
120°
100°
80°
60°

FIGURE 1 WARF COVERAGE. Receiver at Los Banos, CA; transmitter at Lost Hills, CA.

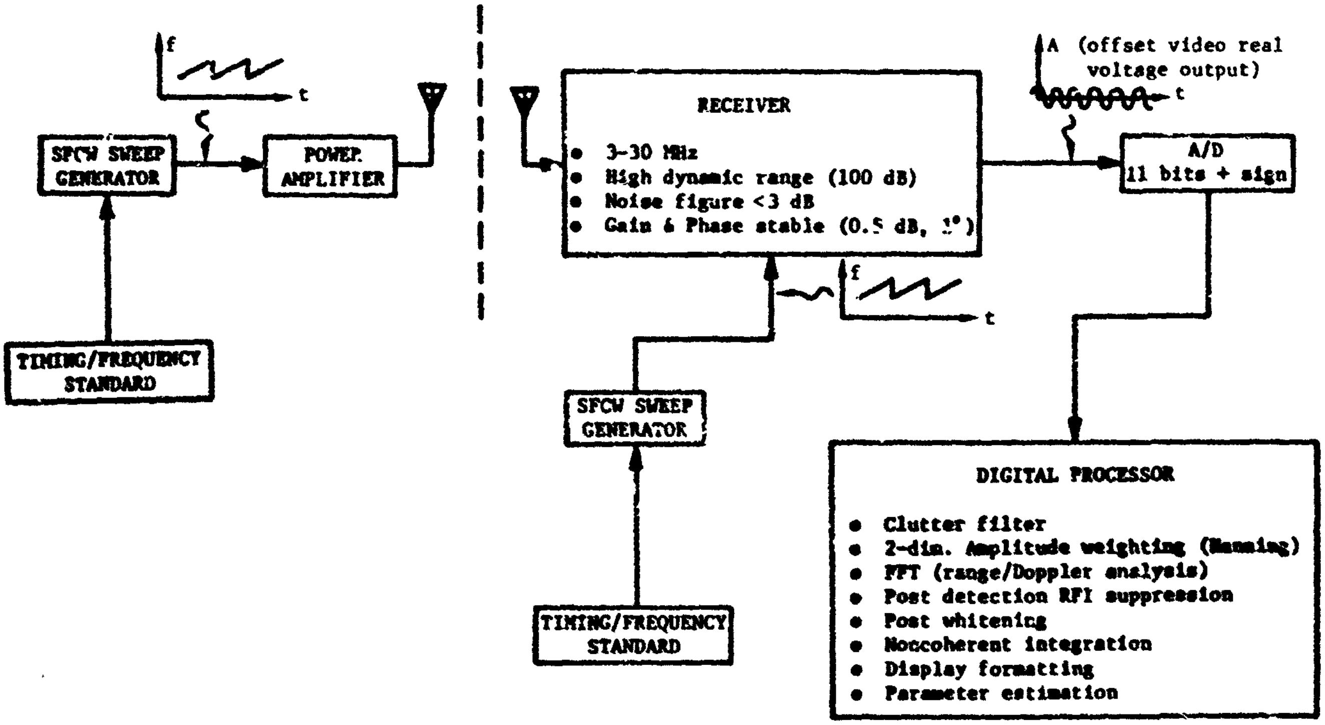
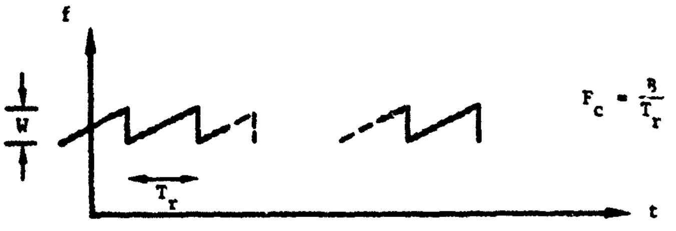

f
t
SFCW SWEEP GENERATOR
POWER AMPLIFIER
TIMING/FREQUENCY STANDARD
RECEIVER
• 3-30 MHz
• High dynamic range (100 dB)
• Noise figure <3 dB
• Gain & Phase stable (0.5 dB, 1°)
A (offset video real voltage output)
t
A/D
11 bits + sign
f
t
SFCW SWEEP GENERATOR
TIMING/FREQUENCY STANDARD
DIGITAL PROCESSOR
• Clutter filter
• 2-dim. Amplitude weighting (Hanning)
• FFT (range/Doppler analysis)
• Post detection RFI suppression
• Post whitening
• Noncoherent integration
• Display formatting
• Parameter estimation

FIGURE 2 SIMPLIFIED BLOCK DIAGRAM OF WARF WAVEFORM GENERATION AND RECEPTION.


f
W
T_r
$F_c = \frac{B}{T_r}$
t

Waveform

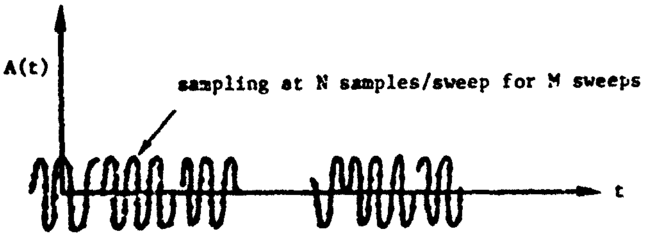

A(t)
sampling at N samples/sweep for M sweeps
t

FIGURE 3 RECEIVER OUTPUT. A/D converted at sample frequency f_s Synchronization provided to capture N samples per sweep for M sweeps.

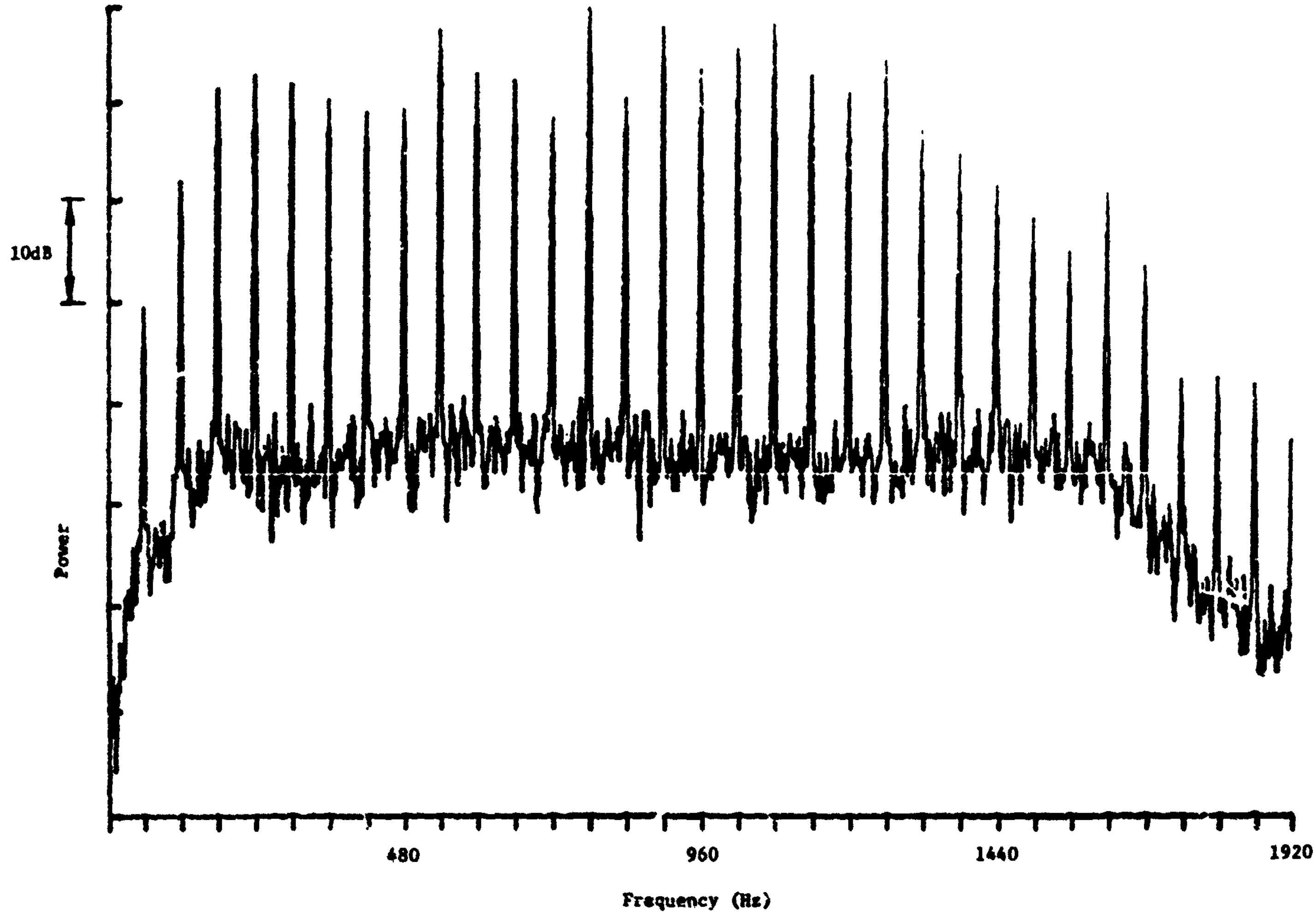


FIGURE 4 SPECTRUM COMPUTED FROM REPETITIVE, SWEEP-FREQUENCY CONTINUOUS WAVE (SFCW) SIGNAL. Waveform

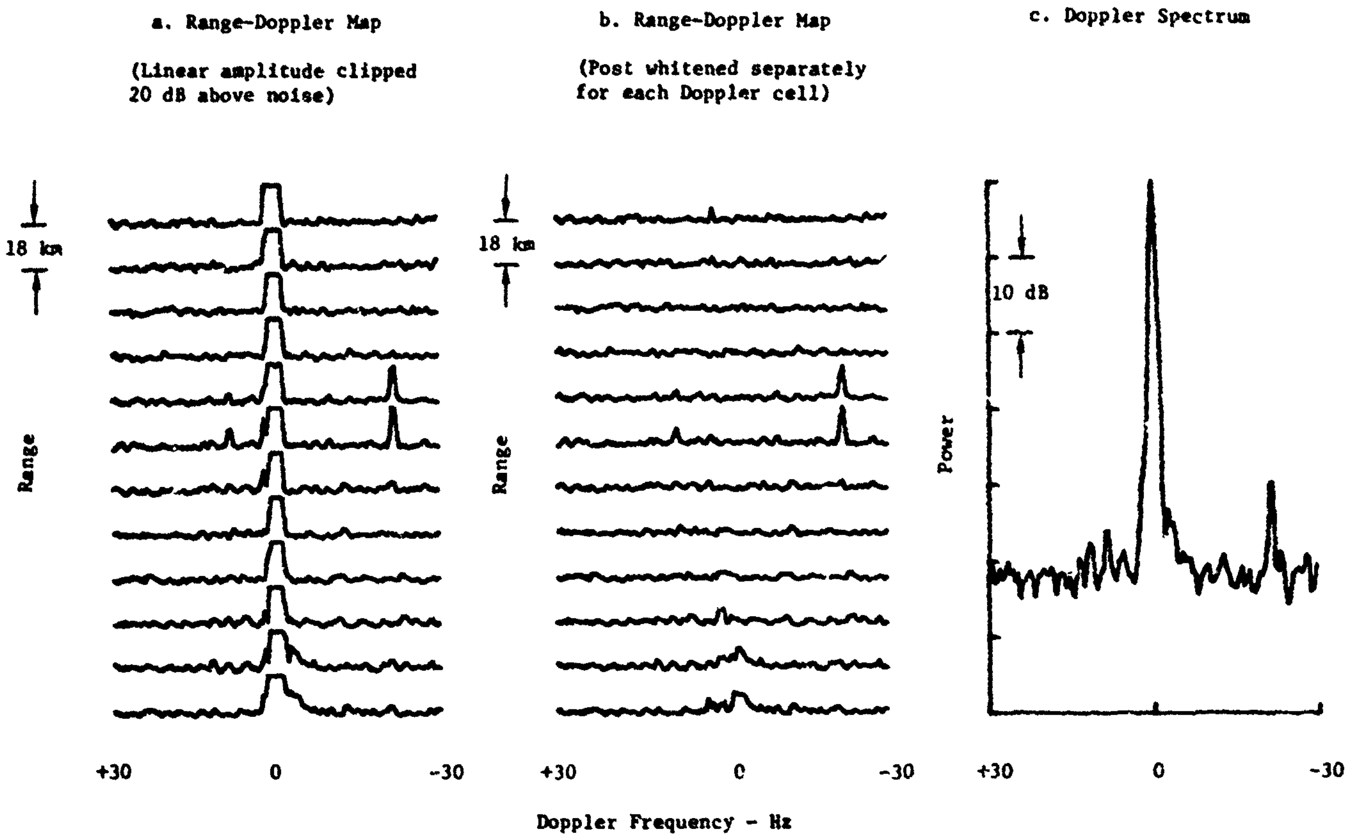


FIGURE 5 REPRESENTATIVE RANGE-DOPPLER MAPS AND DOPPLER SPECTRUM AIRCRAFT ECHO AT -25 Hz.

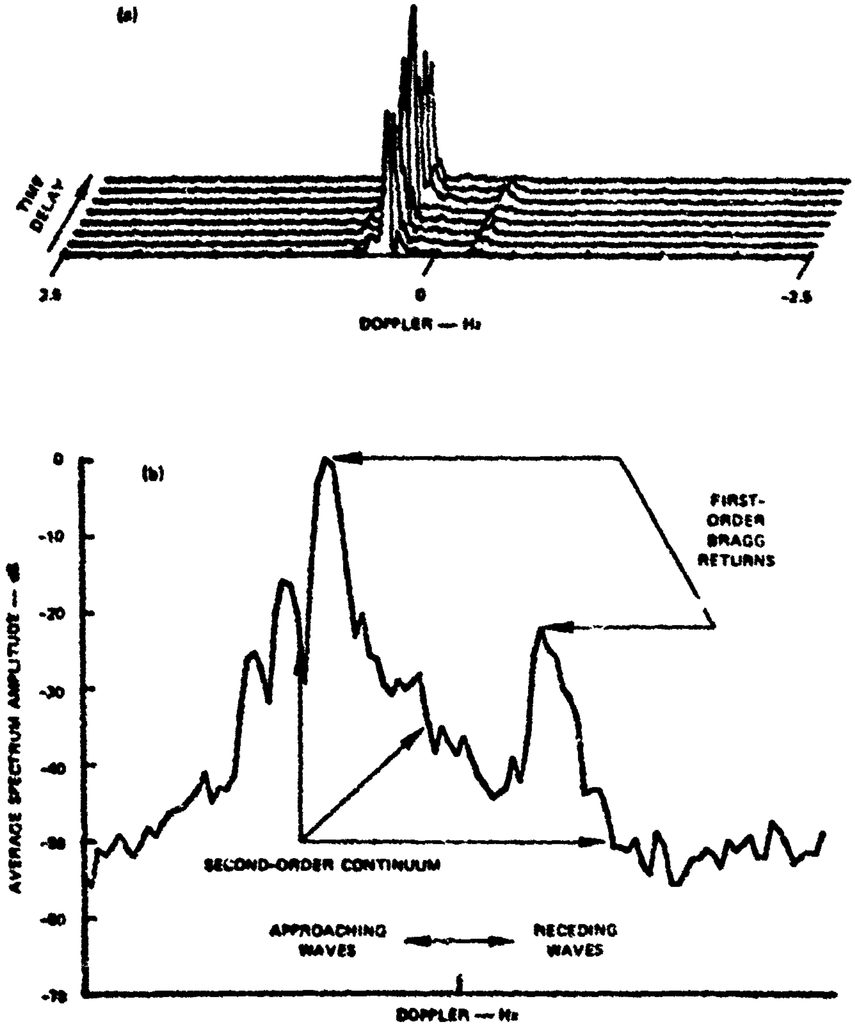
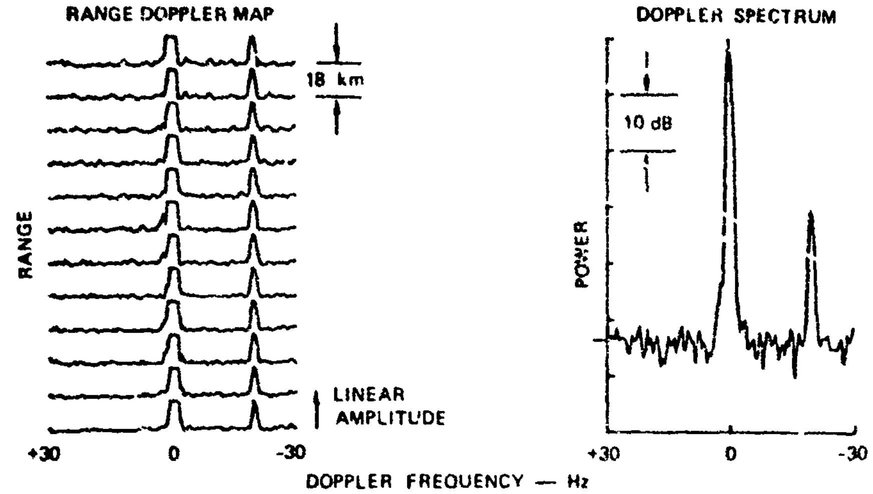
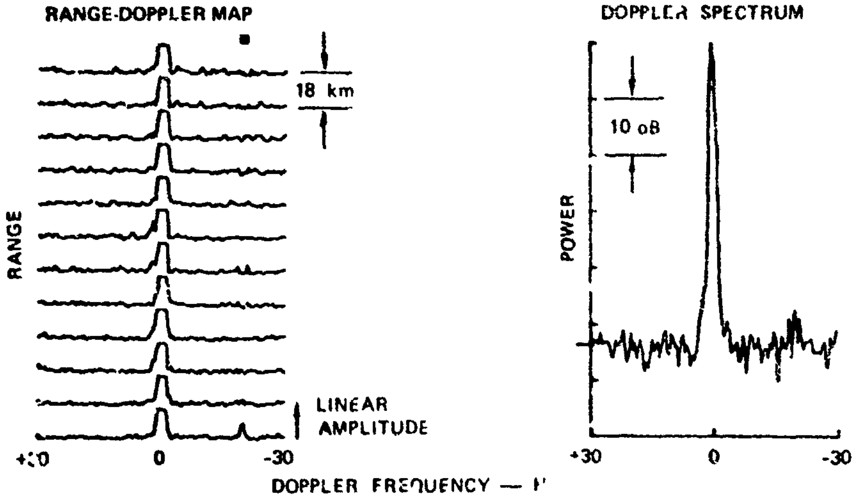


FIGURE 6 RANGE-DOPPLER-PROCESSED SEA-ECHO DOPPLER SPECTRUM. The mean Doppler spectrum is an average of Doppler spectra recorded at different range lines separated by 3 km. The first-order echoes produced by a resonant interaction between the radio waves and the ocean waves is sensitive to changes in the wind-direction field. The second-order sideband structure surrounding the stronger Bragg line is sensitive to changes in the directional ocean wave spectrum (Maresca, J. W., Jr. and C. T. Carlson, 1978).



(a) RADAR DWELL PROCESSED WITHOUT BRISA



(b) RADAR DWELL PROCESSED WITH BRISA

FIGURE 7 EXAMPLE OF INTERFERENCE SUPPRESSION USING BRISA PROCESSING.

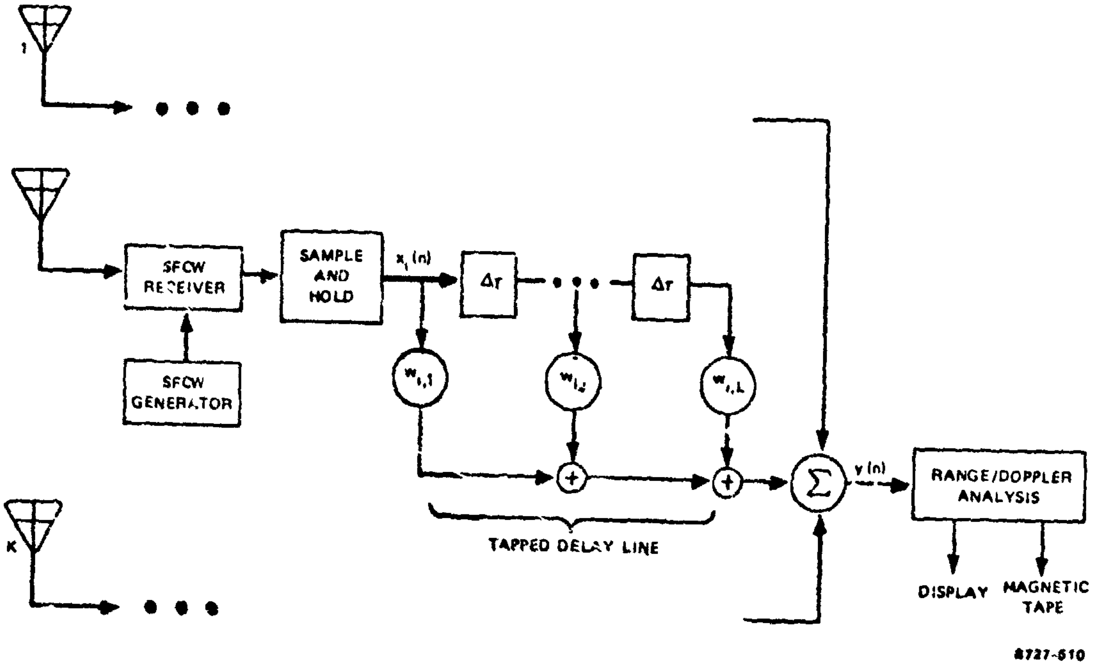

1
SFCW RECEIVER
SFCW GENERATOR
SAMPLE AND HOLD
$x_i(n)$
Δτ
Δτ
$w_{i,1}$
$w_{i,2}$
$w_{i,L}$
TAPPED DELAY LINE
K
Σ
y(n)
RANGE/DOPPLER ANALYSIS
DISPLAY
MAGNETIC TAPE
8727-610

FIGURE 8 BLOCK DIAGRAM OF DIGITAL ARRAY PROCESSOR.

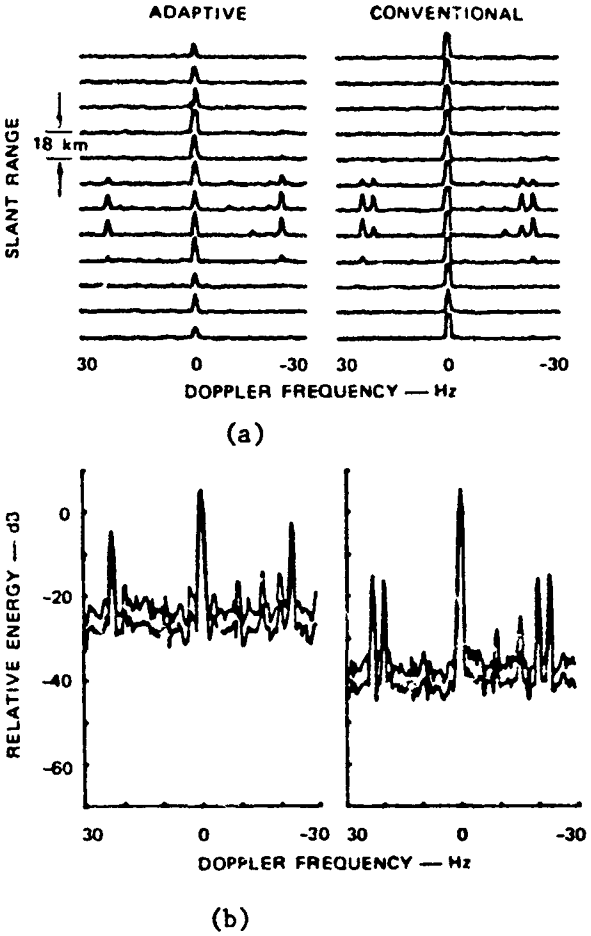


FIGURE 9 DETECTION OF IN-BEAM REPEATER (23 Hz) IN PRESENCE OF OFF-AZIMUTH REPEATER (20 Hz). Off-azimuth repeater is 1.7° away in azimuth. Dwells separated by 1 min, 8 elements, 4 taps, $\alpha = 0.1$. (a) Range/Doppler Maps. (b) Doppler Spectra.

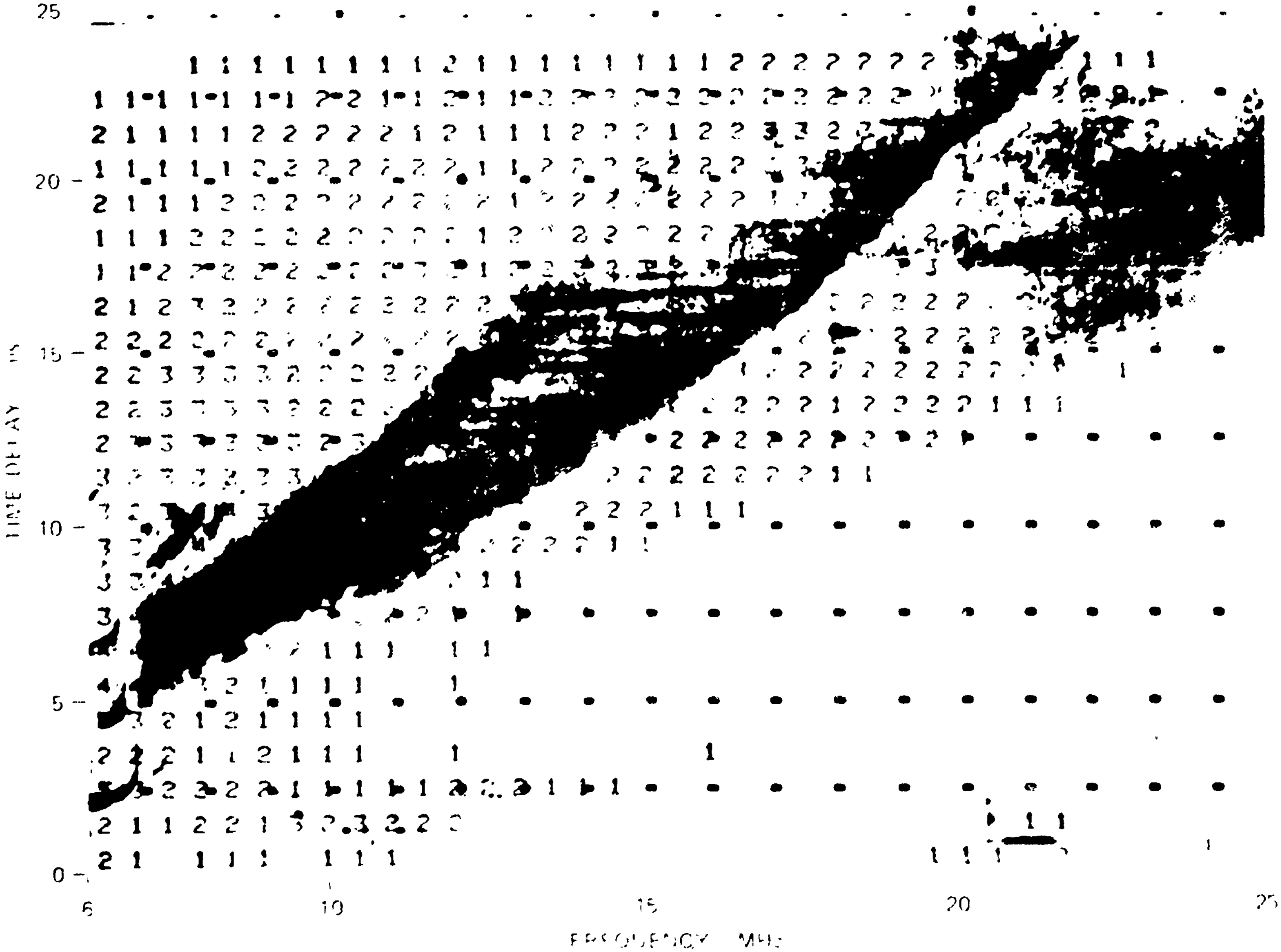


FIGURE 10 QUANTIZED CLUTTER-TO-NOISE RATIO MEASUREMENT SUPERIMPOSED ON A CONVENTIONAL BACKSCATTER IONOGRAM
Values are in tens of dB for comparison purposes only.

DISCUSSION

K. Bibl, US

Your dephasing technique used for suppression of coherent interference, also reduces the range ambiguity of long range echoes.

Author's Reply

Yes that is true. Note that the range ambiguity is the inverse of the waveform repetition frequency, low PRF's are chosen partly to insure that no energy returns from ambiguous ranges.

C.S. Goutelard, Fr

Vous utilisez, comme forme d'onde, une excursion linéaire de fréquence qui présente une fonction d'ambiguité en "lame de couteau" interdisant la mesure simultanée de la distance et de la vitesse des cibles. Vous contournez cette dificulté en utilisant des sauts discrets de fréquence de l'ordre de 60Hz mais qui introduisent alors une périodicité d'égale valeur dans l'ambiguité de fréquence.

Vous mentionnez avoir étudié d'autres types de formes d'ondes. Pouvez-vous préciser ces points, notamment les diverses formes d'onde susceptibles d'améliorer l'ambiguité dans les diverses applications que vous envisagez.

Author's Reply

Modifications to the waveform (transmitted and/or received) such as time domain weighting, blanking or phase deconvolation certainly affect the range/Doppler ambiguity function. In fact we have rejected several candidate techniques on the basis of poor sidelobes or ambiguities. For example, simple blanking of the received signal produces undesirable range sidelobes. A smooth weighting function on the blanking signal reduces these sidelobes.

N.C. Gerson, US

Have you observed first and second circumferential transit (first and second round-the-world echoes)? Do you have any statistics or estimates of path loss for both?

Author's Reply

Yes we have, but I don't have any statistics to present at this time.

HF SKYWAVE RADAR ESTIMATES OF THE TRACK, SURFACE WIND AND WAVES OF HURRICANE ANITA[1]

Joseph W. Maresca, Jr. and Christopher T. Carlson
SRI International
Menlo Park, California 94025

SUMMARY

The track of Hurricane Anita and the intensity of the surface wind and waves throughout the storm were estimated from high frequency (HF) skywave radar sea echo Doppler spectra measurements made over 3000 km away from the storm at the SRI-operated Wide Aperture Research Facility (WARF). The WARF radar measurements were made from 29 August 1977 through landfall on 2 September 1977, during both the day and the night as Anita intensified from a tropical storm to an intense hurricane moving westward through the Gulf of Mexico. Twenty-one maps of the surface wind directions within 250 km of the eye were generated from the WARF measurements. Each wind direction measurement is an estimate of the average conditions over an ocean area 15 km in range by 25 km in cross-range. A track was computed from estimates of the hurricane center made from each wind direction map. The WARF radar position estimates were compared to positions coincident in time along the smooth track produced from reconnaissance aircraft, shore-based radar and satellite cloud photographic measurements by the National Hurricane Center (NHC) and agreement was found to within ±19 km.

Hurricane Anita passed near the NOAA National Data Buoy Office (NDBO) moored buoy EB-71 (26.0°N, 93.5°W). The WARF radar estimates of the surface wind speed and direction, the significant wave height and the one-dimensional wave spectrum in all four quadrants of the storm, including the region of maximum winds were compared to available in situ measurements made at the buoy. Agreement between the WARF radar and EB-71 estimates of wind direction, wind speed and significant wave height was 7°, 0.4 m/s, and 0.5 m, respectively. Daily WARF estimates of the longshore coastal surface current on the fringe of Hurricane Anita made over a seven-day period were compared with those estimates made at a moored current meter located downcoast after correcting for a phase difference. The magnitude of the longshore current speeds at both locations were within 10 cm/s.

1. INTRODUCTION

We measured significant wave height, surface wind speed and direction and surface current speeds for the first Gulf of Mexico hurricane of the 1977 season using a high frequency (HF) skywave radar. The radar measurements were made from California by using the SRI-operated Wide Aperture Research Facility (WARF). We recorded sea backscatter for Hurricane Anita at distances more than 3000 km from the WARF, by means of single F-layer ionospheric reflection. We compiled real-time maps of the surface wind direction field within a radial distance of 200 km of the storm center, then estimated the hurricane position from these radar wind maps, and developed a track for Anita over a 4-day period between 29 August and 2 September 1977 as the storm moved westward across the Gulf of Mexico. The radar track was computed from 1[illegible] independent position estimates made before Anita crossed the Mexican coast, and was subsequently compared to the official track produced by National Hurricane Center (NHC). Agreement between the WARF position estimates and coincident temporal positions on the NHC smooth track was ±19 km. At approximately 0000 on 1 September 1977, Anita passed within 50 km of the National Data Buoy Office (NDBO) open ocean moored buoy EB-71 (26.0°N, 93.5°W), and provided us with the opportunity to compare WARF estimates of the significant wave height, and surface wind speed and direction in all four quadrants of the storm with those made at the buoy. Agreement between the WARF and EB-71 measurements was within 10%. We measured the surface currents along the western Louisiana coastline by using the WARF radar. Our measurements show good agreement with those made by moored current meters.

The purpose of this paper is to describe the capability of remotely monitoring hurricanes and other open ocean storms by using an HF skywave radar. We will describe the important aspects of the WARF skywave radar, the sea echo Doppler spectra, the method of analysis used to estimate the wind wave and surface current parameters, and the accuracy of these radar-derived quantities. Further details on Anita and other hurricanes can be found in Maresca and Carlson (1977, 1978).

2. WARF SKYWAVE RADAR

The Wide Aperture Research Facility (WARF) (SRI brochure, 1977) is a high-resolution, experimental, high-frequency skywave radar located in central California. The radar is bistatic and operates in the HF band between 5 and 30 MHz. Ocean areas are illuminated by a 20-kW swept-frequency continuous-wave (SFCW) signal from a transmitter site located at Lost Hills, California. The energy reflected from the surface beam is received 185 km to the north at Los Banos, California. The receiving antenna array is 2.5-km long and consists of a double linear array of 256 whip antennas producing a nominal 1/2° azimuthal beamwidth at 15 MHz. The signal propagates to and from remote ocean patches by means of one or more ionospheric "reflections."

The WARF coverage area is shown in Figure 1. The radar can be directed either east or west, and can be electronically steered in azimuth ±32° from boresight anywhere within the coverage area in 1/4° increments. Position accuracy is a function of midpath ionospheric height estimates whose uncertainty in the midpath height results in a nominal position accuracy of approximately 20 km. At any one location, the accuracy between consecutive measurements in range and azimuth is an order of magnitude better. WARF has multiple-beam capability, and sea backscatter is usually received simultaneously at four adjacent ocean areas from four different beams separated by 1/4°. The size of the ocean scattering patch is a function of the beam-width, the range, the range cell separation, and the number of range cells averaged together. The size of the minimum scattering patch at a range of approximately 2000 km is 3 km in range by 15 km in azimuth.

---

[1] A portion of this paper was presented at the Sixteenth Conference on Coastal Engineering held in Hamburg, Germany (August 1978) and appeared in the proceedings of the conference.

3. SEA ECHO DOPPLER SPECTRUM

The sea backscatter received at the WARF is coherently processed in range and Doppler to produce a sea-echo Doppler spectrum. We usually process 21 independent Doppler spectra spaced at 3-km range intervals. These spectra are obtained simultaneously at each of four adjacent radar beams. A total of 84 independent Doppler spectra are obtained for each coherent time period. We compute an average spectrum from a subset of these Doppler spectra, depending on the type of measurement and the time and space scales associated with the ocean surface features. An example of a mean sea-echo Doppler spectrum produced by averaging 112 spectra obtained from four consecutive 102.4 s coherent time periods, over a scattering patch consisting of 21 range cells and 3 adjacent beams, is shown in Figure 2.

The sea-echo Doppler spectrum shown in Figure 2 is characterized by two dominant first-order echoes surrounded by a second-order continuum. Crombie (1955) interpreted the first-order echoes in terms of simple Bragg scattering that represented a resonant response between radio waves of wave number k_o and ocean waves of wave number $k = 2k_o$. The radar measures the relative power and Doppler of the ocean waves traveling radially toward or away from the radar. The power ratio of the two first-order echoes are indicative of the wave direction of the waves of wave number k. Because k is usually large ($k > 0.5$), it is assumed that the wind direction is identical to the direction of these waves. An shift in Doppler of the first-order echoes from their theoretical positions when the ionospheric Doppler is zero is indicative of the magnitude of the surface current.

The wave-height spectrum is derived from the second-order structure surrounding the first-order echoes. For hurricanes, the power in the second-order echoes is large. As the total wave energy increases, the amplitude of the second-order echoes increases as illustrated in Figure 3. Barrick (1972 a,b) derived theoretical expressions that accurately model the HF scattering process to second order. For a specific directional wave spectrum, the model computes the Doppler spectrum. The effects of the wind direction, wave directionality and the wave frequency spectrum on the modeled Doppler spectrum have been extensively studied through the use of this model.

4. HURRICANE DATA SAMPLING

Data sampling during a hurricane is divided into two tasks to optimize the sampling time and the data quality. The spectral resolution, directly related to the coherent integration time, can be much coarser for first-order measurements than for second-order measurements. Wind direction estimates are computed from the first-order echoes, and can be computed considerably more rapidly than wave height and wind speed estimates, which are computed from the second-order echoes. Usually, the longer the coherent integration time, the greater the influence the ionosphere has on the quality of the data.

The quality of the recorded sea backscatter depends on the ionospheric conditions over short periods--on the order of minutes. High-quality sea backscatter is obtained if the radio waves propagate by means of a strong, single, stable, coherent ionospheric layer. Sometimes the signals may be received at the same time from two or more different paths (multipath). In this case, the second or succeeding signals will be reflected from different parts of the ocean and different parts of the ionosphere, and will contaminate the sea echo received from the first path. If the ionosphere is changing in time or space during the coherent radar dwell (time period), further degradation of the data will occur. The ability to predict the ionospheric conditions would enable the radar operator to minimize the contaminating effects of the ionosphere, improve the quality of sea backscatter, and reduce the sampling time. The ionospheric soundings provide some data quality information. The vertical and oblique incidence soundings are taken every 10 minutes; a complete sounding requires approximately 3 minutes. The coherent radar measurements made at WARF require between 10 and 100 seconds to complete. Because the time required to complete a sounding is greater than the time required to record the sea backscatter data, assessment of the data quality is difficult for rapidly changing ionospheres. Therefore, real-time output of the data from the WARF site minicomputer is used to verify data quality.

The wind direction measurement is not extremely sensitive to ionospheric contamination because only the amplitude of the two strong first-order echoes must be measured. A coherent integration time of 12.8 seconds (0.078-Hz resolution) is sufficient to resolve the peaks of the first-order echoes. We can map the wind-direction field in a hurricane by scanning in range and azimuth. It is possible to routinely map the surface-wind-direction field of a hurricane in about 10 minutes. Once the surface-wind-direction map is made, the storm center can be identified for tracking purposes, and regions of interest can be selected for more extensive monitoring of wind speed and wave height anywhere within the storm.

The measurement of surface currents by HF radar is slightly more sensitive to contamination of the Doppler spectrum by ionospheric multipath or smearing than the wind direction measurements. The surface current measurements are based on the Doppler shift of the first-order echoes relative to an echo of known Doppler. To accurately resolve the Dopplers of these echoes, coherent integrations times of 102.4 seconds (0.01 Hz resolution) are used to process the data. The accuracy of the surface current estimate is a function of the accuracy of determining the peak of the first-order echoes.

The significant wave height and wind speed measurements are sensitive to ionospheric contamination because these quantities are estimated from the second-order echoes surrounding the stronger first-order echoes. This contamination is the largest source of error in these measurements. A coherent integration time of 102.4 seconds is required to resolve the second-order echoes. The ionosphere does not generally support coherent integration time periods of this length. Multipath and ionospheric smearing can seriously degrade the weaker second-order echoes. Because of this contamination, we are not able to routinely estimate wave height for each 102.4-second time period as we are able to calculate the surface current for each 102.4-second period or wind direction for each 12.8-second time period. A sampling strategy that combines careful propagation management through selection of frequencies, which result in a stable, coherent, single propagation path, and signal processing that minimizes the contaminating effects of the ionosphere are used to obtain a data set suitable for analysis. Recent work by SRI and NOAA (Georges, T. M. and Maresca, J. W., Jr.) as resulted in improved methods of collecting high-quality data by sorting the data according to a spectral sharpness index. The effect of ionospheric contamination, however, is less severe for data recorded during large waves generated during a hurricane. The amplitude of the second order echoes containing the wave height information may be stronger than the contamination effects, and thus, wave height can be calculated despite the contamination.

For the Anita wind direction measurements, we divided the data into 16 groups and analyzed three consecutive 12.8 second coherent radar dwells. Each wind direction estimate was calculated from an average of 15 Doppler spectra. At a range of 3000 km, the size of each scattering patch was [illegible] km x [illegible] km. It would be desirable to compute wave height and wind speed from a similar data set, but this is not presently possible. Longer coherent time periods and more independent samples of the spectra are required to obtain a high quality sample. We could collect the data over a small scattering patch by averaging over a long time, or we could increase the scattering patch size and average in space. Averaging in space is preferable because it reduces the total time required to obtain a mean Doppler spectrum. For the Anita wave height and wind speed measurements, we analyzed the data from three of the adjacent azimuth cells and [illegible] continuous range cells. The total scattering patch was [illegible] km x [illegible] km. Several consecutive integration periods are required to record the data.

5. WIND DIRECTION

HF skywave radar has been used to map the surface wind fields associated with large weather systems (Barnum, J. R., et al., 1977) and tropical storms (Maresca, J. W., Jr. and Barnum, J. R., [illegible]). The radar measured surface-wind directions are derived from the predominant direction of ocean gravity waves, approximately 10-m long. The waves satisfying the first order Bragg scattering condition, [illegible], are assumed to be tightly coupled to the wind for time scales on the order of tens of minutes. This assumption is reasonable for the high wind speed conditions associated with hurricanes. Available directional wave spectra measurements (Longuet-Higgins, M. S., et al., 196[illegible]; Mitsuyasu, H., et al., 19[illegible]; [illegible], 1969) indicate that the dominant wave direction is representative of the predominant wind direction. For open ocean conditions, agreement between the WARF radar and shipboard or other measurements of wind direction is ±10°. (Stewart, R. H. and Barnum, J. R., 1975) For hurricane winds, the agreement between coincident wind direction measurements made by the NOAA National Data Buoy Office (NDBO) data buoys and the WARF radar is better than 10° (Maresca, J. W., Jr. and Carlson, C. T., 197[illegible], 19[illegible]; Maresca, J. W., Jr. and Barnum, J. R.)

The radar measures the relative power between the approaching and receding waves that satisfy the Bragg scattering condition. If a cosine directional distribution (Longuet-Higgins, M. S., et al., 1963)

$$G(\theta) = \cos^{s} \frac{\theta}{2} \qquad (1)$$

is assumed, then the relative power of the approaching and receding waves measured by the radar is sufficient to estimate θ with an ambiguity about the beam direction. This left-right ambiguity is resolved by the predictable cyclonic surface circulation within the hurricane. The shape of G(θ) is controlled by the spreading parameter, s, where θ is the angle between the radar beam and the wind direction. For open ocean conditions, we have estimated s from several models (Mitsuyasu, H., et al., 19[illegible]; Stewart, R. H. and Barnum, J. R., 1975). For the extreme hurricane winds, the values of s estimated from these models are too low. Based upon previous hurricane analysis and comparisons of wind direction at NDBO data buoys, we used values of s between 1.0 and 2.0. No attempts were made to account for variations in s as a function of location within a hurricane.

6. SURFACE CURRENT

The speed of the surface currents in the upper one meter of ocean detected remotely by HF radar can be estimated from the phase speed, or Doppler, of the ocean waves seen in the first order sea echoes (Barrick, D. E., et al., 197[illegible], 197[illegible]; Stewart, R. H. and Joy, J. W., [illegible]; Maresca, J. W., Jr. and Georges, T. M., [illegible]). The measured phase velocity of these ocean waves can be different from the theoretical phase velocity predicted by first order water wave theory. This difference is caused by the advection of the waves by a current. Stewart and Joy (1974) showed that the surface current [illegible]

$$V_{CR} = [illegible] \qquad (2)$$

where λ is the ocean wave length and Δf is the shift in Doppler of the first order echoes. For a radar frequency of 15 MHz, the Doppler shift of [illegible] Hz, the magnitude of the surface current directed radially along the radar axis is [illegible] m/s at a depth of [illegible] m. Unlike the skywave radar wind direction measurement, the effects of the ionospheric motion must be known to make a skywave measurement of current. The entire Doppler spectrum can be shifted by ionospheric motion, as has been shown (Maresca, J. W., Jr. and Carlson, C. T., 1977; Maresca, J. W., Jr., et al., 1976; Maresca, J. W., Jr. [illegible]). The [illegible] echoes received from an HF repeater, land or oil platforms, or other discrete scatterers along the coast of Mexico are sufficient to remove the effects of the ionospheric motion from the data. After removal of the ionospheric effect from the Doppler spectrum, we assume that any deviation from the theoretically predicted Doppler shift of the first-order peaks is due to the surface current.

The radar measured component of the surface current is directed along the radar axis. With only one radar measurement, the absolute component of the current cannot be derived. A second independent measurement by a radar with overlapping coverage would be required for an absolute direction measurement. However, the radar axis is approximately parallel with the Gulf of Mexico coastline in the region of measurement reported here and provides a reasonable estimate of longshore surface current velocity.

7. SIGNIFICANT WAVE HEIGHT

Barrick derived an integral expression that predicts the Doppler spectrum for a specified directional wave spectrum input. Recent efforts have succeeded in inverting this integral expression to compute the total rms wave height (Barrick, D. E., 19[illegible]; Maresca, J. W., Jr. and Georges, T. M., [illegible]), one dimensional wave frequency spectrum (Barrick, D. E., 197[illegible]; Lipa, B., 197[illegible] a,b; Barrick, D. E. and Lipa, B. J., 19[illegible]) and the directional distribution (Lipa, B. J., 197[illegible]; Barrick, D. E. and Lipa, B. J., 19[illegible]). Barrick's

(1977a, c) expressions have been used to analyze skywave radar data recorded for a Pacific Ocean storm (Maresca, J. W., Jr. and Georges, T. M.) and tropical storms (Maresca, J. W., Jr. and Carlson, C. T., 1977, 1978; Maresca, J. W., Jr., 1978)

We used a power law derived from simulated data by Maresca and Georges to compute rms wave height by relating the ratio of the total second-order and first-order power to the rms wave height:

$$k_0 h = a R_2^b \tag{3}$$

where $0.2 < k_0 h < 1.0$, h is the rms wave height; k_0 is the radar wave number; R_2 is the ratio of the total second- to total first-order power; and a = 0.8 and b = 0.6 are constants. This average expression was derived from theoretical simulations of the Doppler spectra for different radar-to-wind directions, directional distributions, functional forms of the wave-frequency spectrum, and operating radar frequencies. Equation (3) is accurate to within 10%. Discussion of the errors can be found in Maresca and Carlson (1977, 1978) and Maresca and Georges.

8. WIND SPEED

Historically, wave models have been developed to predict wave height and the wave spectrum from an input wind field. The accuracy of these models is dependent upon the accuracy of the input winds. Hasselmann et al. (1976) proposed a one-dimensional parametric wind-wave model for fetch limited growing wind-sea conditions. Ross and Cardone (Ross D. B., 1976; Cardone, V. J. and Ross, D. B., 1978; Ross, D. B. and Cardone, V. J., 1978; Cardone, V. J., et al., 1977) empirically derived a power-law expression for hurricanes based on the form proposed by Hasselmann et al. (1976) that relates the nondimensional wave energy, $\tilde{E}$, by using wind, wave, and fetch measured during hurricanes Ava, Camille, and Eloise. For hurricanes,

$$\tilde{E} = 2.5 \times 10^{-5} \tilde{R}^{0.45} \tag{4}$$

where $\tilde{E} = Eg^2/W^4$; $\tilde{R} = rg/W^2$; $E = h^2$; and $H_s = 4h$. In $\tilde{E}$ and $\tilde{R}$, E is the total wave energy, h is the rms wave height: H_s is the significant wave height; r is the radial distance from the eye to the measurement point that accounts for fetch; g is the gravitational acceleration; and W is the wind speed. Solving for wind speed in Eq. (4), we obtain

$$W = \left(\frac{h^2 g^2}{2.5 \times 10^{-5} \, (rg)^{0.45}} \right)^{0.323} . \tag{5}$$

The wind-wave model used to derive Eq. (5) is applicable for slow moving storms in which $W \geq 15$ m/s and $\tilde{R} \leq 3 \times 10^4$. For the unusual case where the storms move very fast or very slow, Ross and Cardone (1978) showed that significant differences in the modeled and measured wave heights occur.

We used Eq. (5) to calculate wind speed for Anita and compared our results with the wind speeds measured at NDBO buoys and by reconnaissance aircraft. The radial fetch (r) was measured from the WARF-derived wind maps, and the wave height (h) was computed using Eq. (3). The radar-derived W is not an instantaneous wind speed estimate; it is a temporal and spatial average of the winds. Our radar-derived W was compared to the 15-minute wind speed averages made at NDBO moored data buoys.

9. MEASUREMENTS AND RESULTS

Hurricane Anita formed as a tropical depression in the Gulf of Mexico at about 1200Z on 29 August 1977. Anita developed into a tropical storm at approximately 0600Z on 30 August 1977, and about 12 hours later intensified into the first Gulf of Mexico hurricane of the 1977 season. As Anita moved west across the Gulf, winds in excess of 75 m/s were recorded. Five days of skywave data beginning 29 August 1977, were recorded prior to Anita's landfall, [illegible] September 1977, approximately 248 km south of Brownsville, Texas. Twenty-one radar wind maps were compiled at WARF. The first 4 wind maps were not used in the radar-derived track presented here because the radar showed two distinct centers during this early period. On 30 August 1977, the storm intensified and developed one center. The wind maps were updated 3 to 5 times per day during both daytime and nighttime periods and were used to develop the WARF-derived track. Figure 4 shows the radar-derived positions in relation to the official NHC smooth track produced from reconnaissance aircraft measurements, visible and infrared satellite cloud photographs, and shore-based microwave Doppler radar. The relative agreement between the WARF position estimates and the interpolated temporal position estimates along the smooth track is ±19 km.

There are two potential sources of error associated with the WARF hurricane position fixes: the absolute position error of the radar consisting of range and azimuth errors, and the errors associated with locating the storm center from the radar wind direction measurements. We estimate the range errors of the radar caused by errors in determining the ionospheric height at midpath to be 20 km. If a coastal scan is included as part of collecting the wind map data, the land echo can be used as a reference to more accurately determine the ionospheric height, and therefore, reduce this error. We estimate the error in azimuth caused by ionospheric tilting to be 20 km. These range and azimuth errors can be reduced significantly by installing an HF repeater along the coast which receives signals and transmits them back with a known frequency shift. When we assume similar mean ionospheric conditions within 200 km of the storm center, the entire wind map can be translated in azimuth and range to correct for the absolute position error. The location of the wind direction measurement with respect to the storm center is generally not affected by these position errors. The error associated with determining the storm center from the radar maps is about 20 km. The error is caused by the left/right ambiguity in the wind direction measurement. The average maximum error from these two potential sources of error is about 40 km. In comparing the WARF position fixes to the NHC track we found relative differences of between 5 and 50 km, and these relative differences can be attributed to the sources of error just discussed.

Anita passed 50 km south of NDBO buoy EB-71 at about 0000Z on 1 September 1977. Two WARF-derived wind maps were made at 2140Z on 30 August 1977 and 0120Z on 1 September 1977, which brackets this time period. One of these wind maps is shown in Figure 5. Also shown on Figure 5 is the surface wind direction field derived from data recorded by NDBO buoy EB-71. These buoy-measured wind directions were recorded at 2-hour intervals during the period ±18 hours of Anita's passing EB-71. The buoy-derived wind field was computed by a time space conversion that assumed uniform wind direction and lateral storm motion during this period. We compared the buoy-derived wind directions to the WARF-derived wind directions; agreement was within 19°. Agreement between the WARF-derived wind direction estimate coincident in time and space with the buoy wind direction estimate was 1°.

Between 2314Z on 31 August 1977 and 0020Z on 1 September 1977, WARF measurements were made at five locations surrounding the center of the storm. The location of each measurement relative to the storm center was interpolated from the two wind maps. We computed the wind direction by using Eq. (1), wind speed by using Eq. (5), and wave height by using Eq. (3) at each location (see Table 1), and compared these measurements to a buoy-derived wind and wave field. The maps of the spatial distribution of the wind direction, wind speed, and wave height were compiled from NDBO EB-71 data buoy measurements. Each parameter was plotted in relation to the storm center; they are shown in Figures 6, 7, and 8. We assumed that Anita moved uniformly with no change in the meteorological conditions during the period 18 hours before and 18 hours after passing the buoy.

Table 1

WARF ESTIMATES OF SIGNIFICANT WAVE HEIGHT, WIND SPEED, AND WIND DIRECTION

Point	Lat (°N)	Long (°W)	Time (GMT)	r* (km)	N*	H_s* (m)	W* (m/s)	ϕ* (°N)
A	25.7	92.9	2314	35	80	5.8	26.7	277.5
A	25.7	92.9	2343	35	112	5.2	22.8	-
B	26.3	92.1	2324	75	80	6.0	24.4	95.1
C	26.3	93.1	2358	65	35	5.8	24.4	70.2
D	25.7	92.1	0003	65	134	5.1	22.5	168.8
E	25.2	91.1	0020	180	49	4.6	18.1	137.2

*r = Radial Distance
N = Number of Spectra Averaged
H_s = Significant Wave Height
W = Wind Speed
ϕ = Wind Direction

The significant wave height shown in Figure 6 was measured at the buoy every 3 hours; the wind direction and wind speed shown in Figures 7 and 8 were measured at the buoy every 2 hours. During this 36-hour period, Anita began to intensify, and under our assumption of uniform lateral storm motion the validity of the buoy-derived wind and wave fields are suspect. Exact comparison of the EB-71 and WARF measurements are difficult because of the differences in the time, location, and area of ocean monitored. On Figure 6 we also included the wave forecast for significant wave height computed by Cardone et al. (1977) for comparison.

WARF wind and wave height estimates were too far away from the buoy-derived quantities for direct comparison, but the agreement between the WARF- and buoy-derived wind and wave fields was reasonable. The wave forecast was compared to both the buoy- and WARF-derived wave heights. We found good agreement between the forecast, buoy, and WARF estimates of wave height along the 4.6-m contour east of the hurricane. The buoy estimate west of the hurricane along the 4.6-m contour line suggests the radius of this contour line was too large. WARF estimates of wave height at Points B and C, located at the extremes of the forecast region of highest waves, were lower than the forecast. The good agreement of Point E with the nearby buoy estimate, and the consistency of the WARF estimates at Point B and C suggest that the 7.6-m contour should have been smaller. The WARF estimates of wave height at Points A and D are located between the 6.1-m contours. Again, the buoy- and WARF-derived wave heights suggested that the 6.1-m contour line was too large. We should also note that any contamination of the sea-echo Doppler spectra by the ionosphere would result in radar wave-height estimates that would be too high. The composite of wave height data obtained from the forecast and EB-71 buoy indicated the validity of the WARF wave height estimates.

The agreement between the WARF-derived estimates of wind speed and the buoy-derived estimates of wind speed is good. There are three principal errors associated with the WARF wind speed estimates: error in estimating the radial fetch, error in estimating the rms wave height, and error in the parametric model. We computed the error in calculating wind speed for a ±0.5-m error in estimating wave height for significant wave height of 5.5 m (9.1% error) and for radial fetches of 30, 50, 70, and 100 km. The errors were less than 1.6 m/s. We also computed the error in calculating wind speed for a ±20-km error in estimating the radial fetch for a significant wave height of 5.5 m and radial fetches of 30, 50, 70 and 100 km. For radial fetches greater than 30 km, a +20-km error causes an error of less than 2 m/s in wind speed. For radial fetches greater than 50 km, a -20-km error causes an error of less than 2 m/s in wind speed. This represents less than an 8% error. These errors are typical of the WARF estimates of the significant wave height and radial fetch measurements. The errors associated with the model are discussed by Ross and Cardone. (1978) For Anita, the mean and rms differences between the Cardone et al. (1977) parametric model forecast and measured wave heights at EB-71 is 0.21 ±0.83 m. This includes errors in measuring wave height at the buoy and in radial fetch from the conventional position fixes. We also calculated the wind speed using Eq. (5) for some of the buoy-measured wave heights shown in Figure 6, and compared the calculated wind speed measurements to wind speeds measured [illegible] the buoy (Table 2). The data is indicative of the accuracy we could expect from the WARF estimates of wind speed using Eq. (5). For these data, we believe the largest sources of error in the comparison were the uncertainty in the radial distance to each point caused by compiling the map over a 36-hour period, and the assumption of a symmetrical distribution of the winds.

Table 2

COMPARISON OF WIND SPEED VALUES CALCULATED FROM EQ. (5) DERIVED FROM EB 71 SIGNIFICANT WAVE HEIGHT MEASUREMENTS

Lat (°N)	Long (°W)	H_s^* (m)	r^* (km)	W_c^* (m/s)	W_m^* (m/s)	Difference (m/s)
25.5	94.8	2.5	217	11.9	7.6	+4.3
25.7	94.4	2.9	174	13.5	9.0	+4.5
25.7	93.9	3.1	124	12.9	11.3	+1.6
25.8	93.4	5.5	69	23.4	17.4	+6.0
26.0	92.5	6.5	24	30.4	34.1	-3.7
26.3	91.5	4.7	126	19.4	23.1	-3.7
26.6	90.7	4.6	212	17.7	18.1	-0.4

$^*H_s$ = Significant Wave height
r = Radial Distance
W_c = Computed Wind Speed
W_m = Measured Wind Speed

We computed wave height and wind speed estimates several times daily over the life of Anita. Figure 9 and Table 3 show daily wave height and wind speed estimates made at several locations within the storm from 30 August through 2 September. Anita intensified over this period as reflected in the increasing wave heights. Also shown on Figure 9 is the nearest in situ observation of the wave height measured by EB-71 at 0000Z on 1 September. Using Eq. (5) we estimated wind speeds from the WARF radar measurements of wave height and radial fetch. The nearest in situ observation of the surface wind speed was also made at EB-71 at 0000Z on 1 September 1977 (point B on Figure 8). No other surface observations of wind speed were available for comparison to the radar data. However, NOAA reconnaissance aircraft 5-second averaged wind speed estimates made at an altitude of 440 m were available for this time period (Sheets, R. C., 1978). Direct comparison of the aircraft wind speeds and the WARF radar wind speeds is difficult because of different altitudes, averaging times, and locations of the measurements. We reduced the aircraft flight level wind speed to a surface level wind speed for comparison to the WARF radar estimates. The aircraft measurements were made along N-S and E-W axes directed through the storm center. The aircraft measurements nearest the WARF radar measurements were used to estimate the surface wind speed. We computed the surface wind speed from the aircraft measurements using a ratio relating the upper level gradient wind to the surface level wind. Using the two layer Cardone (1969) marine boundary layer model, Elsberry et al. (1974) computed the ratio of the wind at the top of the upper layer to the wind at the top of the surface layer for different regions of the hurricane, different surface roughness and different ratios of heat conductivity to eddy viscosity. This ratio ranges from about 0.5 to 0.85. The lower values represent regions near the peak winds. We assumed the 440-m aircraft wind was representative of the wind at the top of the upper layer and reduced it to the surface level using a ratio of 0.7.

The agreement between the WARF radar and aircraft winds corrected to the surface is good. This comparison is not intended as a test of the radar technique. The purpose of this comparison was to demonstrate that the WARF radar estimates of wind speed are quite reasonable and are considerably different than the upper level wind speed estimates.

Table 3

COMPARISON OF WIND SPEEDS COMPUTED FROM EQ. (5) AND MEASURED BY NOAA RECONNAISSANCE AIRCRAFT

	WARF RADAR							NOAA AIRCRAFT			
Point	Lat (°N)	Long (°W)	Date	Time (GMT)	$^*r$ (km)	H_s (m)	Surface W (m/s)	Date	Time (GMT)	440-m W (m/s)	Surface W (m/s)
1	27.7	92.3	8-30	205[illegible]	188	5.0	19.0	8-30	13[illegible]6	23.0	16.1
2	26.3	92.1	8-31	2324	72	6.0	24.6	8-31	143[illegible]	34.0	23.8
3	25.7	96.8	9-1	2358	118	7.2	25.7	9-2	0622	32.0	22.4
4	24.4	95.9	9-2	0530	62	7.9	30.0	9-2	0622	41.6	29.1**

---

* r = radial distance from WARF measured center to location of wave height measurement

** average of North and South aircraft wind speeds

Surface current measurements were made at the WARF radar during the periods 30 August through 2 September 1977 and 4-5 September 1977. The measurements were made along the western portion of the Louisiana coastline as shown by the shaded region in Figure 10a. Two days after Anita crossed land, tropical storm Babe, a weaker, short-lived storm developed. The tracks of Anita and Babe are also shown. We used the land echo appearing in the Doppler spectra to remove ionospheric motion effects from our data. Smith (197[illegible]) measured the current speed and direction during this same period at a location approximately 21.5 km off the Port O'Connor, Texas coast. Two recording current meters were deployed at two and ten meters above the bottom in approximately 17 meters of water. The longshore component of current measured by Smith on a 2-hour interval at this location is shown by the solid line in Figure 10b. Positive values indicate motion toward 62°. The measurements of surface current made by the WARF radar are shown by dots in Figure 10 (b).

Before Anita developed, the current-meter record shows currents of 10 to 20 cm/s moving toward 62°. As the peripheral winds of Anita impacted the coastal regions, the currents reversed direction in response to the winds and steadily increased to a maximum value of approximately 80 cm/s directed toward 242°. After the storm made landfall on 2 September, the magnitude of the current decreased until late on 4 September. At this time, another increase in the current magnitude was observed. Finally, the current reversed direction back to 62° on 6 September. The two current maxima found on 1 September and 5 September were caused by hurricane Anita and Babe. The WARF radar estimates are in good agreement with the current-meter observations. Again, direct quantitative comparison is not possible because of the large separation in distance between the two measurements. The WARF radar measurements were made approximately 290 km up coast from the current-meter records. Anita tracked approximately parallel to the coastline. The perpendicular distance between the two measurement points and the track are approximately equal, and we therefore observe, as expected, similar magnitudes but a different phase of the longshore current at each measurement point. The WARF radar estimates of the current precede those measured by the meter. In addition, the currents generated by Babe are greater at the location of the WARF radar measurements than at the location of the current meter.

12. SUMMARY

Spatially averaged hurricane wind-speed, wind-direction, and wave-height estimates made at the WARF for Anita were compared to point measurements made at NDBO buoys and by reconnaissance aircraft. Agreement was within the nominal measurement accuracy of all the sensors. Surface current measurements made by other sensors coincident with the radar measurements are rare. Comparison of point current measurement made nearly 290 km apart during Anita by the WARF radar and moored open-ocean current meters show reasonable agreement. The WARF data set is not limited to the results presented in this paper. Other analyses of the radar data that were not obtained in the vicinity of the buoy are also available. These experiments indicate that during a hurricane, HF skywave radar can provide operational surface data that are as accurate as the more recognized in-situ measurements. The supportive surface data supplied by the WARF radar would prove particularly useful for tracking during early formative stages of hurricanes when multiple centers may be observed or when cirrus shielding may obscure visual location by satellite cloud photography. The high resolution, large coverage area, real-time steering and continuous monitoring capabilities are unique to skywave radar. The hurricane data obtained from skywave radar complements data obtained from satellites, aircraft, and buoys.

ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research (AFOSR) under Contract No. F9620-76-C-0023 and the National Oceanic and Atmospheric Administration (NOAA) under Contract No. 03-7-022 3511. We gratefully acknowledge the help of C. Glassmeyer, W. Preuss, G. Tomlin, and C. Powell of SRI in collecting data at WARF; D. Westover and W. Weir for writing the real-time sampling and post-processing software and B. Richards and J. King for typing the manuscipt.

REFERENCES

BARNUM, J. R., J. W. Maresca, Jr., and S. M. Serebreny, 1977, "High-Resolution Mapping of Oceanic Wind Fields with Skywave Radar," IEEE Trans. on Antennas and Propagation, AP-25, pp. 128-132.

BARRICK, D. E., 1972a, "First-order Theory and Analysis of MF/HF VHF Scatter From the Sea," IEEE Trans On Antennas and Propagation, AP-20, pp. 2-10.

BARRICK, D. E., 1972b, "Remote Sensing of Sea State by Radar," Remote Sensing of the Troposphere, V. E. Derr, ed., U.S. Government Printing Office, Washington, D. C.

BARRICK, D. E., J. M. Headrick, R. W. Bogle, and D. Crombie, 1974, "Sea Backscatter at HF: Interpretation and Utilization of the Echo," Proceedings of the IEEE, 62, pp. 673-680.

BARRICK, D. E. 1977a, "Extraction of Wave Parameters from Measured HF Sea-Echo Doppler Spectra," Radio Science, Vol. 12, pp. 415-424.

BARRICK, D. E., M. W. Evans, and B. L. Weber, 14 October 1977, "Ocean Surface Currents Mapped by Radar," Science, Vol. 198, pp. 138-144.

BARRICK, D. E., 1977b, "The Ocean Waveheight Nondirectional Spectrum from Inversion of the HF Sea-Echo Doppler Spectrum," Remote Sensing of Environment, Vol. 6, pp. 201-227.

BARRICK, D. E. and B. J. Lipa, 1978, "Ocean Surface Features Observed by HF Coastal Ground-wave Radars: A Progress Review," Ocean Wave Climate, M. D. Earle and A. Malahoff, Eds.,(Plenum Publishing Company, New York).

CARDONE, V. J., 1969, "Specification of the Wind Distribution in the Marine Boundary Layer for Wave Forecasting," Report TR 9-1, Geophys. Sci. Lab., New York University NTIS, No. AD702490.

CARDONE, V. J., D. B. Ross, and M. R. Ahrens, December 13-16, 1977, "An Experiment in Forecasting Hurricane Generated Sea States," Proceedings of the 11th Technical Conference on Hurricanes and Tropical Meteorology, Miami, Florida.

CARDONE, V. J., and D. B. Ross, 1978, "State of Art Wave Predictions and Data Requirements," in Ocean Wave Climate, M. D. Earle and A. Malahoff, Eds. (Plenum Publishing Company, New York).

CROMBIE, D. D., 1955, "Doppler Spectrum of Sea Echo at 13.56 Mc/s," Nature, Vol. 175, pp. 681-682.

ELSBERRY, R. L., N. A. S. Pearson, L. B. Corgnati, Jr., 1974 , "A Quasi-Empirical Model of the Hurricane Boundary Layer," *J. Geophys. Research*, Vol. 79, pp. 3033-3040.

EWING, J. A. 1969 , "Some Measurements of the Directional Wave Spectrum," *J. Marine Research*, Vol. 27, pp. 163-171.

GEORGES, T. M. and J. W. Maresca, Jr., "The Effect of Radar Beamwidth on the Quality of Sea Echo Doppler Spectra Measured with HF Skywave Radar," accepted by *Radio Science*.

HASSELMANN, K. D. B. Ross, P. Muller, and W. Sell, 1976 , "A Parametric Wave Prediction Model," *J. Phys. Oceanog.*, Vol. 6 pp. 200-228.

LIPA, B. J., 1977a , "Derivation of Directional Ocean-wave Spectra by Integral Inversion of Second-Order Radar Echoes," *Radio Science*, Vol. 12, pp. 425-434.

LIPA, B. J., 1977b , "Inversion of Second-order Radar Echoes from the Sea," *J. Geophys. Research*, Vol, 83, pp. 959-962.

LONGUET-HIGGINS, M. S., D. E Cartwright, and N. D. Smith, 1963 , "Observations of the Directional Spectrum of Sea Waves Using Motions of a Floating Buoy," *Ocean Wave Spectra*, pp. 111-136, (Prentice-Hall: Englewood Cliffs, New Jersey).

MARESCA, J. W.,Jr., J. R. Barnum and K. L. Ford, 1976 , "HF Skywave Radar Measurements of Coastal and Open Ocean Surface Currents" (Abstract , paper presented at the 1976 Annual Meeting of USNC/URSI University of Massachusetts, Amherst, Massachusetts, October 11-15.

MARESCA, J. W., Jr. and K. L. Ford, 1976 , "HF Skywave Radar Measurementof the Surface Current in the North Pacific Ocean," Technical Note RML-TN-9, Stanford Research Institute, Menlo Park, California.

MARESCA, J. W., Jr. and C. T. Carlson, 1977 , "Tracking and Monitoring Hurricanes by HF Skywave Radar over the Gulf of Mexico," Technical Report 1, SRI International, Menlo Park, California.

MARESCA, J. W., Jr., 1978 , "High Frequency Skywave Radar Measurements of Waves and Currents Associated with Tropical and Extra-Tropical Storms," *Ocean Wave Climate*, eds. M. D. Earle and A. Malahof (Plenum Publishing Company, New York).

MARESCA, J. W., Jr. and C. T. Carlson, 1978 , "Tracking and Monitoring Hurricanes by HF Skywave Radar Over the Gulf of Mexico," Final Report, SRI International, Menlo Park, California.

MARESCA, J. W., Jr. and R. Padden, "HF Radar Measurements of the Tidal Currents in San Francisco Bay," in preparation.

MARESCA, J. W., Jr. and T. M. Georges, "HF Skywave Radar Measurement of the Ocean Wave Spectrum," submitted to *J. Geophys. Research*.

MARESCA, J. W., Jr. and J. R. Barnum, "Remote Measurements of the Position and Surface Circulation of Hurricane Eloise by Skywave Radar," accepted by *Monthly Weather Review*.

MITSUYASU, H., F. Tasai, T. Syhara, S. Mizuno, M. Ohkusu, T. Honda, and K. Rikiishi, 1975 , "Observations of the Directional Spectrum of Ocean Waves Using a Clover Leaf Buoy," *J. Phys. Oceanog.*, Vol. 5, No. 4, pp. 750-760.

ROSS, D. B., 1976 , "A Simplified Model for Forecasting Hurricane Generated Waves" (Abstract), *Bull. Am. Meteorol. Soc.*, presented at Conference on Atmospheric and Oceanic Waves, Seattle, Washington, March 29-April 2, 1976 American Meteorological Society.

ROSS, D. B. and V. J. Cardone, 1978 , "A Comparison of Parametric and Spectral Hurricane Wave Prediction Products," to be published in the Proceedings of "NATO Symposium on Turbulent Fluxes Through the Sea Surface, Wave Dynamics and Prediction," Marseilles, France, September 12-16, 1977 (Plenum Publishing Company, New York).

SHEETS, R. C., 1978 , "Hurricane Anita - A New Era in Airborne Research," *Mariner's Weather Log*, Vol. 22, No. 1, pp. 1-8.

SMITH, N. P., 1978 , "Longshore Currents on the Fringe of Hurricane Anita," *J. Geophys. Research.*, Vol. 83, No. C12, pp. 6047-6051.

"SRI Remote Measurements Laboratory Research Capabilities" brochure, Stanford Research Institute, Menlo Park, California 1977 .

STEWART, R. H. and J. W. Joy. 1974 , "HF Radio Measurements of Surface Currents," *Deep-Sea Research*, Vol. 21, pp. 1039-1049.

STEWART, R. H. and J. R. Barnum, 1975 "Radio Measurements of Oceanic Winds at Long Ranges: An Evaluation," *Radio Science*, Vol. 10, pp. 853-857.

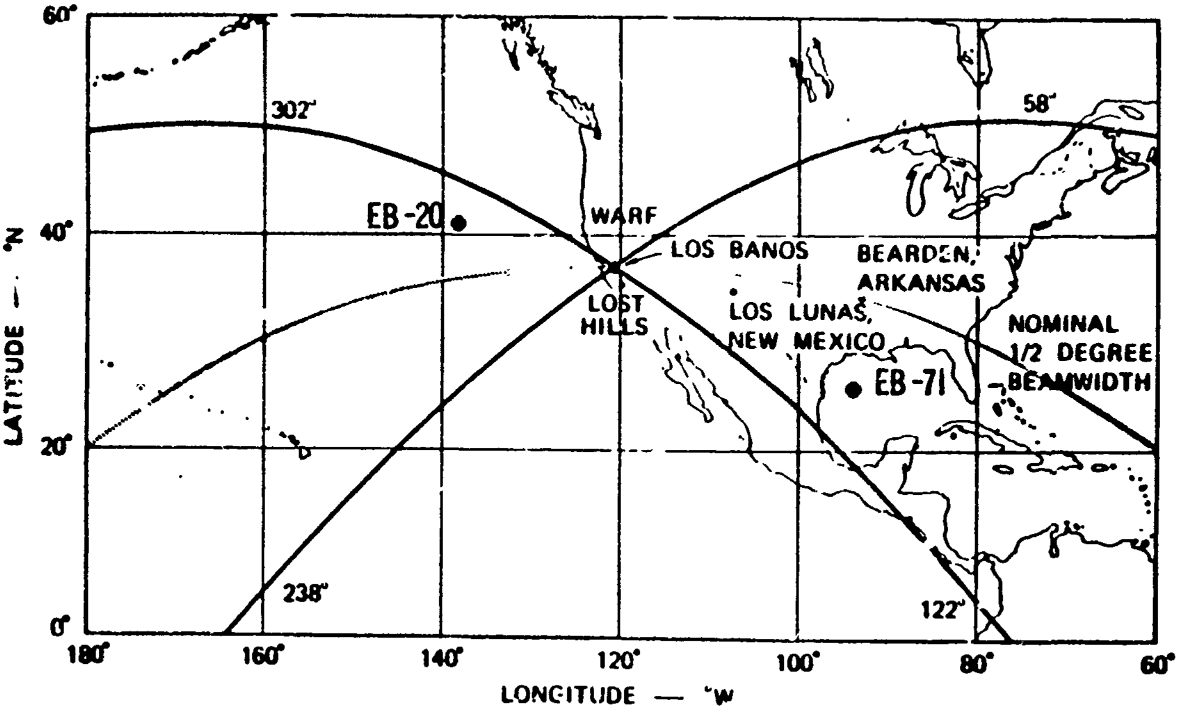

60°
40°
20°
0°
180°
160°
140°
120°
100°
80°
60°
LATITUDE — °N
LONGITUDE — °W
302°
58°
238°
122°
EB-20
WARF
LOS BANOS
BEARDEN, ARKANSAS
LOST HILLS
LOS LUNAS, NEW MEXICO
EB-71
NOMINAL 1/2 DEGREE BEAMWIDTH

FIGURE 1 COVERAGE AREA OF THE WARF HF SKYWAVE RADAR. All Anita measurements were made west of 88° W in the Gulf of Mexico.

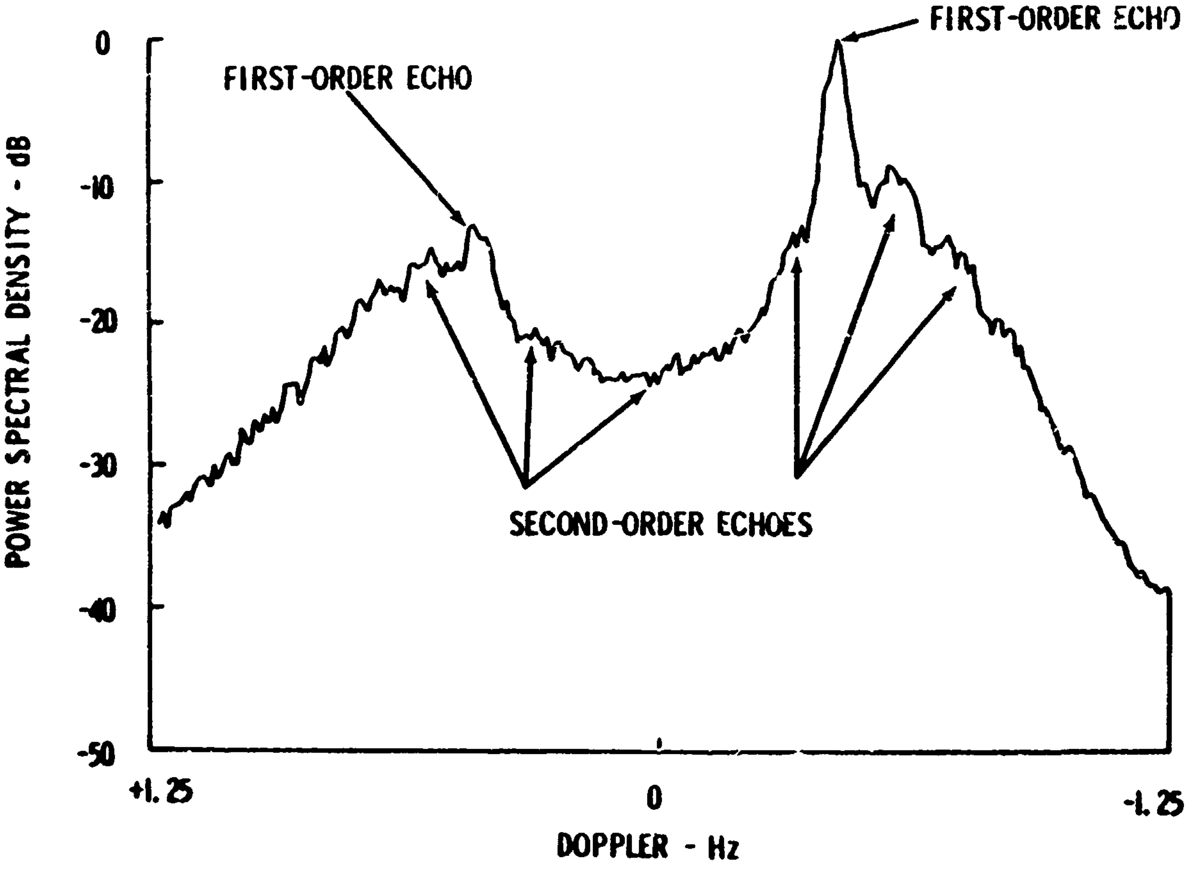


FIGURE 2 AVERAGE SEA ECHO DOPPLER SPECTRUM RECORDED WITHIN 35 KM OF THE CENTER OF HURRICANE ANITA AT 2343Z ON AUGUST 31, 1977

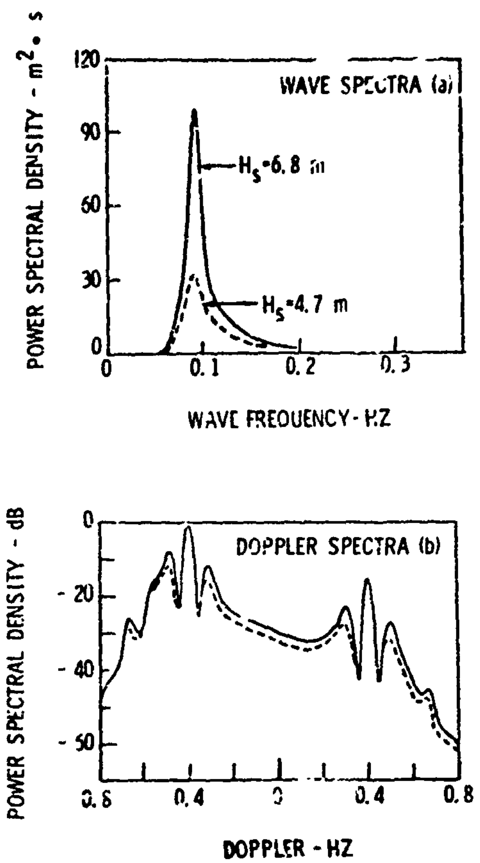


FIGURE 3 EXAMPLE OF TWO SYNTHETIC DOPPLER SPECTRA (b) PRODUCED FROM TWO INPUT WAVE SPECTRA (a) WITH THE SAME DIRECTIONAL DISTRIBUTION AND RADAR-TO-WIND DIRECTION, BUT DIFFERENT TOTAL WAVE ENERGY (0.02 Hz resolution)

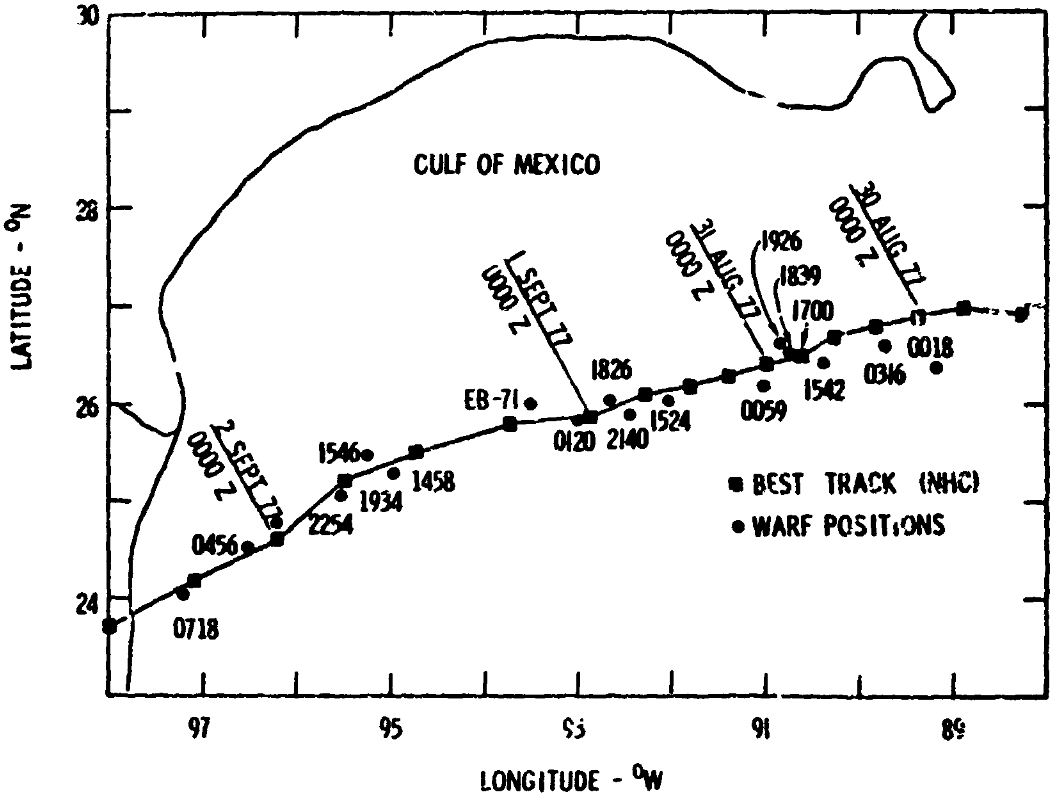

30
28
26
24
LATITUDE - °N
97
95
93
91
89
LONGITUDE - °W
GULF OF MEXICO
30 AUG 77 0000 Z
31 AUG 77 0000 Z
1 SEPT 77 0000 Z
2 SEPT 77 0000 Z
1926
1839
1700
0018
0316
1542
0059
1826
1524
2140
0120
EB-71
1546
1458
1934
2254
0456
0718
■ BEST TRACK (NHC)
● WARF POSITIONS

FIGURE 4 WARF-MEASURED TRACK OF HURRICANE ANITA PRODUCED FROM THE RADAR WIND MAPS

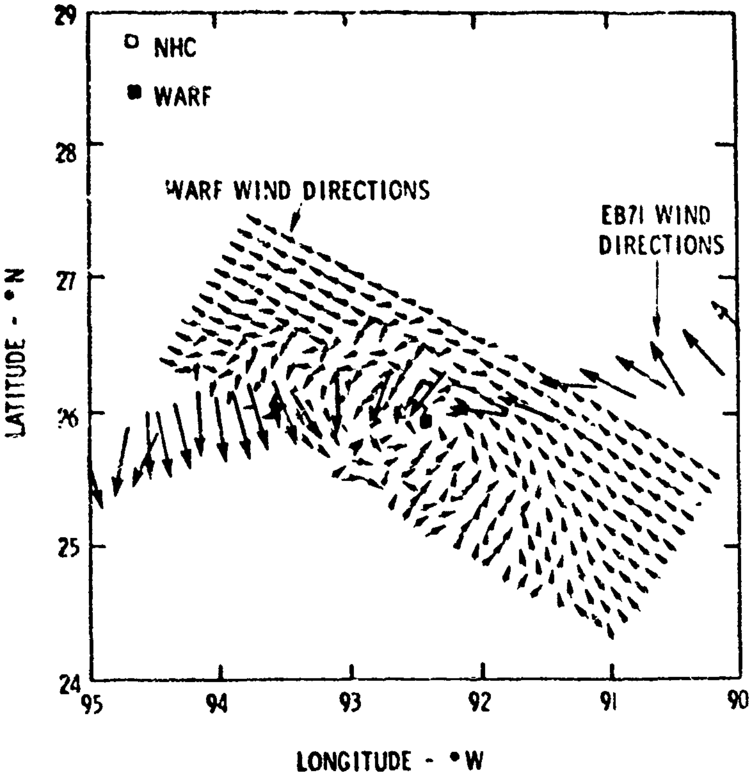

○ NHC
■ WARF
WARF WIND DIRECTIONS
EB71 WIND DIRECTIONS
LATITUDE - °N
29
28
27
26
25
24
95
94
93
92
91
90
LONGITUDE - °W

FIGURE 5 WARF-DERIVED WIND DIRECTION MAP MADE FOR ANITA AT 2140Z ON AUGUST 31, 197

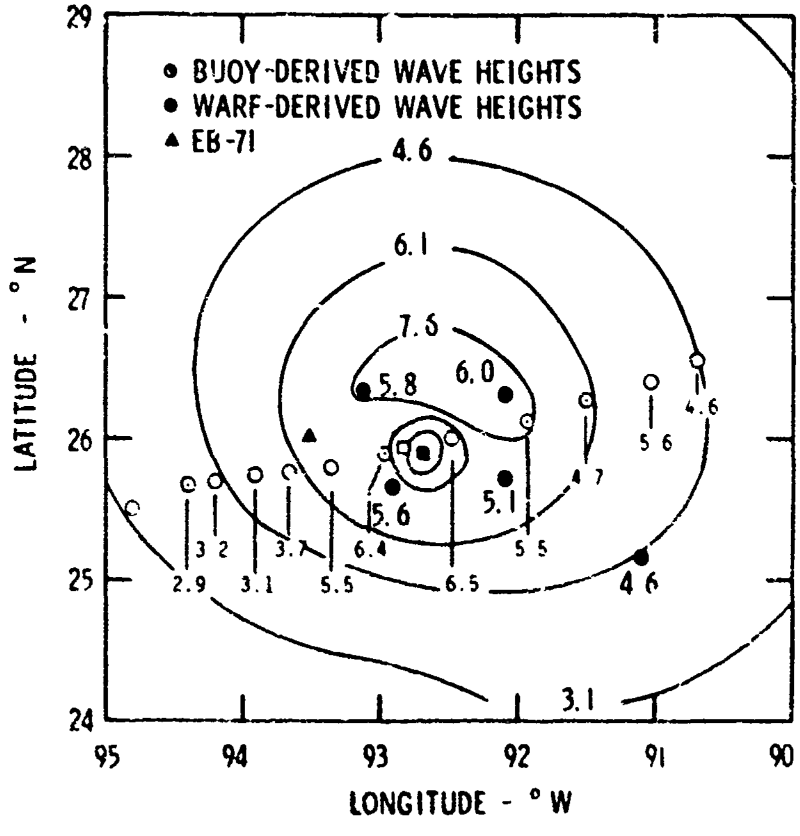

⊙ BUOY-DERIVED WAVE HEIGHTS
● WARF-DERIVED WAVE HEIGHTS
▲ EB-71
4.6
6.1
7.6
5.8
6.0
4.6
5 6
4 7
5.1
5.6
3 2
3.7
6.4
5 5
2.9
3.1
5.5
6.5
4 6
3.1
LATITUDE - °N
29
28
27
26
25
24
95
94
93
92
91
90
LONGITUDE - °W

FIGURE 6 COMPARISON OF THE WARF-DERIVED SIGNIFICANT WAVE HEIGHTS (m) MEASURED BETWEEN [illegible] ON AUGUST 31, 1977 and 0020Z ON SEPTEMBER 1, 1977 AND THE EB-71 DERIVED SIGNIFICANT WAVE HEIGHTS MADE BETWEEN 0600Z ON AUGUST 31, 1977 AND 1800Z ON SEPTEMBER 1, 1977. The wave height contours are reproduced from Figure 9 of Cardone, V. J., et al., (1977). The letter designations are given on Figure 7.

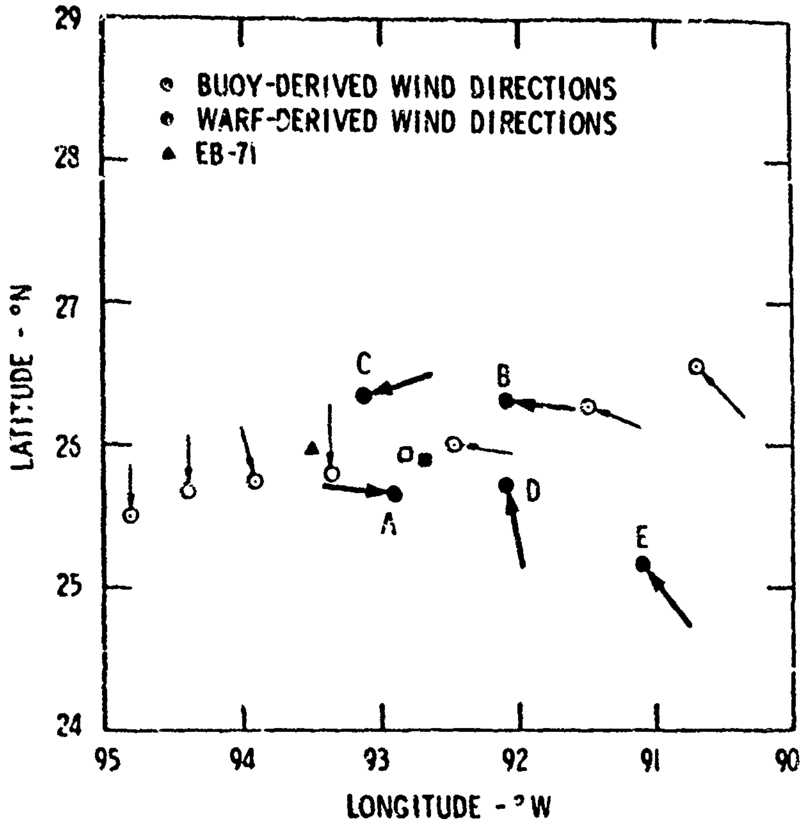


FIGURE 7 COMPARISON OF THE WARF-DERIVED WIND DIRECTIONS (→) MADE BETWEEN 2314Z ON AUGUST 31, 1977 AND 0020Z ON SEPTEMBER 1, 1977 AND THE EB-71-DERIVED WIND DIRECTIONS (→) MADE BETWEEN 0600Z ON AUGUST 31, 1977 AND 1800Z ON SEPTEMBER 1, 1977

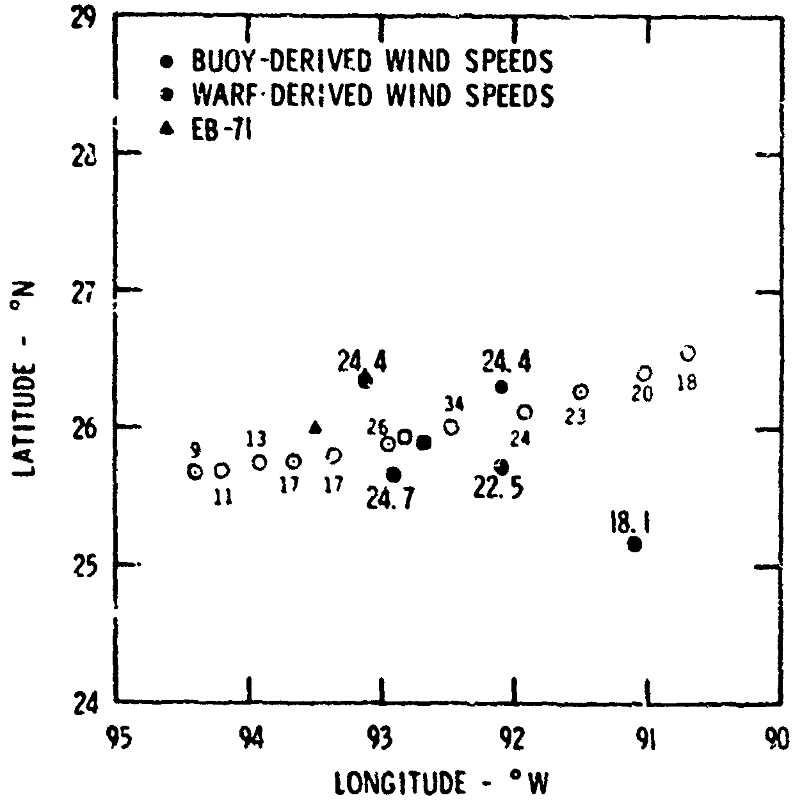


FIGURE 8 COMPARISON OF THE WARF-DERIVED WIND SPEEDS (m/s) MEASURED BETWEEN 2314Z ON AUGUST 31, 1977 AND 0020Z ON SEPTEMBER 1, 1977 AND THE EB-71-DERIVED WIND SPEEDS (m/s) MEASURED BETWEEN 0600Z ON AUGUST 31, 1977 and 1800Z ON SEPTEMBER 1, 1977

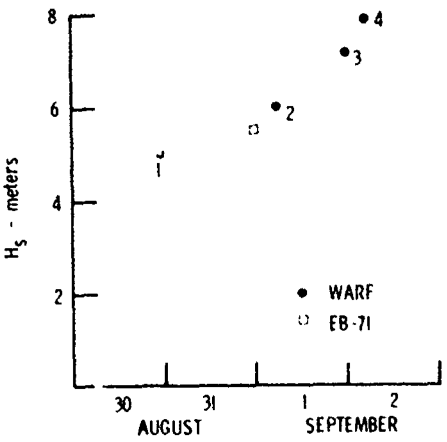


FIGURE 9 DAILY ESTIMATES OF SIGNIFICANT WAVE HEIGHT FOR HURRICANE ANITA MADE AT THE WARF RADAR

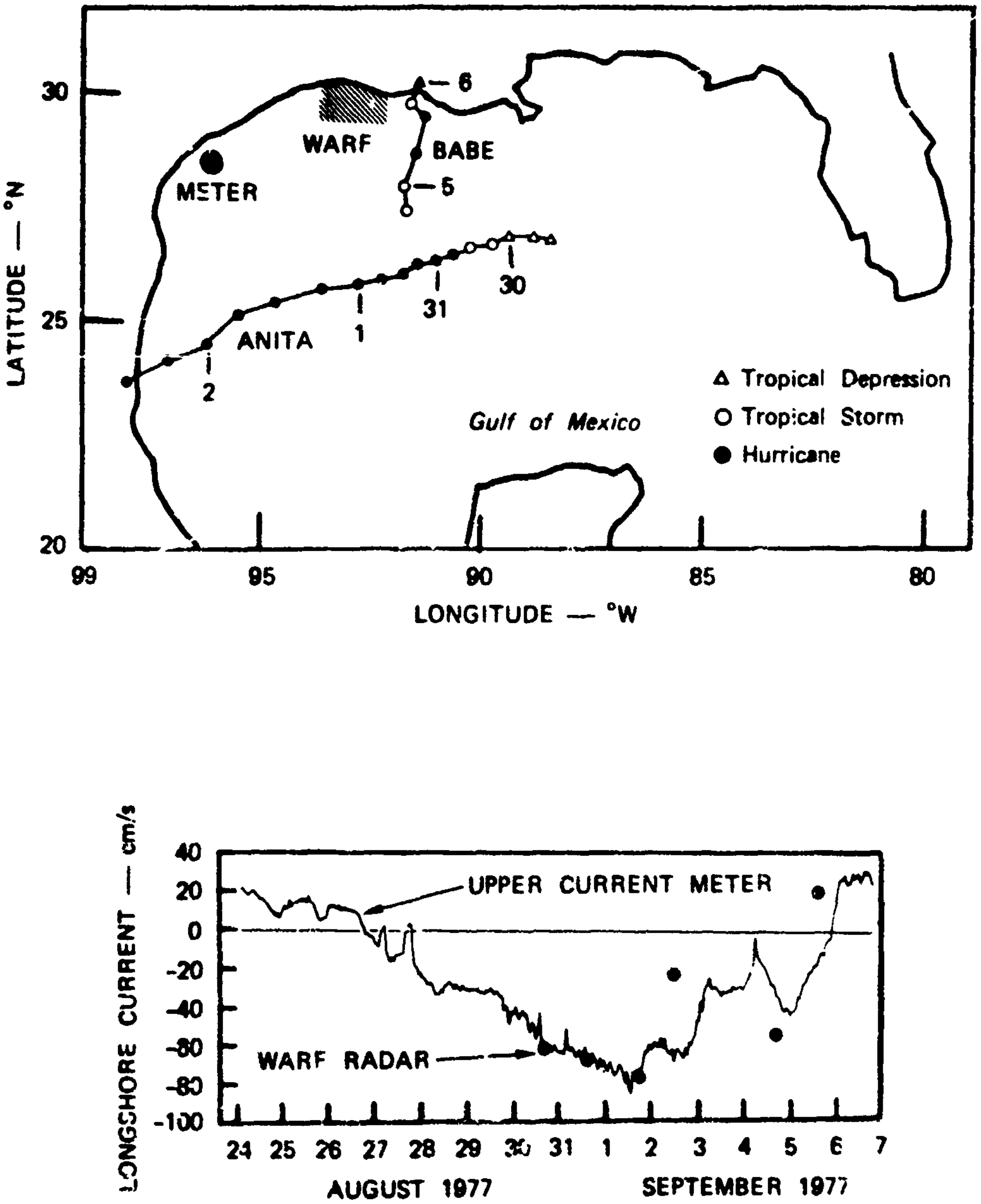

30
25
20
LATITUDE — °N
99
95
90
85
80
LONGITUDE — °W
WARF
METER
BABE
6
5
ANITA
30
31
1
2
Gulf of Mexico
△ Tropical Depression
○ Tropical Storm
● Hurricane
40
20
0
-20
-40
-60
-80
-100
LONGSHORE CURRENT — cm/s
UPPER CURRENT METER
WARF RADAR
24
25
26
27
28
29
30
31
1
2
3
4
5
6
7
AUGUST 1977
SEPTEMBER 1977

FIGURE 10 LOCATION (a) AND RESULTS (b) OF WARF RADAR AND CURRENT METER OBSERVATIONS MADE DURING HURRICANES ANITA AND BABE (30 AUGUST - 6 SEPTEMBER, 1977). Continuous current record reproduced from Figure 2 of Smith, N. P., (1978).

OCEAN SWELL PARAMETERS FROM NARROW-BEAM HF RADAR SEA ECHO

Belinda Lipa
SRI International
Menlo Park, California

Donald Barrick
NOAA Wave Propagation Laboratory
Boulder, Colorado

ABSTRACT

This paper describes inversion methods for HF radar sea-echo Doppler spectra, giving parameters of the ocean wave spectrum in the important long wavelength region. Radar spectra exhibiting very narrow spikes in the higher-order structure adjacent to the first-order lines are indicative of ocean swell with a single dominant wavelength; it is shown how to interpret such a radar spectrum to give swell period, direction and rms waveheight. When the higher-order echo peaks are more spread in frequency, a simple model for the swell yields a closed-form expression for the four second-order swell peaks; parameters identifiable from this model, in addition to swell rms height, direction, and dominant period, include the angular spread and frequency spread of the swell. Finally, when the swell Doppler peaks are quite spread, it is shown how to use integral inversion to give Fourier coefficients of the ocean-wave spectrum as a function of ocean wavelength. At a given wavelength these coefficients are interpreted in terms of a cardioid model for the angular distribution to give the mean wave direction and the angular spread. In normal surface-wave experiments, the major source of error or noise is the random surface height of the sea; we describe the resulting statistics of the radar spectrum and trace the propagation of uncertainty to the derived ocean parameters.

1. INTRODUCTION.

(a) HF Radar Scatter

Crombie (1955) discovered the mechanism behind HF radar sea scatter nearly 25 years ago by spectrally analyzing the received time series. Using the gravity-wave dispersion equation that relates the velocity to the square root of the ocean wavelength, he correctly deduced that the two, sharp, symmetrically positioned Doppler peaks were produced by those ocean wavetrains exactly half the radio wavelength, moving toward and away from the radar. We now call these obvious peaks the "Bragg lines" since the mechanism producing them is the first-order Bragg effect. When theories confirmed this mechanism (Wait, 1966; Barrick, 1972), and showed that the strength of these echo peaks is proportional to the heights of the corresponding Bragg-scattering ocean wavetrains, scientists became enthusiastic about the prospect of using HF radars to measure sea-state parameters. In order to directly observe the interesting, long ocean waves (e.g., swell) that are the essence of "sea state" directly by this first-order Bragg effect, one would have to operate the radar in the lower MF region; huge antenna size requirements, heavy spectrum utilization in this region, and ionospheric problems all dictate against such a system (Barrick, 1978; Barrick and Lipa, 1979a). At mid and upper HF, the first-order Bragg peaks are produced by the less interesting, shorter waves, and hence using these alone, radio oceanographers resigned themselves to extracting only wind-direction information because the short waves align themselves quickly with the wind (Long and Trizna, 1973; Stewart and Barnum, 1975).

The more sophisticated radar systems and digital signal processors of the mid-sixties showed a lower-level spectral continuum surrounding the first-order Bragg peaks that was definitely established as sea echo. Barrick (1971, 1972) derived a theoretical expression for this continuum, showing that it was explained as second-order ocean wave-wave interactions. This power spectral density vs. frequency manifested itself as an integral containing a product of two waveheight directional spectra, the arguments of which were the double Bragg-interacting wave vectors. This gave scientists hope that the integral could be inverted, yielding directional information of the more interesting, longer ocean waves (swell) from measurements at a single HF frequency. Hasselmann (1971), however, observed that if the kernal of the integral (called the "coupling coefficient") were a constant, the second-order spectral continuum would nearly replicate the ocean nondirectional spectrum, thereby yielding no long-wave directional information. Barrick (1977a,b) developed and tested an approximate technique that yielded this nondirectional information for longer waves (i.e., waveheight, dominant wave period, and the nondirectional spectrum) by combining the two second-order peaks surrounding each Bragg line; indeed, the results proved to be relatively insensitive to direction. Maresca and Georges (1979) developed an even simpler method to obtain waveheight by combining the energy in both second-order spectral sidebands; again, this process renders the results insensitive to direction. All of this suggested that extracting directional information of long waves from the second-order echo might prove elusive (even though swell height and period were obtainable).

Lipa (1977, 1978) was the first to show that directional information can be derived from the second-order echo, even though she concentrated on the wind-wave region (rather than swell), and used a model that separated the directional and frequency dependences as independent multiplicative factors. Often the directional dependence is dependent on wavenumber, as when one or more swells and wind waves with different wavenumbers arrive from different directions; hence this model is not generally valid. An example is illustrated in the sea-echo Doppler spectrum of Figure 1, where second-order wind-wave and swell peaks are both present, and are from different directions.

This paper shows that swell direction (as well as height and period) may indeed be derived from the second-order sea-echo spectral peaks. This reversal in outlook results from closer examination of models of the radar Doppler spectrum produced by swell, which shows that: (i) the positions of the echo swell peaks vary slightly with a swell direction; (ii) the amplitudes of the echo swell peaks depend fairly strongly on the coupling coefficient, and this in turn is shown here to vary significantly with swell direction. Consequently, we have been able to develop analytical techniques for extracting swell height, direction, and period, and in some cases, angular and frequency width of the swell. This is done for a variety of models, applicable for situations from one extreme where the swell has a very narrow frequency and angular distribution to the case where swell directions and frequencies span a wider range. Closed-form expressions are derived for two situations, and a general inversion technique is presented for a third. Expressions permitting estimates of errors in these parameters are derived and presented.

(b) The Nature of Swell

"Swell" is used loosely by mariners to refer to the longest waves present at a given time on the ocean, in contrast to "chop," or the shorter, less organized waves. Oceanographers technically define swell as waves produced by storms in another area and/or at an earlier time (Kinsman, 1965). In contrast to "wind waves" that are being constantly generated by local winds, swell waves appear more organized (their quasi-sinusoidal nature permits the identification of a direction and period) and generally have longer periods. Although severe storms or hurricanes will often directly arouse these long-period waves, recent studies (Hasselmann et al., 1976) show that they are also generated from energy transfer due to higher-order nonlinear interactions among the shorter waves. For example, 20 m/s (∼40 knot) winds should directly excite waves with periods, T, no longer than 12.8 s (i.e., to first order $T = 2\pi v/g$, where g is the acceleration of gravity and v is the wave phase velocity, whose greatest value corresponds to waves which just keep up with the wind). Swells with periods between 10 and 18 seconds are nearly always observed on the oceans even though local winds may be quite low.

For this analysis, we include as swell any waves whose periods exceed 9 s (or whose frequencies are less than 0.11 Hz). While this definition generally includes waves oceanographically categorized as swell above, we select this definition here because waves lying in this region of the ocean spectrum allow considerable simplification of the second-order Doppler spectrum above mid-HF. As a result, we can linearize the integral equation and in some cases employ simple swell models to derive closed-form expressions.

The simplest swell model is the limit of "old" swell, where the finite-sized storm generating area recedes to infinity in distance and past time; in this limit the swell waveheight directional spectrum becomes an impulse function vs. wave frequency and direction. Such a model is usually applicable when the storm area was greater than 2000 km distant from the observer, and will be shown to produce a Doppler spectrum also consisting of impulse functions. "New" swell refers to waves from a storm area whose distance from the observer was not much different from the dimensions of the patch itself; this is somewhat analogous to being in the near field of an antenna. An example is taken from a situation existing at a recent U.S. Pacific coastal HF radar experiment site, where observed swell was explained by after-the-fact meteorological analyses (called "hindcasts"). These hindcasts — while admittedly crude pictures — showed that swell in this case was generated from a storm approximated to lie in a rectangular area 650 km x 740 km, whose leading edge at the wind cessation was 925 km from the radar. Simple hindcasting methods showed that at the onset, waves from this storm area were ∼1.07 m high (rms), had a mean period of ∼14 s (wave frequency ∼.071 Hz) but a spread of ∼.02 Hz about this mean; in addition, wave directional spread about the mean was ∼40°. For this situation, the "impulse function" model for swell was seen to yield inaccurate results, but a narrow-band finite-beam swell model analyzed below produced the correct interpretation. For even closer "near-field" swell situations, the swell spectrum is so spread that even these finite-width models are inadequate; in this case, full inversion of the integral is called for.

2. THE RADAR CROSS SECTION.

We consider narrow-beam radar systems with vertical polarization directed at or near grazing incidence. Vertical polarization is the only case of importance because: (i) ground or surface-wave radars require vertical polarization to attain reasonable propagation distances; (ii) the horizontally polarized component of scatter from the sea is several orders of magnitude lower than vertical, so that even skywave radars effectively discriminate only the vertical mode for sea echo. Barrick (1972) gives the following relation for the first-order average radar cross section per unit surface area $\sigma(\omega)$ at a Doppler shift ω from the carrier:

$$\sigma^{(1)}(\omega) = 2^6\pi k_o^4 \sum_{m=\pm1} S(-2m\vec{k}_o)\delta(\omega - \omega_B) \tag{1}$$

where $\vec{k}_o$ is the radar wave vector of magnitude k_o, S is the directional ocean wave spectrum and ω_B is the first order Bragg frequency given by

$$\omega_B = \sqrt{2gk_o} \tag{2}$$

with g the gravitational acceleration. The first-order spectrum described by (1) consists of two impulse (i.e., Dirac-delta) functions symmetrically placed about the carrier frequency at positions defined by the deep-water dispersion equation, and with amplitudes proportional to the directional ocean-wave spectrum at the Bragg wave vectors $\pm 2\vec{k}_o$, as shown in Fig. 1. In practice, the first-order lines are displaced from their ideal positions because of underlying ocean currents and spread over a finite frequency band due to current turbulence and system effects.

The analogous second-order radar cross section and the directional ocean-wave spectrum are related by a nonlinear integral equation given by Barrick (1972). We express this equation in the following form:

$$\sigma^{(2)}(\omega) = 2^6\pi k_o^4 \sum_{m,m'=\pm1} \int_o^\infty \int_{-\pi}^{\pi} |\Gamma|^2 \delta(\omega - m\sqrt{gk} - m'\sqrt{gk'}) S(m\vec{k}) S(m'\vec{k}') k\,dk\,d\theta \tag{3}$$

where the vectors $\vec{k}$ and $\vec{k}'$ obey the relation

$$m\vec{k} + m'\vec{k}' = -2\vec{k}_o \ . \tag{4}$$

The quantity Γ is a coupling coefficient which includes the effects of both hydrodynamic and electromagnetic nonlinearities carried to second order in a perturbation theory.

$$\Gamma_{m,m'} = \frac{1}{2}\left[\frac{(\vec{k}\cdot\vec{k}_0)(\vec{k}'\cdot\vec{k}_0)/k_0^2 - 2\vec{k}\cdot\vec{k}'}{\sqrt{\vec{k}\cdot\vec{k}'} + k_0\Delta}\right] - \frac{i}{2}\left[k + k' - \frac{(kk' - \vec{k}\cdot\vec{k}')}{mm'\sqrt{kk'}}\left(\frac{\omega^2+\omega_B^2}{\omega^2-\omega_B^2}\right)\right] \qquad (5)$$

Barrick and Snider (1977) have shown that the sea echo voltage spectrum after N sample averaging is χ-square distributed with N degrees of freedom. They also show that Doppler spectra (both first and second order) become uncorrelated for time intervals greater than approximately 25 seconds for radar frequencies greater than 10 MHz, and that Doppler spectra from different range cells are statistically independent to spatial separations as small as 5 km. These experimental results may be used to optimize the amount of incoherent averaging. In practice, at least 10 and normally more than 100 spectra are incoherently averaged before analysis.

The first-order echo is used to linearize the integral equation as follows in an approximation valid at upper HF for higher sea states (e.g., at 25 MHz for rms ocean waveheights greater than 0.4 m). For small values of k/k_0 it follows from (4) that the vector $\vec{k}'$ is approximately equal to the Bragg wavevector. The second ocean spectral factor in the integral equation (3) may therefore be removed by normalizing the second-order spectrum by the power in the neighboring first-order peak. This normalization scheme, proposed and implemented by Barrick (1977a,b) linearizes the integral and removes unknown system gains and path losses from the data.

Because of previously discussed frequency smearing, we integrate the second-order spectrum over a finite frequency window δ. The energy in the first-order line is obtained by integrating over a frequency window of width δ_h, which can conveniently be taken as the width between the half-power points. We then define the following experimental parameters:*

$$r_{m,m'}(\omega) = \frac{\int_{\omega-\delta/2}^{\omega+\delta/2} \sigma^{(2)}(\omega)d\omega}{\int_{m\omega_B-\delta_h/2}^{m\omega_B+\delta_h/2} \sigma^{(1)}(\omega)d\omega} . \qquad (6)$$

When we employ models to describe the swell (as in the first two cases considered below), δ is the width of the second-order swell peak. In this case, we have integrated out the frequency, and are concerned only with the energy in the Doppler peaks; for this case, we redefine $r_{m,m'}(\omega)$ in (6) above as $\mathcal{R}_{m,m'}$. When we do complete inversion of the integral, however, (as in the third case below), $\delta = \delta_h$ because the first-order peak width determines the ultimate frequency resolution attainable in the inversion process. Substituting expressions (1) and (2) for the first and second-order radar cross sections into (6) gives the approximate linear integral equation.

$$R_{m,m'}(\omega) = \int_{\omega-\delta/2}^{\omega+\delta/2}\int_{-\infty}^{\infty}\int_{0}^{2\pi} |\Gamma_{m,m'}|^2\,\delta(\omega - m\sqrt{gk} - m'\sqrt{gk'})\,S(m\vec{k})\,k\,d\omega\,dk\,d\theta . \qquad (7)$$

We also define and employ the derivative of (7), i.e., $R'_{m,m'}(\omega)$; this is identical to (7) but with the integral over ω removed. This quantity describes the shapes of four radar echo frequency peaks vs. Doppler frequency, ω, based on the linearizing approximation discussed above. These four peaks — as represented by the four combinations $m,m' = \pm 1$ — will be seen to be positioned two each around the two first-order Bragg lines, as shown experimentally in Fig. 1. In contrast with previous impressions in the literature, the coupling coefficient in the above expression varies significantly with direction. If this were not so, it would be easy to show (i.e., if $\Gamma_{m,m'}$ were constant) that the four Doppler echo swell peaks defined by (7) would all have the same amplitudes and shapes, making it difficult to extract many swell parameters. More relevant details about the coupling coefficient are found in the Appendix.

In the next section, we show how to invert equation (7) to give the ocean-wave directional spectrum. The standard error in the derived parameters depends on the data covariance matrix. To derive this, we note that $R_{m,m'}$ is defined in equation (6) to be the ratio of sums of χ^2 power spectral samples and is therefore itself a random variable which follows an F distribution (Barrick, 1979). If the numerator and denominator are sums of M and N independent samples respectively, in the limit of large N the F distribution becomes χ^2 with an effective number of degrees of freedom ν_e given by

---

* Lower-case R's are used to denote measured quantities while upper case are theoretical, script R's denote total energy in the peak while block R's denote the actual peak function itself.

$$1/I_e = 1/M + 1/N \; . \tag{8}$$

Barrick and Lipa (1979b) show theoretically that power spectral values at different Doppler frequencies are statistically independent. The covariance matrix of a set of ratios $R_{i,i'}$ is therefore diagonal with elements given by:

$$\langle \Delta R_{i,i'} \Delta R_{j,j'} \rangle = \frac{R^2_{i,i'}}{I_e} \, \delta_{ii'} \, \delta_{jj'} \; . \tag{9}$$

3. APPLICATION TO OCEAN SWELL.

We consider ocean waves with temporal periods larger than 9 seconds. Such waves, when generated by storms at great distances, result in a wave field characterized by a very narrow band of wave frequencies and directions; the ocean-wave spectrum may then often be modeled as impulse functions in wavelength and direction. For a storm center at closer range, the wave direction and frequency spread is small but finite, and a narrow-band, narrow-beam model yields a useful closed-form result. Finally, a mixture of swell and wind waves may be present with a spectrum which is broadly distributed in both wavelength and direction. We show how analysis of the radar spectrum allows one to distinguish between these possibilities, and determine ocean-wave parameters and their associated statistical error.

(a) Swell of a Single Wavelength and Direction

We write the total waveheight spectrum at a wavevector $\vec{k}$ as the sum of a continuous wind-wave spectrum $S_w(\vec{k})$ and a swell component which is an impulse function in wavevector space:

$$S(\vec{k}) = S_w(\vec{k}) + H_s^2 \delta(\vec{k} - \vec{k}_s) \tag{10}$$

where H_s is the rms waveheight of the swell and $\vec{k}_s$ is the swell wavevector which has magnitude k_s and propagation angle θ_s with respect to the radar beam (i.e., away from the radar). When this spectrum is substituted into the integral equation (7), the right hand side separates into a continuous wind-wave spectrum, cross-spectral terms due to the interaction of swell and wind waves given by

$$R_{m,m'}(\omega) = H_s^2 \int_{\phi-\pi/2}^{\phi+\pi/2} \int_0^{\infty} \int_0^{2\pi} \Gamma^2_{m,m'} \, \delta(\omega - m\sqrt{gk} - m'\sqrt{gk'}) \, \delta(m\vec{k} - \vec{k}_s) \, k \, dk \, d\theta \, d\phi \; , \tag{11}$$

and terms due to the swell interacting with swell. The latter terms can be seen to be a very resonant phenomenon which — for given swell period and direction — can only be observed at a single radar operating frequency that occurs at lower HF, well below our range of consideration here (i.e., upper HF). The first set of terms representing wind-wave/wind-wave interactions occurs at Doppler frequencies further removed from the Bragg lines than the swell peaks (see Fig. 1) and thus lie outside the range of consideration here; their inversion is treated elsewhere (Barrick and Lipa, 1979a,b). Thus the cross-spectral terms as represented by (11) (where the wind-wave spectrum has already been divided out in the normalization defined by (6)) contain the swell information that will be used in this paper.

The integration in (11) may be performed immediately using the delta function constraints on the integration variables. Four sharp spikes occur in the Doppler spectrum with frequencies and powers given by

$$\omega_{m,m'} = m\omega_s + m'[\omega_B^4 + 2\omega_s^2\omega_B^2\cos\theta_s + \omega_s^4]^{1/4} \; , \tag{12}$$

where

$$\omega_s = \sqrt{gk_s} \; , \; \omega_B = \sqrt{2gk_0} \, , \text{ and}$$

$$R_{m,m'} = 2 H_s^2 |\Gamma_{m,m'}|^2 \; , \tag{13}$$

where $\Gamma_{m,m'}$ is evaluated at $\vec{k}_s$ and $\vec{k}'$ obtained from (4). For typical HF radar frequencies $k_s \ll k_0$, and it follows from (12) that the swell peak frequencies are close to the Bragg frequency $\omega_B = \sqrt{2gk_0}$, normally occurring in the null between the first-order line and the peak of the wind-wave spectrum (see Fig. 1). We define $\Delta\omega^+$ to be the radian frequency displacement between the two swell peaks surrounding the approaching Bragg line; likewise for $\Delta\omega^-$ with respect to the two swell peaks surrounding the receding Bragg line. Estimates of swell spatial wavenumber and direction may be made from the peak frequencies (12) through the equations

$$k_s = \left(\frac{1}{16g}\right)(\Delta\omega^+ + \Delta\omega^-)^2 \tag{14}$$

$$\theta_s = \cos^{-1}\left(8\omega_B \frac{(\Delta\omega^+ - \Delta\omega^-)}{(\Delta\omega^+ + \Delta\omega^-)^2}\right) . \tag{15}$$

The values of $\Delta\omega_{m,m'}$ are calculated by finding the centroid frequency of the spectral peak between the half-power points. Barrick (1979) has calculated a general expression for the standard deviation of a centroid frequency for a spectrum which has χ^2 statistics. This expression is insensitive to the exact shape of the peak; if we therefore assume a Gaussian form for convenience, he shows that the standard deviation of the frequency estimate is

$$\mathrm{Sd}(\omega) = \frac{\Delta}{2}\sqrt{\frac{N}{K}} \quad , \tag{16}$$

where N is the number of samples contained within the half-power width, K is the number of independent spectra used in the sample average, and Δ is the frequency spacing (radian/s) between adjacent spectral points. The corresponding errors in the estimates (14) and (15) for k_s and θ_s follow from linear error propagation theory (Brandt, 1970)

$$\mathrm{Sd}(k_s) = \frac{\Delta}{2}\sqrt{\frac{k_s\ N}{g\ K}} \quad , \tag{17}$$

and

$$\mathrm{Sd}(\theta_s) = 8\omega_B\Delta\sqrt{\frac{N}{K}}\left|\frac{\sqrt{5(\Delta\omega^+)^2 - 6\Delta\omega^+\Delta\omega^- + 5(\Delta\omega^-)^2}}{\sin\theta_s\ (\Delta\omega^+ + \Delta\omega^-)^3}\right| \quad \text{for } \sin\theta_s \neq 0. \tag{18}$$

The uncertainty in the estimate of wavenumber is generally small; that in angle can be large because of the differencing of small frequency shifts in equation (15).

In an independent analysis based on the amplitudes of the swell peaks, expressed theoretically in equation (13), we use the maximum likelihood method (Brandt, 1970) to give estimates of swell waveheight and separate independent estimates of the angle. The following expression is minimized with respect to θ_s and H_s:

$$I(\theta_s,H_s) = \sum_{m,m'} \frac{(\tau_{m,m'} - R_{m,m'})^2}{\langle\Delta R^2_{m,m'}\rangle} \quad , \tag{19}$$

where $\tau_{m,m'}$ are the measured values of the energy in the four Doppler peaks, defined by (6), and where the variance of $R_{m,m'}$ is defined by equation (9). Thus we minimize the sum of the weighted squared deviations of the experimental values of $\tau_{m,m'}$ (including random fluctuations) from the ideal functional form to give the optimum values $\tilde{\theta}_s$ and $\tilde{H}_s$. The residual $I(\tilde{\theta}_s,\tilde{H}_s)$ obeys an approximate χ^2 distribution and we use a χ^2-test on the validity of the ocean-wave spectral model, (10). If the value of $I(\tilde{\theta}_s,\tilde{H}_s)$ is too large, the assumption of this model must be reconsidered and more general models used. These models are considered later in this section.

To obtain the standard deviations in the derived parameters, we expand equation (13) in a Taylor series about $\tilde{\theta}_s,\tilde{H}_s$ to give the approximate expression for the variation in the residual with distance from the optimum in the absence of noise:

$$\Delta R_{m,m'} = 4H_s|\Gamma_{m,m'}|^2\delta H_s + 2H_s^2\frac{\partial(|\Gamma_{m,m'}|^2)}{\partial\theta_s}\ \delta\theta_s \tag{20}$$

which may be expressed in the matrix form:

$$\Delta R = M\Delta\xi \tag{21}$$

where

$$\Delta R = \begin{bmatrix} \Delta R_{1,1} \\ \Delta R_{1,-1} \\ \Delta R_{-1,1} \\ \Delta R_{-1,-1} \end{bmatrix} ; \quad \Delta\xi = \begin{bmatrix} \delta H_s \\ \delta\theta_s \end{bmatrix} ; \qquad M = \begin{bmatrix} 4H_s|\Gamma_{1,1}|^2 & 2H_s^2\dfrac{\partial(|\Gamma_{1,1}|^2)}{\partial\theta_s} \\ \vdots & \vdots \\ 4H_s|\Gamma_{-1,-1}|^2 & 2H_s^2\dfrac{\partial(|\Gamma_{-1,-1}|^2)}{\partial\theta_s} \end{bmatrix} .$$

Brandt (1970) shows that when the elements of $\Delta\xi$ are determined by the maximum likelihood method of minimizing the expression (19), the covariance matrix C_ξ follows from the covariance matrix C_R defined in equation (9) through the equation

$$C_\xi = (M^T C_R^{-1} M)^{-1} \tag{22}$$

and the standard deviations of $\hat{H}_s$ and $\tilde{\theta}_s$ are the square roots of the diagonal elements.

(b) Swell of Narrow Frequency and Angular Spreads

We now consider a model in which the swell is assumed to have a finite but narrow Gaussian spread in wavenumber about a mean wavenumber, k_s, and is spread in angle according to a cardioid function of power s about a mean direction θ_s. Thus we define the model for this ocean waveheight spectrum as

$$S(\bar{k}) = S_w(\bar{k}) + \frac{H_s^2 \Gamma(s/2 + 1) \exp[-\frac{(k-k_s)^2}{2\sigma_k^2}] \cos^s\left(\frac{\theta-\theta_s}{2}\right)}{2^{3/2}\pi\sigma_k k \Gamma(s/2 + 1/2)} , \tag{23}$$

where $\Gamma(x)$ is the gamma function of argument x. The second term is the swell spectral model used, and it is normalized such that its energy is equal to the mean-square swell waveheight

$$\text{i.e.,} \qquad \int_0^\infty k\,dk \int_{-\pi}^{\pi} d\theta S_s(k,\theta) = H_s^2 \ .$$

This model could be applicable when the Doppler swell peaks are spread considerably beyond the half-power width of the first-order peak, but are still relatively narrow with respect to their frequency shift from this first-order echo. This could also be discovered because the large residual given by (19) indicates that the double-impulse-function model is not providing a good fit to the data.

The half-power wavenumber width, k_H, and the half-power angular spread, θ_H, of this model are simply related to the parameters σ_k and s by

$$k_H = 2\sqrt{2\ln 2}\,\sigma_k \qquad \text{and} \qquad \theta_H = 4\cos^{-1}[(1/2)^{1/s}]. \tag{24}$$

Often the nondirectional waveheight spectrum is measured with a buoy from a time series of the waveheight. In this case, a temporal spectrum is obtained rather than a spatial spectrum (i.e., vs. ω rather than k). For a narrow-band swell peak, the half-power widths of the two are related by means of the gravity-wave dispersion equation as

$$\omega_H \simeq \frac{g k_H}{2\omega_s} = \frac{\sqrt{2\ln 2}\, g\sigma_k}{\omega_s} , \tag{25}$$

where $\omega_s \equiv \sqrt{g k_s}$.

Upon substituting (23) into (7) for $R'_{m,m'}(\omega)$, we can arrive a closed-form solution for the swell-echo Doppler peaks by eliminating the double integral as follows: (i) we employ the delta-function constraint along with the relevant Jacobian to eliminate one of the integrals; (ii) we employ asymptotic techniques (i.e., the saddle-point approximation) to perform the remaining integration. The approximation involved in the second step produces the following result which is no more than 5% in error under the following "narrow-band" swell assumptions (for swell whose period is greater than 9s and radar frequencies greater than 10 MHz): $\omega_H/\omega_s < 0.3$ and $\theta_H < 60°$ ($s > 20$).

$$R'_{m,m'}(\omega) = \frac{2}{\omega_{DH}} \sqrt{\frac{\ln 2}{\pi}}\, R_{m,m'} \exp\{-4\ln 2(\omega-\omega_{m,m'})^2/\omega_{DH}^2\} , \tag{26}$$

where the peak positions $\omega_{m,m'}$ are identical to those of the impulse function model given by (12). The above expression is normalized so that the integral over ω yields $R_{m,m'}$; this quantity is given by

$$K_{m,m'} = \frac{2H_s^2}{A(s)} \int_{-\pi}^{\pi} |\Gamma_{m,m'}(\theta)|^2 \cos^s\left(\frac{\theta-\theta_s}{2}\right) d\theta \, , \tag{27a}$$

where

$$A(s) = \frac{2\sqrt{\pi}\,\Gamma(s/2 + 1/2)}{\Gamma(s/2 + 1)} = \int_{-\pi}^{\pi} \cos^s(\theta/2)\, d\theta \, . \tag{27b}$$

The square of the Doppler peak half-power width ω_{DH} is comprised of two terms, one resulting from the finite frequency spread of the waveheight spectrum and one produced by its finite angular spread (i.e., $\omega_{DH}^2 = \omega_{DHF}^2 + \omega_{DH\theta}^2$), where

$$\omega_{DHF} = \omega_H \left| m + m' \frac{\omega_s^3 + m\omega_B^2\omega_s\cos\theta_s}{[\omega_B^4 + 2m\omega_B^2\omega_s^2\cos\theta_s + \omega_s^4]^{3/4}} \right| , \tag{28a}$$

$$\omega_{DH\theta} = \frac{\theta_H}{2} \left| \frac{\omega_s^2\omega_B^2\sin\theta_s}{[\omega_B^4 + 2m\omega_B^2\omega_s^2\cos\theta_s + \omega_s^4]^{3/4}} \right| , \tag{28b}$$

where all ω's are in radians/s, and θ_H is in radians. All quantities in these equations have been defined previously in terms of the swell parameters.

There are several interesting implications of (26) for the four Doppler peaks and the subsequent equations that should be noted:

(i) The peak centroids or maxima occur at the same positions, $\omega_{m,m'}$, as those for the impulse function model. Hence one can solve for the swell period (i.e., k_s or ω_s) and direction, θ_s, by solving (14) and (15) for the spacings between the Doppler swell-peak centroids.

(ii) The frequency spread of the swell, ω_H, does not appreciably affect the energy distributions in the Doppler peaks, as represented by $K_{m,m'}$ in (27). The wave directional spread, however, does affect this quantity. In the limit of the impulse function in angle ($s \to \infty$), (27a) is seen to reduce identically to (13). In practice, having determined ω_s, θ_s (and knowing ω_B), one would numerically solve the integral (27a) (see the Appendix), and then employ maximum likelihood methods discussed in the preceding section (but with (27a)) to determine swell height, H_s, independent estimates of direction, θ_s, and estimates of the angle spreading parameter, s.

(iii) Of the two contributors to finite Doppler spectral peak width, the one produced by finite swell frequency spread dominates for nearly all practical swell situations, as seen by comparing (28a) with (28b). For example, for the "new" swell geometry described as typical in the Introduction, it can be seen that $(\omega_{DHF}/\omega_{DH\theta})^2 \approx 10$ implying that $\omega_{DH} \approx \omega_{DHF}$. Hence the width of the swell echo Doppler peaks, ω_{DH}, can be related directly to the frequency width spread of the swell spectrum, ω_H, directly through (28a).

(c) General Inversion for Swell

If the radar spectrum close to the first-order lines has a broad distribution, rather than displaying narrow peaks, we use a general inversion method which applies to both swell and wind-generated seas. The ocean-wave spectrum is expanded as a finite Fourier series in angle with coefficients which are functions of wavenumber:

$$S(\vec{k}) = \frac{1}{2} \sum_{i=-n}^{n} c_i(k)\, tf_i(\theta) \tag{29}$$

where

$$tf_i = \cos(i\theta), \; i \geq 0$$
$$= \sin(i\theta), \; i < 0 \; .$$

We assume the ocean-wave spectrum is constant within N wavenumber bands $k_j - \delta_j/2$ to $k_j + \delta_j/2$ and substitute (29) into the integral equation (7), which reduces to the linear form for c_i:

$$R_{m,m'}(\omega) = \sum_{i=-n}^{n} \sum_{j=1}^{N} K_{ijmm'}(\omega) c_i(k_j) \tag{30}$$

where the coefficients are given by

$$K_{ijmm'}(\omega) = \frac{1}{2\pi} \int_{\omega-\delta/2}^{\omega+\delta/2} \int_{k_j-\delta_j/2}^{k_j+\delta_j/2} \int_{0}^{2\pi} |\Gamma_{m,m'}|^2 \delta(\eta - m\sqrt{gk} - m'\sqrt{gk'}) cf_i(\theta) k\, d\theta\, dk\, d\omega \quad . \tag{31}$$

The integral in (31) is performed numerically using a method developed by Lipa (1977). The values of the coefficients $K_{ijmm'}$ for negative values of i are very small because the coupling coefficient is an even function of angle, and the frequency contours defined by the delta function constraint are almost circular. For $i < 0$, $cf_i(\theta)$ is an odd function of angle and it follows that the integral in (31) is nearly zero. It is therefore impossible to obtain the corresponding values of c_i for $i < 0$ by inverting (30). This constraint applies to a narrow-beam radar with a single look direction, it can be circumvented by using two or more look angles, or by use of the rotating broad beam antenna described by Barrick and Lipa (1979b). Values of the even Fourier coefficients may be obtained by inverting the linear equations (30) using methods developed by Phillips (1962) and Twomey (1963) and previously applied to radar sea-echo inversion by Lipa (1977) and Barrick and Lipa (1979b). The number of Fourier coefficients that can be derived with adequate accuracy from a given data set depends on the random fluctuations of the data and the magnification of error that occurs during the inversion process. We have inverted measured data to give the first three even Fourier coefficients $c_0(k)$, $c_1(k)$, $c_2(k)$ with adequate accuracy (Lipa et al., 1979).

The nondirectional coefficient $c_0(k)$ is merely the ocean waveheight nondirectional spectrum, since it follows from (29) that $c_0(k)$ is the integral of the directional wave spectrum over angle. The directional coefficients $c_1(k)$ and $c_2(k)$ do not admit to a quick interpretation in terms of important directional parameters such as the dominant or mean wave direction. Barrick and Lipa (1979b) show how they may be interpreted in terms of the cardioid model for the angular distribution which gives the dependence at a given wavelength as $\cos^s\left(\frac{\theta-\theta_s}{2}\right)$ where θ is the wave angle at a selected direction. This model requires that the following relations be obeyed:

$$\frac{c_1(k)}{c_0(k)} = \frac{2s}{(s+2)} \cos\theta_s \qquad ; \qquad \frac{c_2(k)}{c_0(k)} = \frac{2s(s-2)}{(s+2)(s+4)} \cos 2\theta_s \; . \tag{32}$$

Estimates of s and θ_s at each wavenumber may be determined by fitting these equations to the values of $c_i(k)$ obtained by inversion of the radar spectrum. The angle θ_s can then be interpreted as the mean wave direction. Since only the even coefficients c_0, c_1, and c_2 are derived from the above general inversion process, a unique determination of θ_s and s is not always possible without additional information.

4. CONCLUSIONS.

Our analysis reverses the impressions given by prior works on the extraction of sea-state parameters from HF radar echoes, in that it shows that long-wave directions can be determined from the second-order echo. The resulting techniques also produce more accurate estimates of swell height and period than those reported previously. The reason for this is that both the second-order Doppler peak positions and their amplitudes (the latter being proportional to the coupling coefficient) are shown to vary with swell direction. In other words, by recognizing that swell direction enters into the problem, more accurate estimates of height and period are obtainable, as well as the direction itself. A model suitable for narrow band/beam swell allows the direct extraction of the swell angular beamwidth and frequency bandwidth also. Expressions for the errors in these estimates are derived and presented.

The three techniques presented here — including the two closed-form models we derive — have been successfully tested against measured data from three separate experiments; they have been compared with several alternate sources of "surface truth" [Lipa et al., 1980]. The analysis of the vast amount of data there is extensive, and draws heavily upon the models and methods presented here. Our purpose in this paper is to derive and present simple models and methods that have been proven against data, along with a discussion of how and when to use them. We derive error estimate expressions for our results, and show how to use maximum likelihood principles specifically for the swell models employed here. Thus the paper is complete for understanding and should provide all the tools a user needs to extract swell from narrow-beam second-order HF sea-echo Doppler spectra.

A couple of limitations should be noted. First, a single narrow beam — while able to extract a surprising amount of swell directional information — is unable to distinguish from which side of the beam the swell originates (because of the appearance of only $\cos\theta_s$ everywhere, and not $\sin\theta_s$). Therefore unless one can scan the beam (or employ multiple beams), he is forced to guess or use additional information to remove this ambiguity. Secondly, it has been assumed that only a single swell from a single direction (but allowing for a spread in angle) is present at a given wave frequency for the two models derived. If two Doppler peaks can be resolved in the data representing different swell frequencies and directions, then each can be handled separately using the models. If, however swells from different directions are sufficiently close in wave frequency that they cannot be separately distinguished in the Doppler records, then the single-swell models are likely to give spurious results. Such a situation would be signalled by the χ^2-test applied to the residual, (19), indicating a poor fit of the model. Also, other internal consistencies normally observed (e.g., the agreement of directions obtained from peak positions with those obtained from maximum likelihood on the peak amplitudes) would fail. In this case, full inversion (i.e., the third method) must be employed.

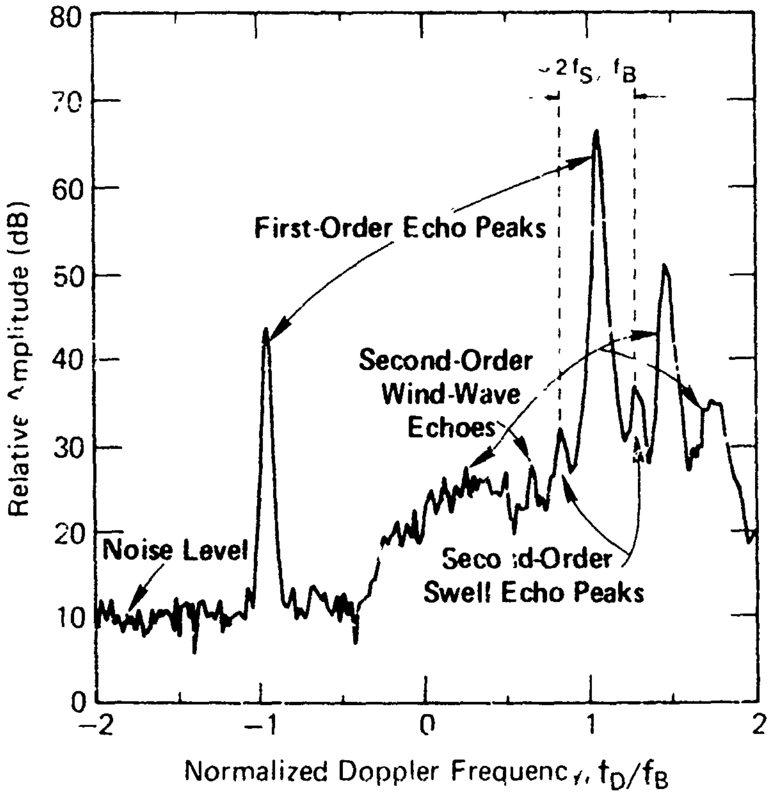


Figure 1. Narrow-beam sea-echo Doppler spectrum at 9.4 MHz measured at San Clemente Island on December 4, 1972, 1114-1156. Thirteen 204.8-second spectra were averaged to produce this result, illustrating the first-order Bragg peaks, second-order swell peaks, and second-order wind-wave peaks. Independent buoy observation showed 14-second swell of .76 m significant height, along with wind waves having 1.83 m significant height.

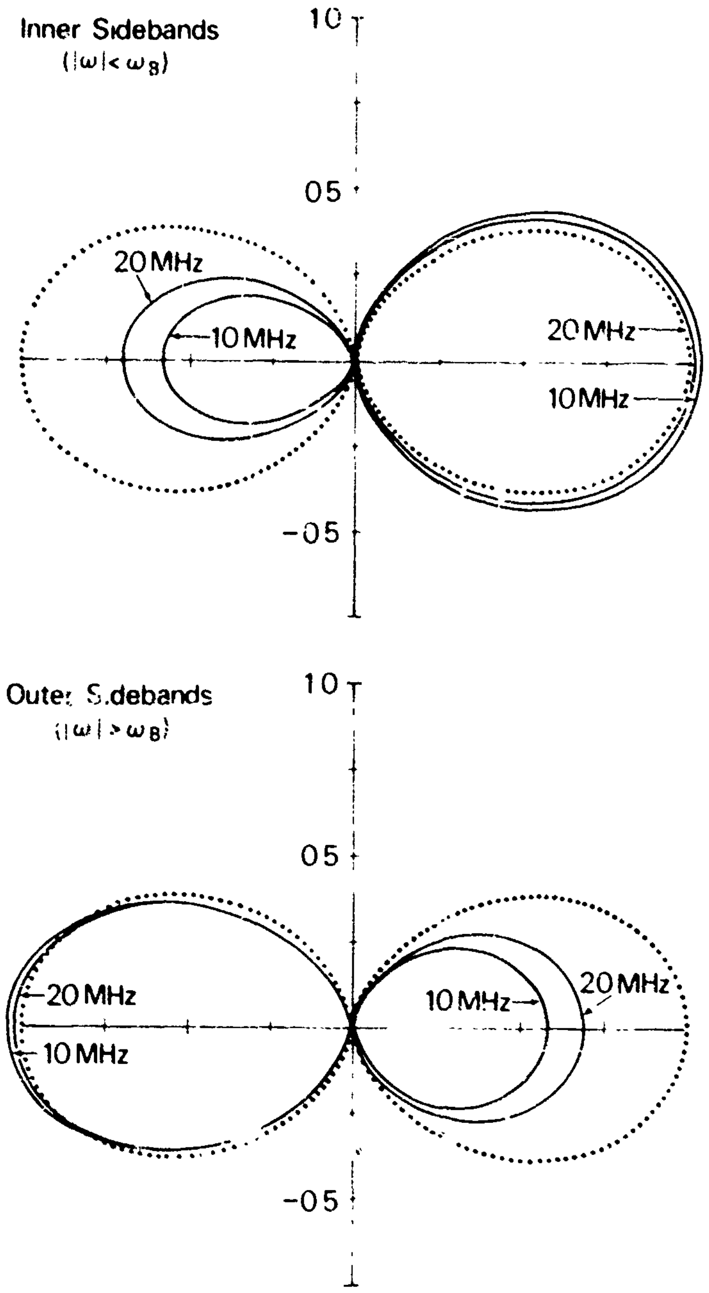


Figure 2. Polar plots of normalized coupling coefficient (divided by k_0^2), $|\Gamma_{m,m'}|^2$, for 13-second swell at two radar frequencies. Plots are vs. θ_s, the angle between the radar look direction and the swell propagation direction. Dotted curves give $\cos^2\theta_s$ function for comparison. For positive Doppler sidebands, θ_s in above plots is measured from +x axis; for negative sidebands, from -x axis.

5. REFERENCES

Barrick, D. E., Dependence of second-order sidebands in HF sea echo upon sea state, 1971 IEEE G-AP International Symposium Digest, Sept. 21-24, Los Angeles, Calif., 194-197 (1971).

Barrick, D. E., Remote sensing of sea state by radar, in Remote Sensing of the Troposphere, ed. V. E. Derr, U.S. Government Printing Office, Washington, D.C. (1972).

Barrick, D. E., The ocean waveheight nondirectional spectrum from inversion of the HF sea-echo Doppler spectrum, Remote Sensing Environ., 6, 201-227 (1977a).

Barrick, D. E., Extraction of wave parameters from measured HF sea-echo Doppler spectra, Radio Sci., 12, 415-424 (1978).

Barrick, D. E., Parameter extraction accuracy from random HF sea-echo Doppler spectral samples, in press (1979).

Barrick, D. E., and Lipa, B. J., Ocean surface features observed by HF coastal ground-wave radars: A progress review, *Ocean Wave Climate*, Plenum, New York, 129-152 (1979a).

Barrick, D. E., and Lipa, B. J., A compact transportable HF radar system for directional coastal wave-field measurements, *Ocean Wave Climate*, Plenum, New York, 153-201 (1979b).

Barrick, D. E. and Snider, J. B., The statistics of HF sea-echo Doppler spectra, IEEE Transactions on Antennas and Propagation, AP-25, 19-28 (1977).

Brandt, S., *Statistical and Computational Methods in Data Analysis*, North Holland, Amsterdam (1970).

Crombie, D. D., Doppler spectrum of sea echo at 13.56 Mc/s, *Nature*, 175: 681-682 (1955).

Hasselmann, K., Determination of ocean wave spectra from Doppler radio return from the sea surface, *Nature*, 229 16-17 (1971).

Hasselmann, K., D. B. Ross, P. Müller, and W. Sell, A parametric wave prediction model, J. *Phys. Oceanography*, 6, 200-228 (1976).

Kinsman, B., *Wind Waves*, Prentice-Hall, Englewood Cliffs, N.J., 676 pp (1965).

Lipa, B. J., Derivation of directional ocean wave spectra by integral inversion of second-order radar echoes, *Radio Sci.*, 12, 425-436 (1977).

Lipa, B. J., Inversion of second-order radar echoes from the sea, J. *Geophys.* Res., 83, 959-962 (1978).

Lipa, B. J., Barrick, D. E., Maresca, J. W., Teague, C. C., Ocean swell parameters from HF radar sea echo, II experiment, in preparation (1979)

Lipa, B. J. and D. E. Barrick, Ocean swell parameters from narrow-beam HF radar sea echo, including FORTRAN codes for the coupling coefficient, NOAA Tech. Memorandum, 1979.

Long, A. E. and Trizna, D. B., Mapping of North Atlantic winds by HF radar sea backscatter interpretation, IEEE Transactions on Antennas and Propagation, AP-21, 680-685 (1973).

Maresca, J. W. Jr. and Georges, T. M., HF skywave radar measurement of the ocean wave spectrum, accepted for publication in *J. Geophys. Res.*, (1979).

Phillips, D. L., A technique for the numerical solution of certain integral equations of the first kind, J. Ass. *Comput.* Mach., 9, 84-97 (1962).

Stewart, R. H. and Barnum, J. R., Radio measurements of oceanic winds at long ranges: an evaluation, *Radio Sci.*, 10, 853-857 (1975).

Twomey, S., On the numerical solution of Fredholm integral equations of the first kind by inversion of the linear system produced by quadrature, J. *Ass. Comput.* Mach., 10, 97-101 (1963).

Wait, J. R., Theory of HF ground wave backscatter from sea waves, J. Geophys. Res., 71, 4839-4842 (1966).

APPENDIX

In addition to the positions of the four Doppler swell echo peaks, their amplitudes and energy depend upon swell period and direction. This is most obvious from the impulse function model of (13), and also for the narrow beam/band model from (27). It is the coupling coefficient that produces these directional effects on the amplitudes. Conversely, if one were to treat the coupling coefficient as a constant, independent of swell direction and period as suggested in previous works, measured data could not be explained by the resulting model, and the most important swell parameter, namely height, could not be accurately extracted.

To illustrate the strong directional dependence of the coupling coefficient, we show plots of this quantity, $|\Gamma_{m,m'}|^2$, vs. θ_s in Fig. 2 for a typical HF/swell example: 13 s swell, at both 10 and 20 MHz. These plots are normalized to k_0^2, the radar wavenumber. Shown also is the $\cos^2\theta_s$ function for comparison.

The amplitudes of the four spectral peaks for the impulse-function model of swell are directly proportional to this coupling coefficient, from (13). One can clearly see therefore that these amplitudes will vary greatly, even disappearing when the swell propagates perpendicular to the radar beam. This strong variation is observed in measured data and is responsible for the success of swell parameter extraction.

The reason for the asymmetric front-to-back shape lies with the existence of the first term (the electromagnetic interaction effect) of the coupling coefficient, (5). Even though the hydrodynamic interaction term (the second term) is larger, one clearly cannot neglect the electromagnetic term. For if the first term were zero, the coupling coefficient would have the $\cos^2\theta_s$ form and there would be no front-back asymmetry. In that event, the odd directional cosine coefficients c_1, c_3, etc contained in the waveheight spectrum (29) would be indeterminable. In other words, one would not be able to distinguish swell at θ_s from swell at $180°-\theta_s$. As it is, a single narrow beam can extract c_0, c_1, c_2, c_3, and possibly higher odd coefficients if the data is not too noisy.

The physical mechanism underlying this asymmetry is of interest. In one direction (that with the larger lobe), the radical in the first term of (5) is real, i.e., $\sqrt{\vec{k}\cdot\vec{k}'} = \sqrt{|\vec{k}\cdot\vec{k}'|}$. This means that the intermediate electromagnetic wave interacting with the two ocean waves of wavevectors k and $\vec{k}'$ (one of which is the swell wavevector) is "propagating", and the electromagnetic and hydrodynamic terms are then seen to add in quadrature. In the other direction with the smaller lobe, however, the argument of the radical is negative and hence the radical is imaginary, i.e., $\sqrt{\vec{k}\cdot\vec{k}'} = \sqrt{-|\vec{k}\cdot\vec{k}'|} = i\sqrt{|\vec{k}\cdot\vec{k}'|}$. In this case the intermediate electromagnetic wave is not propagating, but "evanescent"; then the electromagnetic and hydrodynamic terms are no longer in quadrature, but are in phase and subtracting. In the former case, an approximate factor for the coupling coefficient at the larger lobe maximum is $1 + (\frac{\omega_s}{\omega_B})^2$, while in the latter the appropriate factor at the smaller lobe maximum is $[1 - \frac{\omega_s}{\omega_B}]^2$. At the point where $\vec{k}\cdot\vec{k}' = 0$, the intermediate electromagnetic wave is bound to the surface, propagating along this boundary with no energy scattered upwards; for this singular situation the ocean waves with $\vec{k}$ and $\vec{k}'$ are propagating at right angles to each other, which has given rise to the terminology "corner reflector" effect. The bound intermediate wave for this case is known in electromagnetic theory as the "Wood's anomaly."

Looking at Fig. 2, it is tempting to suggest that the upper and lower plots are identical if flipped around. One might go further and say that the coupling coefficient can be modeled accurately by a $\cos^2\theta_s$ pattern whose amplitude factor in one direction is $[1 + (\frac{\omega_s}{\omega_B})^2]$ and in the other is $[1 - \frac{\omega_s}{\omega_B}]^2$; this would result in a closed-form solution to (27). However, this simplification would produce identical sets of Doppler peaks surrounding each of the two first-order Bragg peaks, and this has been observed in measured data not to be the case for $|\cos\theta_s| < 0.5$ [Lipa et al., 1980]. Closer examination of the coupling coefficient, as plotted in Fig. 2, shows that for example, at $|\cos\theta_s| = 0.5$ (i.e., $\theta_s = 60°$ and $120°$) the values of the smaller lobes differ as much as 25% from each other. This is sufficient to produce significant error in the swell parameter extraction process. Therefore, one should employ the exact expression for the coupling coefficient in (13), (27) and in the full integral inversion process. The coupling coefficient can be calculated quickly on the computer, and Lipa and Barrick [1979] provide a simple FORTRAN code that has been successfully tested for calculating this function.

SUMMARY OF SESSION VI

TUNNEL TRANSMISSION AND SURFACE WAVES

The concept that electromagnetic waves can be guided by tunnels is not surprising. However, in an air filled tunnel the frequency must be greater than some cut-off value that typically is of the order of 50 MHz. When the tunnel contains axial conductors, the situation is changed dramatically, and there is no cut off as such, although wave guide type modes also propagate at the higher frequencies.

The use of shielded cables with intentional leakage in the sheath was seen to play a major role in optimizing the communication range and also allowing for remote access if the cable to transceiver distances are not too great. This leaky feeder concept is finding a real place in mining communication and other limited access corridors. We were fortunate in having the majority of the leading workers in the field as participants. Actually, much of the discussion took place outside of the lecture hall.

A special bonus was a late paper by **Dr Antonio Fernandes**, of the Instituto Superior Technico in Lisbon. The paper deals with a comparison of the characteristics of slotted and loose braid coaxial cables.

This special session was augmented by several papers (e.g. 39 and 41) that did not deal explicitly with leaky feeders, but they were also concerned with propagation in non ionized regions of the earth. These papers dealt with HF wave transmission through the earth and along the earth's surface.

PRINCIPLES OF H.F. COMMUNICATION IN TUNNELS USING OPEN TRANSMISSION LINES AND LEAKY CABLES

James R. Wait, David A. Hill and David B. Seidel
Cooperative Institute for Research in Environmental Sciences
University of Colorado
Boulder, Colorado 80309

SUMMARY

A rather ingenious communication technique has been developed for use in coal mines. The idea is to use a two-wire transmission line or equivalent loosely braided coaxial cable that can be suspended from the upper wall (Wait, July, 1975). A transmitter placed in the vicinity of the line excites a strongly unbalanced mode that normally would propagate like a coaxial or TEM mode with relatively high attenuation. The key step in the Belgian system, and the closely related French system, is to convert this to a balanced mode that is much less attenuated because the return current flows mostly in the second wire rather than through the surrounding rock. The physical basis of this system has many similarities with the leaky feeder devices used in British coal mines when the mode conversion is inadvertent. We have examined a number of relevant problems of this type. Brief summaries are given in the following with appropriate references to where details can be found.

The braided coaxial cable was studied as a communication scheme in a mine tunnel. A simplified rectangular waveguide model was adopted for the tunnel, and the shield of the cable was assumed to behave as a single inductive transfer impedance. Specific results on the attenuation of the monofilar and bifilar (or coaxial) modes of propagation, taking into account the possible existence of a thin lossy film on the cable, were obtained (Mahmoud and Wait, Jan., 1976). In order to estimate the maximum possible range of communication, we considered the coupling factors of these modes to transmitting and receiving dipoles inside the tunnel, and we presented results on these factors for various cable parameters and over a wide range of frequencies (Mahmoud and Wait, Jan., 1976).

The transmission loss for a leaky coaxial cable communication system in a circular tunnel was calculated for an idealized model. The transmitting and receiving antennas were electric dipoles that could be located anywhere within the tunnel. For typical cable parameters, the optimum frequency was found to be in the range from 2 to 20 MHz. Calculations revealed that the use of a sparse braid and a high velocity cable will strongly improve the cable-dipole coupling with only a small penalty of a higher attenuation rate (Hill and Wait, April, 1976).

The previous development was extended to permit attenuation calculations when a braided cable is located close to a tunnel wall. This is an important case in mine communications utilizing leaky feeders. We found that the attenuation rate for the bifilar mode was hardly affected at all by the finite conductivity of the wall. On the other hand, the monofilar mode suffers a very high attenuation when the cable approaches the wall (Hill and Wait, July 1976). In a subsequent analysis, we examined the effect of the dependence of the braid transfer impedance on the propagation constant for a coaxial cable located in a circular tunnel (Wait and Hill, Jan., 1977).

The expression for the gap admittance of a coaxial cable within a circular tunnel was derived using a quasi-static method. Coupling between the TEM mode within the cable and the monofilar mode within the tunnel was computed, and 25% of the cable power was typically transferred to the tunnel from a single cable gap. The results have application to the use of leaky coaxial cables for communication in mine tunnels (Hill and Wait, Oct., 1975).

An idealized model of a trolley wire communication system was considered. To facilitate the analysis, the tunnel cross section was taken to be semicircular and the surrounding rock medium was homogeneous with a finite conductivity. The metallic rails or other conductors on the flat floor of the tunnel were represented by a plane metallic reflecting surface of infinite extent. The trolley wire and associated feeder line were assumed to be equivalent to a single cylindrical conductor that was parallel to the axis of the tunnel but located anywhere within the cross section. The appropriate modal equation was then solved for the propagation constant of the dominant mode. It was shown that the attenuation rate increases significantly as the trolley wire approaches the tunnel wall. However, even when the unloaded trolley wire was within 20 cm (8 in) of the wall, the attenuation rate at 200 kHz was still less than 2 dB per kilometer (Hill and Wait, Nov., 1976).

In a related study, the general theory for transmission in a circular tunnel containing a thin axial conductor was employed to calculate the attenuation rate of the propagating mode. The remarkable property was that the attenuation rate was approximately proportional to frequency and it did not depend critically on the wall conductivity for typical conditions (Wait and Hill, June, 1976).

A general analysis was given for the transmission of electromagnetic waves along a circular tunnel in the presence of axial conductors. One of these conductors was an idealization of a trolley wire with its bonded feeder line that may be located anywhere in the tunnel. The other was a metal rail that was located within the homogeneous rock medium at a finite distance from the tunnel wall. A mode equation for the propagation modes was obtained that was used to obtain numerical results for the attenuation rate in the frequency range 50 to 800 kHz. It was shown that the attenuation rate is only weakly dependent on the burial depth of the rail conductor, at least in the range up to 50 cm. Also, it was found that the rail conductor for the model adopted carries an appreciable fraction of the total return current when typical rock conductivities are assumed. A thin layer of insulation on the rail appears to have a negligible effect. In fact, there was little change even if the rail was within the tunnel and just above the tunnel floor (Wait and Hill, March, 1977).

The previous modal analysis for a circular tunnel containing a thin wire was shown to be greatly simplified at low frequencies provided the wall conductivity was high. The explicit formula for the propagation constant, so obtained when specialized to a perfectly conducting wire, was in agreement with a quasi-static analysis (Wait, May, 1977).

Propagation along a thin wire located in an air-filled tunnel was considered in the limiting case where the tunnel radius becomes indefinitely large but the distance of the wire to the tunnel wall remained finite. It was found that the results were consistent with the corresponding analysis for a thin wire located over a plane homogeneous earth (Wait, Nov., 1976).

We analyzed the fields excited by a dipole in a circular tunnel that contains a dielectric coated axial conductor. The integral form of the solution is generally valid provided the outer radius of the dielectric coated conductor is small compared with its distance to the tunnel wall. The propagation mode that behaves as an axial surface wave at high frequencies was examined for a frequency range from 1 MHz to 1 GHz. The attenuation reached a maximum value at frequencies of the order of 100 MHz where the absorption loss by the tunnel wall was most significant. In the range from 1 to about 10 MHz, the attenuation rate was approximately proportional to the frequency and was only weakly dependent on the conductivity of the region external to the tunnel walls. The results were consistent with earlier investigations for surface wave lines located over a homogeneous half space (Wait and Hill, Dec., 1976 and April, 1977).

The electromagnetic fields of dipole emitters located inside an empty tunnel or cylindrical cavity were first formulated. Both electric and magnetic types were considered. The results were used to deduce an expression for the change of self-impedance due to the presence of the tunnel walls. The input resistance R of the electric dipole was approximately a constant for frequencies less than about 25 MHz for a typical tunnel radius of 2m. In contrast, for the same conditions, R for the magnetic dipole or small loop varied approximately as the square of the frequency. The theory was extended to allow for the presence of an axial conductor or cable within the tunnel. The presence of the cable increases both R for a transversely oriented electric dipole and a circumferentially oriented magnetic dipole but R decreased for an axial electric dipole. Not surprisingly, the axial magnetic dipole is only weakly influenced by the presence of the cable (Hill and Wait, Dec., 1976, April, 1978).

Radio frequency transmission in a semi-circular tunnel containing a coaxial cable was again considered (Seidel and Wait, July, 1978). The general formulation accounted for both the ohmic losses in the tunnel wall and a thin lossy film layer on the outer surface of the dielectric jacket of the cable. Using the quasi-static approximation (Wait, May, 1977) it was found that the propagation constants of the low-frequency transmission line modes could be obtained from the solution of a cubic equation. However, for the special case when the conductivity thickness product of the lossy film layer vanishes, this cubic equation reduced to a quadratic. The spatially dispersive from of the braid transfer impedance (Wait and Hill, Jan., 1977) was also accounted for. It was shown that the quasi-static theory is well justified for frequencies as high as 100 MHz for typical tunnel geometries. Finally, special characteristic impedances were derived for the various modes of the equivalent multi-conductor transmission line.

Using a further application of the quasi-static theory (Wait, May, 1977), it is shown that the effect of ohmic losses in the wall, for two axial conductors (Hill and Wait, Sept., 1977), can be represented as series impedances per unit length in equivalent coupled transmission line circuits. This opens the way for a systematic calculation of the mode conversion phenomena that occur when one of the axial conductors has a discrete shunt load to the ground plane. Explicit formulas are obtained for the modal conversion coefficients. These are used to discuss the effect of a shunt load on the trolley wire on the performance of a radio frequency communication system (Wait and Hill, March, 1978). It is confirmed that the unloaded dedicated communication line provides for a low loss mode that is hardly affected by the shunt load on the adjacent trolley wire in the tunnel.

The role of controlled mode conversion in leaky feeder mine-communication systems was also considered (Seidel and Wait, Sept., 1978). The controlled interchange of energy between the propagation modes was used to optimize the communication range and allow for coupling to and from portable hand-held antennas. In particular, we showed that, for this model, the optimum lengths of the leaky cable sections or mode converters are not predicted adequately by the approximate coupled mode theory. However, it appears that simple modifications of this latter approach can be made that achieve compatibility (private communications from P. Delogne).

Using a somewhat different approach (Wait, July, 1978), we then considered guided electromagnetic waves in a periodically non-uniform tunnel. A boundary value analysis was outlined for the transmission in a circular tunnel that contained a thin axial conductor where the latter was allowed to vary in a periodic fashion throughout its length. Simplifications of the formal mode equation were achieved by invoking quasi-static conditions used as in previous studies (Wait, May, 1977). It is shown that the axial non-uniformity can be represented approximately as a modification of the series impedance of the equivalent transmission line of the composite tunnel structure.

Finally, we considered the operating bandwidth of a leaky coaxial cable in a circular tunnel (Hill and Wait, April 11, 1978). We examined both pulse transmission and the bandwidth limitations of the channel. The phase and group velocities of the bifilar mode were found to be nearly independent of frequency between 1 MHz and 100 MHz, and the channel bandwidth was found to be quite large. The effect of mode excitation was also examined by treating the mutual impedance between a pair of short electric dipoles.

REFERENCES

HILL, D.A. and WAIT, J.R., October 1975, "Electromagnetic fields of a coaxial cable with an interrupted shield located in a circular tunnel," Journal of Applied Physics, Vol. 46 No. 10, 4352-4355.

HILL, D.A. and WAIT, J.R., April 1976, 'Calculated transmission loss for a leaky feeder communication system in a circular tunnel", Radio Science, Vol. 11, No. 4, 315-321.

HILL, D.A. and WAIT, J.R., July 1976, "Propagation along a braided coaxial cable located close to a tunnel wall," IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-24, No. 7, 476-480.

HILL, D.A and WAIT, J.R., November 1976, Analysis of radio frequency transmission along a trolley wire in a mine tunnel," IEEE Transactions on Electromagnetic Compatibility, Vol. EMC-18, No. 4, 170-174.

HILL, D.A. and WAIT, J.R., Sept. 1977, "Analysis of radio frequency transmission in a semi-circular mine tunnel containing two axial conductors," IEEE Transac

HILL, D.A. and WAIT, J.R., April 1978, "The impedance of dipoles in a circular tunnel with an axial conductor," Accepted by the IEEE Transactions on Geoscience Electronics, Vol. GE-16, No. 2, 118-126.

HILL, D.A., and WAIT, J.R., Nov. 1978, "Bandwidth of a leaky coaxial cable in a circular tunnel," IEEE Transactions on Communications, Vol. COM-26, No. 11.

MAHMOUD, S.F. and WAIT, J.R., January 1976, "Calculated channel characteristics of a braided coaxial cable in a mine tunnel," IEEE Transactions on Communications, Vol. COM-24, No. 1, 82-87.

SEIDEL, D.B., and WAIT, J.R., July 1978, "Transmission modes in a braided coaxial cable and coupling to a tunnel environment," IEEE Transactions on Microwave Theory and Technique, Vol. MTT-26.

SEIDEL, D.B., and WAIT, J.R., Sept., 1978, "Role of controlled mode conversion in leaky feeder mine-communication systems," IEEE Transactions on Antennas and Propagation, Vol. AP-26, No. 5, 690-694.

WAIT, J.R., July 1975, "Theory of EM wave propagation through tunnels," Radio Science, Vol. 10, No. 7.

WAIT, J.R., Nov. 1976, "The eccentrically located wire in a cylindrical cavity in a conducting medium and the limit of a planar boundary," Radio Science, Vol. 11, No. 11, 897-899.

WAIT, J.R., May 1977, "Quasi-static limit for the propagating mode along a thin wire in a circular tunnel," IEEE Transactions on Antennas and Propagation, Vol. 25, No. 3, 441-443.

WAIT, J.R., July, 1978, "Guided electromagnetic waves in a periodically non-uniform tunnel," IEEE Transactions on Antennas and Propagation, Vol. AP-26, No. 4, 623-625.

WAIT, J.R. and HILL, D.A., June 1976, "Low-frequency radio transmission in a circular tunnel containing a wire conductor near the wall," Electronics Letters, Vol. 12, No. 13, 346-347.

WAIT, J.R. and HILL, D.A., Dec. 1976, "Attenuation on a surface wave G-line suspended within a circular tunnel," Journal of Applied Physics, Vol. 47, No. 12, 5472-5473.

WAIT, J.R. and HILL, D.A., December 1976, "Impedance of an electric dipole located in a cylindrical cavity in a dissipative medium," Applied Physics, Vol. 11, No. 4, 351-356.

WAIT, J.R. and HILL, D.A., January 1977, "Influence of spatial dispersion of the shield transfer impedance of a braided coaxial cable," IEEE Transactions on Microwave Theory and Techniques, Vol. 25, 72-74.

WAIT, J.R. and HILL, D.A., March 1977, "Radio frequency transmission via a trolley wire in a tunnel with a rail return," IEEE Transmission on Antennas and Propagation, Vol. AP-25, No. 2, 248-253.

WAIT, J.R. and HILL, D.A., April 1977, "Electromagnetic fields of a dipole source in a circular tunnel containing a surface wave line," International Journal of Electronics, Vol. 42, No. 4, 377-391.

WAIT, J.R. and HILL, D.A., March 1978, "Analysis of the dedicated communication line in a wire tunnel for a shunt-loaded trolley wire," IEEE Transactions on Communications, Vol. COM-26, No. 3, 355-361.

NOTE: The bibliographies in the above publications contain extensive references to the work of other investigators such as: B.A. Austin, J.C. Beal, D.C. Chang, J. Chiba, Q.V. Davis, P. Degauque, P. Delogne, L. Deryck, A.G. Emslie, R. Gabillard, R.W. Haining, E.F. Kuester, R.L. Lagace, R. Liegeois, N.A. Mackay, D.J.R. Martin, R.G. Olsen and R.L. Spencer. It was not our intention here to summarise all related work.

DISCUSSION

L.B. Felsen, US

To what extent is the concept of lossy anisotropic surface (sheath) impedance justified to represent the actual more complicated wire helix configuration? Are these quantitative comparisons concerning the range of validity of the surface impedance concept?

To what extent do conclusions obtained in this regard from planar sheaths or grids carry over to the cylindrical sheath at the long wavelength (in terms of sheath radius) under consideration?

Author's Reply

The surface transfer impedance description of the sheath, of course, is an approximation to relate the mean current in the sheath to the mean longitudinal field. As shown by K C Casey and others, the transfer impedance is spatially dispersive, so strictly speaking the impedance to use will depend on the mode type. This subject has been discussed by Wait and Hill (1977).

The answer to the second question is that the planar and the cylindrical configurations should not have any quantitative similarity, but surely the situations are analogous.

D.J. Gale, UK

There is a variation to the approach of considering pitch angles and numbers of helical wire wraps. For example, by bunching wires, one can change the extent of metal cover (or open area per unit length), thus introducing a new variable and one that can be encountered in practice.

Author's Reply

Metal cover (usually called "optical coverage") is not always a good indication for shielding effectiveness of wound or braided cables. For example, if the pitch angle is increased, the optical coverage increases, but the surface transfer impedance also increases thus indicating reduced shielding.

D.J.Gale, UK

The surface transfer impedance concept is certainly appropriate, at least in the HF band. Practical problems can be encountered at higher frequencies in the measurement of this characteristic.

Author's Reply

Our theory also indicates that the surface transfer impedance characterizations will eventually become inappropriate if the frequency is increased to too high a value.

MODE CONVERTERS FOR TUNNELS TRANSMISSION

Prof. P. DELOGNE
Laboratoire de Télécommunications et d'Hyperfréquences
Université Catholique de Louvain
Bâtiment Maxwell 1348 LOUVAIN-LA-NEUVE (BELGIUM)

ABSTRACT

Leaky feeder techniques are used to guide electromagnetic waves in tunnels. These systems are based on the coupling of the coaxial or bifilar mode with the monofilar mode. In contrast with the principle of continuous leaky feeders one can use discrete mode converters spaced along a non leaky cable. After an analysis of the general properties of mode converters the paper gives a detailed description of various types : selective and wideband annular slot converters for coaxial cables, leaky sections, selective and wideband converters for two wire lines.

1. INTRODUCTION

The basic problems associated with the propagation of electromagnetic waves in tunnels are now well understood (DELOGNE, 1973 and 1976) and they have been solved numerically for a number of tunnel geometries (MAHMOUD, 1974 ; SEIDEL, 1978 ; WAIT, 1974). Most developed transmission systems use a coaxial cable or a two wire line which is strung parallel to the tunnel axis. Such a transmission line can guide two TEM-like modes. The normal mode of the transmission line is called the coaxial or bifilar mode : it is barely or not at all influenced by the tunnel environment. The monofilar mode is the one in which the current flows along the two wires of the transmission line and returns along the tunnel wall. This is the useful mode for the coupling to the mobile transceivers [illegible] in the tunnel, for its field lines extend in the whole tunnel space.

The use of the ground as a return conductor for the monofilar mode implies a high specific attenuation for this mode. The fundamental principle involved in all systems is to create energy exchanges between the two modes, a long range being hopefully achieved by use of the much lower specific attenuation of the coaxial or bifilar mode. In leaky coaxial cables, the coupling between the modes occurs continuously across the apertures of the imperfect shield. Although coupled line theory provides an accurate model of this phenomenon, a more correct approach is the decomposition into two eigenmodes.

An alternative solution to the continuous leaky feeder is to use a well shielded coaxial cable in which mode converters are inserted at discrete places. The principle can also be used with a two wire line. An obvious requirement for the mode converters is to provide a good impedance match and a low insertion loss for the coaxial or bifilar mode. Ideally the loss should be due to the mode conversion only. This method provides an excellent flexibility in the design of a system because the converter parameters and spacing can be varied along the path in function of the tunnel cross section, acceptable cable location, distance to the base station, etc.

This paper is concerned with the description of several types of mode converters which are now widely used in mines and in road tunnels. The applications will be restricted to frequencies below the tunnel cut-off. At higher frequencies the waveguide modes can play an important role. It has been shown (DELOGNE, 1976) that the same mode converters are also very useful devices at these frequencies, since they act as efficient directive antennas radiating a part of the power carried by the coaxial or bifilar line.

2. GENERAL PROPERTIES OF MODE CONVERTERS

Considering the existence of two modes on each side, a mode converter should be considered as a four port device. We will use subscripts 1 and 2 to denote the two sides of the converter and subscripts c and m to indicate the coaxial (or bifilar) mode and the monofilar mode, respectively. We call the normalized waves A and B which are flowing into and emerging from the device, respectively. By reciprocity the scattering matrix which relates them should be symmetrical. It has thus the following most general form :

$$\begin{vmatrix} B_{c1} \\ B_{m1} \\ B_{c2} \\ B_{m2} \end{vmatrix} = \begin{vmatrix} R_{c1} & R_{cm1} & T_c & T_{c1m2} \\ R_{cm1} & R_{m1} & T_{m1c2} & T_m \\ T_c & T_{m1c2} & R_{c2} & R_{cm2} \\ T_{c1m2} & T_m & R_{cm2} & R_{m2} \end{vmatrix} \begin{vmatrix} A_{c1} \\ A_{m1} \\ A_{c2} \\ A_{m2} \end{vmatrix} \qquad (1)$$

The meaning of the matrix elements is as follows and is illustrated on fig. 1 :

R_{c1} & R_{c2} : reflection factors of the coaxial mode at sides 1 and 2

R_{m1} & R_{m2} : reflection factors of the monofilar mode at sides 1 and 2

K_{cm1}, K_{cm2} : mode conversion factors in reflection at sides 1 and 2

T_{c1m2} : mode conversion factor in transmission between the coaxial mode at side 1 and the monofilar mode at side 2

T_{c2m1} : same, between the coaxial mode at side 2 and the monofilar mode at side 1.

The mode converters are most frequently lossless devices and the scattering matrix must be self adjoint. Most types have moreover a transverse symmetry plane for the monofilar mode and either a symmetry or an asymmetry plane for the coaxial mode. When combined, these two properties yield :

$$|K_{c1}| = |K_{c2}| = K_c$$

$$|K_{m1}| = |K_{m2}| = K_m$$

$$|K_{cm1}| = |K_{cm2}| = K_{cm} \qquad (2)$$

$$|T_{c1m2}| = |T_{c2m1}| = T_{cm}$$

$$K_c^2 + T_c^2 + K_{cm}^2 + T_{cm}^2 = 1 \qquad (3)$$

$$K_m^2 + T_m^2 + K_{cm}^2 + T_{cm}^2 = 1 \qquad (4)$$

and some phase relationships. The transverse symmetry plane does not necessarily yield the equality of the mode conversion factors in reflection and in transmission, nor the equality of the reflection factors or of the transmission factors for the coaxial (bifilar) and monofilar modes.

For obvious practical reasons, mode converters should have no connection to the tunnel wall. If the mode converter is made with lumped elements, this implies $I_{m1} + I_{m2} = 0$, or

$$A_{m1} \quad B_{m1} \quad A_{m2} \quad B_{m2} \qquad (5)$$

As these conditions must be satisfied for all combinations of incident waves, one has

$$K_{cm1} = -T_{c1m2} \qquad (6)$$

$$K_{cm2} = -T_{c2m1} \qquad (7)$$

$$K_{m1} + T_m = K_{m2} + T_m = 1 \qquad (8)$$

In particular it is seen that a coaxial wave incident on the converter always excites two equal monofilar waves. From this we conclude that a directional mode conversion can only be realized by mode converters with a non-zero electrical length.

Finally if the mode converter is lossless, has a transverse symmetry plane, no ground connection and a zero electrical length, it is described by reflection factors K_c, K_m, transmission factors T_c, T_m and a unique mode conversion factor S with the property

$$K_c^2 + T_c^2 + 2S^2 = K_m^2 + T_m^2 + 2S^2 = 1 \qquad (9)$$

3. ANNULAR SLOT CONVERTERS

3.1. Basic theory and quasi static approximation

An annular slot realizing a complete interruption of the outer conductor of a coaxial cable has been proposed as the main ingredient of a mode converter (DELOGNE, 1972). A detailed analysis of the working of the slotted cable in free space was published (DELOGNE, 1971). Later on, the theoretical model was refined to take account of the cable dielectric jacket (WAIT, 1975) and of the presence of the tunnel wall (HILL, 1976).

A simplified quasi static analysis of an annular slot in the outer conductor of a coaxial cable can be obtained by considering the cable and the tunnel as two transmission lines having a common conductor. The latter is interrupted over a short length. The problem thus reduces to an elementary calculation suggested by Figs. 2.a and 2.b, where Z_m and Z_c are the characteristic impedances of the monofilar and coaxial modes respectively. The only difficult point in this respect is to estimate the value of Z_m. Fortunately it appears that this quantity varies very slowly with the electrical parameters of the ground, with the distance of the cable to the wall and with the frequency. A comparison with the exact electromagnetic solutions available between 1 and 50 MHz shows that calculations based on the value $Z_m = 300$ ohm never yield an error above 1 dB on the mode conversion factor.

A naked slot does not provide a good impedance match and a low insertion loss for the coaxial mode. Indeed the external load impedance "seen" by the slot, $2Z_m$, is rather large and most power flowing inside the coaxial cable is reflected back inside the cable. It is thus necessary to add some lumped circuit elements in order to improve the impedance match and to lower the insertion loss.

3.1. Reactive mode converters

In this type of mode converter the slot impedance is deliberately lowered by connecting a reactance across it. The residual reactive effect can be compensated at the design frequency by inserting a dual reactance in the inner conductor. Fig. 3.a, 3.b show two such designs. Equations (1) to (9) all apply to these converters and the values of K_c, T_c, K_m, T_m and S can easily be obtained from the circuit equations, although the calculations are somewhat tedious. The results for $|K_c|$, $|T_c|$ and $|S|$ are shown on Fig. 3.c as a function of the reduced frequency for

$$a = \frac{Z_c}{Z_m} = 0.5 \qquad (10)$$

It is seen that the choice of the circuit elements allows some control of the converter performance. Typical values for Q are 3 or 4. The circuit provides a moderate bandwidth. The parameters K_m and T_m have not been shown because they are less important for practical applications, but it may help to know that the reflection factor and transmission loss of the monofilar mode are very small.

3.2. Wideband annular slot converters

When mode converters have to work over very wide bands, a transformer can be integrated to the annular slot. In these types of converters the bandwidth is only limited by the transformer frequency response.

A first example is shown on Fig. 4.a. The winding sense is such that the coaxial mode would not develop a voltage across the slot if n_1 and n_2 were equal ; hence no mode conversion would occur. A small imbalance allows to create some conversion. The device has all the properties considered in the last alinea of paragraph 2. It is easy to show that

$$K_c = T_m = \frac{\xi}{2 + \xi} \qquad (11)$$

$$K_m = T_c = \frac{2}{2 + \xi}$$

$$S = \frac{(2\xi)^{1/2}}{2 + \xi}$$

where

$$\xi = \frac{Z_m}{Z_c}(1 - n_1/n_2)^2 \qquad (12)$$

A small imbalance will thus yield a low value for the reflection factor and insertion loss of the coaxial mode, but a high value of the same parameters of the monofilar mode.

If this is not desirable, the converter of Fig. 4.b can be used. In spite of the apparent simplicity of the circuit, the calculation of the scattering matrix is extremely tedious, for there is no symmetry plane. The general form (3) together with the properties (6) to (8) apply. One finds :

$$K_{c1} = \left[2(n_1^2 - n_2^2)x - (n_1 - n_2)^2 x^{-1}\right]/D$$

$$K_{c2} = \left[2(n_2^2 - n_1^2)x - (n_2 - n_1)^2 x^{-1}\right]/D$$

$$K_{m1} = K_{m2} = (n_1 - n_2)^2 x^{-1}/D$$

$$T_c = 4 n_1 n_2 / D$$

$$T_m = 2(n_1^2 + n_2^2)x / D \qquad (13)$$

$$K_{cm1} = T_{cm2} = 2 n_1 (n_1 - n_2)/D$$

$$K_{cm2} = T_{cm1} = 2 n_2 (n_1 - n_2)/D$$

with

$$x = (Z_m/Z_c)^{1/2}$$

$$D = 2(n_1^2 + n_2^2)x + (n_1 - n_2)^2 x^{-1}$$

A small imbalance allows here to have small values of the reflection factor and of the insertion loss for both modes.

4. LEAKY SECTION CONVERTERS

4.1. Coupled line theory of leaky coaxial cables

4.1.1. General theory of coupled lines

A leaky coaxial cable strung along a tunnel can be viewed as a system made of two interrelated transmission lines supporting two coupled modes which are called the coaxial and monofilar modes. Figure 5 shows the general circuit diagram of an elementary length dx. The coupling occurs through a transfer admittance Yt and a transfer impedance Zt. In the present case, the transfer admittance Yt is in general neglected and the transfer impedance is taken as purely inductive

$$Z_t = j\omega m \qquad (14)$$

This model, which can be refined by considering m as frequency dependent (Wait, 1977) appears to provide good results when compared to experiments. Anyway the currents and voltages satisfy the coupled line equations :

$$\begin{aligned} dI_c/dx &= -Y_c V_c + Y_t V_m \quad ; \quad dI_m/dx = Y_t V_c - Y_m V_m \\ dV_c/dx &= -Z_c I_c + Z_t I_m \quad ; \quad dV_m/dx = Z_t I_c - Z_m V_m \end{aligned} \qquad (15)$$

These equations may be written in matrix form using column matrices for the currents and voltages :

$$\begin{aligned} d\overline{I}/dx &= -\overline{Y}\,\overline{V} \\ d\overline{V}/dx &= -\overline{Z}\,\overline{I} \end{aligned} \qquad (16)$$

or, after derivation and elimination :

$$\begin{aligned} d^2\overline{I}/dx^2 &= \overline{Y}\,\overline{Z}\,\overline{I} \\ d^2\overline{V}/dx^2 &= \overline{Z}\,\overline{Y}\,\overline{V} \end{aligned} \qquad (17)$$

The eigenmodes are solutions of the form $\overline{V}(x) = \overline{V}_o \exp(-\Gamma x)$ and $\overline{I}(x) = \overline{I}_o \exp(-\Gamma x)$, where $\overline{V}_o$, $\overline{I}_o$ and Γ are solutions of the eigenvalue problems :

$$\begin{aligned} (\overline{Y}\,\overline{Z} - \Gamma^2 \overline{1})\,\overline{I}_o &= 0 \\ (\overline{Z}\,\overline{Y} - \Gamma^2 \overline{1})\,\overline{V}_o &= 0 \end{aligned} \qquad (18)$$

As $\overline{Y}$ and $\overline{Z}$ are symmetric matrices, $\overline{Y}\,\overline{Z}$ and $\overline{Z}\,\overline{Y}$ are transposed of each other. Thus they have the same eigenvalues Γ^2 and $\overline{I}_o$ and $\overline{V}_o$ are right and left eigenvectors of $\overline{Y}\,\overline{Z}$, respectively. Constructing two modal matrices $\overline{J}$ and $\overline{U}$, the columns of which are the eigenvectors I_o and V_o, respectively, the eigenvalue problems (18) may be written

$$\overline{Y}\,\overline{Z}\,\overline{J} = \overline{J}\,\overline{\mathrm{Diag}\,\Gamma^2} \quad ; \quad \overline{Z}\,\overline{Y}\,\overline{U} = \overline{U}\,\overline{\mathrm{Diag}\,\Gamma^2} \qquad (19)$$

However $\overline{J}$ and $\overline{U}$ may not be chosen independently because of the coupling of $\overline{I}$ and $\overline{V}$ apparent in (16). This imposes the constraint

$$\overline{J}\,\overline{\mathrm{Diag}\,\Gamma} = \overline{Y}\,\overline{U} \qquad (20)$$

or equivalently,

$$\overline{U}\,\overline{\mathrm{Diag}\,\Gamma} = \overline{Z}\,\overline{J} \qquad (21)$$

Finally the general solution of the coupled line equations can be written

$$\begin{aligned} \overline{V}(x) &= \overline{U}\left[\overline{\mathrm{Diag}\,e^{-\Gamma x}}\,\overline{A} + \overline{\mathrm{Diag}\,e^{\Gamma x}}\,\overline{B}\right] \\ \overline{I}(x) &= \overline{J}\left[\overline{\mathrm{Diag}\,e^{-\Gamma x}}\,\overline{A} - \overline{\mathrm{Diag}\,e^{\Gamma x}}\,\overline{B}\right] \end{aligned} \qquad (22)$$

where $\overline{A}$ and $\overline{B}$ are two arbitrary column matrices determined by the boundary conditions.

This very compact solution shows progressive and regressive waves which are the eigenmodes, with generalized matrix amplitudes $\overline{A}$ and $\overline{B}$. In order to avoid any confusion, these modes which are frequently called the coaxial and monofilar eigenmodes will be named here modes 1 and 2. They have propagation constants Γ_1 and Γ_2, respectively. Thus, the row index in $\overline{U}$ and $\overline{J}$ is either c or m, while the column index is either 1 or 2. It appears that the columns of $\overline{U}$ and $\overline{J}$ yield the voltage and current distributions of the eigenmodes. For instance, for the progressive eigenmode 1, we have $\overline{A} = (A_1, 0)$ and $\overline{B} = (0, 0)$, which yields

$$\left[V_c(x), V_m(x), I_c(x), I_m(x)\right] = \left[U_{c1}, U_{m1}, J_{c1}, J_{m1}\right] A_1 \exp(-\Gamma_1 x) \qquad (23)$$

The eigenmatrices $\overline{U}$ and $\overline{J}$ may be normalized but not independently, because of (20). The usual scattering matrix normalization

$$\overline{U}\overline{J} = \overline{E} \tag{24}$$

where $\overline{E}$ is the unit matrix and T indicates the transposition, is such that $\overline{A}^T\overline{A}/2$ and $\overline{B}^T\overline{B}/2$ are the complex powers of the progressive and regressive eigenmodes, respectively.

4.1.2. Application to non-leaky cables

Before applying the above theory to leaky cables, it is useful to first solve the problem for a non-leaky coaxial cable ($y_t = z_t = 0$). This case will be referred to by a subscript 0. The general theory is here rather trivial and reduces to standard transmission line theory. One finds

$$\overline{U}_0 = \begin{vmatrix} Z_{co}^{1/2} & 0 \\ 0 & Z_{mo}^{1/2} \end{vmatrix} \tag{25}$$

$$\overline{J}_0 = \begin{vmatrix} Z_{co}^{-1/2} & 0 \\ 0 & Z_{mo}^{-1/2} \end{vmatrix} \tag{26}$$

with

$$Z_{co} = (z_c/y_c)^{1/2} \quad , \quad Z_{mo} = (z_m/y_m)^{1/2} \tag{27}$$

and

$$\Gamma_{co} = (z_c\, y_c)^{1/2} \quad , \quad \Gamma_{mo} = (z_m\, y_m)^{1/2} \tag{28}$$

For the limited purpose of this paper we will neglect the losses and write, with the adequate subscripts c or m :

$$z = j\,\omega\,\ell \quad , \quad y = j\,\omega\,c \tag{29}$$

where ℓ and c are the specific inductance and capacitance, respectively.

4.1.3. Application to leaky cables

We now consider a leaky cable in a tunnel and assume a purely magnetic coupling : $y_t = 0$, $z_t = j\,\omega\,m$. The eigenvalue problem (19)-(20) with the normalization (24) can easily be solved numerically using a computer, but an analytical solution is extremely cumbersome. Here we are interested by approximations valid for a weak coupling assumption. Introducing two coupling coefficients

$$C_1 = \frac{m}{2\sqrt{\ell_m \ell_c}} \quad \frac{\beta_{co}/\beta_{mo}}{\beta_{co}/\beta_{mo} - 1} \tag{30}$$

$$C_2 = \frac{m}{2\sqrt{\ell_m \ell_c}} \quad \frac{\beta_{co}/\beta_{mo}}{\beta_{co}/\beta_{mo} + 1} \tag{31}$$

this means

$$C_1 \;,\; C_2 \;,\; C_1C_2 \ll 1 \tag{32}$$

After very tedious calculations one finds the approximate expressions

$$\overline{U} \cong \begin{vmatrix} Z_{co}^{1/2} & -Z_{co}^{1/2}(C_1\;C_2) \\ Z_{mo}^{1/2}(C_1+C_2) & Z_{mo}^{1/2} \end{vmatrix} \tag{33}$$

$$\overline{J} \cong \begin{vmatrix} Z_{co}^{-1/2} & -Z_{co}^{-1/2}(C_1+C_2) \\ Z_{mo}^{-1/2}(C_1-C_2) & Z_{mo}^{-1/2} \end{vmatrix} \tag{34}$$

$$\Gamma_1 \cong \Gamma_{co} \quad ; \quad \Gamma_2 \cong \Gamma_{mo} \tag{35}$$

It is clear that the modes 1 and 2 are the coaxial and monofilar eigenmodes, respectively. The meaning of the columns of $\overline{U}$ and $\overline{J}$ given previously and illustrated by (23) should not been forgotten when looking at (33) and (34). It is easily seen that an eigenmode carries relative powers 1 and $(C_1 - C_2)$ on either side of the cable sheath; this last part is the relative power of the leakage field.

4.2. A leaky section as a mode converter

We will now show that a leaky section inserted in a non-leaky cable acts as an efficient mode converter. First we consider a non-leaky to leaky transition as suggested on Fig. 6. This problem can be solved by using solutions of the type (22) on either side of the transition and expressing the continuity of the voltages and currents at the junction. This yields the matrix equations

$$\begin{aligned} \overline{U}_o(\overline{A}_o + \overline{B}_o) &= \overline{U}(\overline{A} + \overline{B}) \\ \overline{J}_o(\overline{A}_o - \overline{B}_o) &= \overline{J}(\overline{A} - \overline{B}) \end{aligned} \tag{36}$$

These may be solved for $\overline{B}_o$ and $\overline{A}$ and this yields the scattering matrix of the junction*. Explicitely and for the weak coupling assumption, we find

$$\begin{vmatrix} B_{co} \\ B_{mo} \\ A_1 \\ A_2 \end{vmatrix} \simeq \begin{vmatrix} -C_1C_2 & C_2 & 1 & C_1 \\ C_2 & C_1C_2 & C_1 & 1 \\ 1 & C_1 & -C_1C_2 & C_2 \\ -C_1 & 1 & -C_2 & C_1C_2 \end{vmatrix} \begin{vmatrix} A_{co} \\ A_{mo} \\ B_1 \\ B_2 \end{vmatrix} \tag{37}$$

Thus, at a transition, an incident mode is

- reflected with a (negligible) coefficient $\pm C_1C_2$
- reflected in the other mode with a coefficient $\pm C_2$
- transmitted in the same mode with a coefficient close to one
- transmitted in the other mode with a coefficient $\pm C_1$

It is now an easy task to calculate the scattering matrix of a mode converter consisting of a short section of a leaky cable inserted in a non-leaky one with equal transmission line parameters as suggested by Fig. 7. The general properties (1) to (4) of geometrically symmetric and non-dissipative converters apply and we find :

$$|K_c| = |K_m| \simeq 0 \tag{38}$$

$$|K_{cm}| = 2C_2 \sin\left[(\beta_{co} + \beta_{mo})L/2\right] \tag{39}$$

$$|T_{cm}| = 2C_1 \sin\left[(\beta_{co} - \beta_{mo})L/2\right] \tag{40}$$

$$|T_c|^2 = |T_m|^2 = 1 - |K^2_{cm}| - |T^2_{cm}| \tag{41}$$

The non-equality of the mode conversion factors in reflection K_{cm} and in transmission T_{cm} is a result of the non zero electrical length of the device. As the phase constants β_{co} and β_{mo} are proportional to the frequency, the bandwidth of T_{cm} around a maximum of the sine function is larger than that of K_{cm}. It is moreover maximum for

$$(\beta_{co} - \beta_{mo})L = \pi \tag{42}$$

The 3 dB-bandwidth of T_{cm} thus extend from $f_o/2$ to $3f_o/2$, where f_o is the design frequency yielding (42).

For the design of a mode converter we may consider that

$$\beta_{mo} \simeq k_o = \omega/(3.10^{8}) \tag{43}$$

$$\beta_{co} \simeq k_o\varepsilon^{1/2} \tag{44}$$

where k_o is the free-space phase constant and ε is the dielectric constant of the cable insulation. It is seen from (30) and (31) that $C_1 > C_2$, thus $|T_{cm}| > |K_{cm}|$. The mode conversion thereby exhibit an intrinsic directivity which can be enhanced by chosing a low value for ε ; this yields an increase of $|T_{cm}|$. This directivity may be very useful in tunnel application, for it reduces the standing waves due to monofilar waves travelling in opposite directions between the mode converters.

---

* We take here account that, in transmission line theory, the A and B waves are the progressive and regressive waves, respectively. In network theory however they are the ingoing and outgoing waves, respectively, and the scattering matrix relates the latter to the former.

5. MODE CONVERTERS FOR TWO-WIRE LINES

5.1. General principle

When an open two-wire line is strung into a tunnel the balanced (bifilar) and unbalanced (monofilar) modes are coupled unless the two wires are strictly symmetric with respect to the wall. In practice however the asymmetry yields an extremely low coupling which may be neglected. This means that we consider that the balanced mode yields a negligible electromagnetic field in the tunnel space.

Groundless mode converters for this type of transmission line can be made by inserting into it lumped impedances and transformers which create a deliberately high asymmetry between the two wires. All the properties (1)-(9) of the scattering matrix are valid.

The procedure for obtaining the scattering matrix is in general long and tedious. It involves the following steps (Fig. 8)

- write the four circuit equations
- convert these into new equations relating the coaxial (bifilar) and monofilar currents and voltages defined by

$$I_m = I + I' \qquad V_m = (V + V')/2$$
$$I_c = (I - I')/2 \qquad V_c = V - V' \qquad (45)$$

- convert these new equations into ones relating the ingoing and outgoing waves, using

$$V_i = Z_i^{1/2} (A_i + B_i)$$
$$\qquad\qquad i = c , m \qquad (46)$$
$$I_i = Z_i^{-1/2} (A_i - B_i)$$

- solve the equations in the unknows B_{c1}, B_{m1}, B_{c2}, B_{m2} to obtain the scattering matrix of the mode converter.

5.2. Applications

Figures 9.a and 9.b show a selective and a wideband mode converter, respectively. They are very similar to those seen previously for a coaxial cable. The scattering matrix of these devices is given in (DERYCK, 1975). We will recall here formulas valid at the center frequency $\omega_0 = (LC)^{1/2}$ of the selective converter or for the wideband converter when $n_1 = n_4$ and $n_2 = n_3$:

$$K_c = K_m = d^2/(d^2 + 1)$$
$$T_c = T_m = 1/(d^2 + 1) \qquad (47)$$
$$S = d/(d^2 + 1)$$

where d is an asymmetry factor given by

$$d = 2\,\omega_0 C\ (Z_c Z_m)^{1/2} \qquad (48)$$

for the selective mode converter and

$$d = \frac{n_1 - n_2}{2(n_1 + n_2)}\ (Z_c/Z_m)^{1/2} \qquad (49)$$

for the wideband converter.

6. INADVERTENT MODE CONVERSIONS

As was said previsouly, the asymmetry of the two wires of a two-wire line with respect to the tunnel wall is in general too small to produce significant mode conversions, unless the line is strung very close to the wall. Strong inadvertent mode conversions may however be due to asymmetric objects close to the line, as for instance hanging devices, etc. These effects have been nicely analyzed in (DERYCK, 1973) and occur also with leaky coaxial cables.

7. CONCLUSIONS

Various mode converters for tunnel transmission have been described. Other versions can easily be designed; some of them are effective as mode converters in a frequency band but are inactive and transparent in another frequency range. It is thus possible to combine them into a multi-frequency system using a single cable.

8. ACKNOWLEDGMENTS

These mode converters have been designed by the author while acting as a consultant to the Institut National des Industries Extractives (INIEX, Liége, Belgium) and are patented by this Institute. Thanks are due to Mrs E. COLLE for typing the manuscript.

9. BIBLIOGRAPHY

DELOGNE, P., 1970, "Les liaisons radioélectriques par câble coaxial dans la mine", Annales des Mines de Belgique, n° 7-8, pp. 967-975.

DELOGNE, P., and LIEGEOIS, R., 1971, "Le rayonnement d'une interruption du conducteur extérieur d'un câble coaxial", Ann. Télécom., France, tome 26, n° 3-4, pp. 85-100.

DELOGNE, P., 1973, "Les liaisons par radio en milieu souterrain", Revue HF, Belg., vol. 9, n° 2, pp. 18-26.

DELOGNE, P., 1976, "Basic Mechanisms of Tunnel Propagation", Radio Science, vol. 11, n° 4, pp. 295-303.

DERYCK, L., 1973, "Etude de la propagation des ondes électromagnétiques guidées dans les galeries souterraines", Thèse de doctorat en sciences, Univ. Liége, Belgium.

DERYCK, L., 1975, "Control of mode conversions on bifilar line in tunnels", Proc. IERE, vol. 45 n° 5, pp. 241-247.

HILL, D.A., and WAIT, J.R., 1975, "Electromagnetic Fields of a Coaxial Cable with an Interrupted Shield Located in a Circular Tunnel", J. Appl. Phys., vol. 46, n° 10, pp. 4352-4356.

MAHMOUD, S.F., and WAIT, J.R. 1974, "Theory of Wave Propagation Along a Thin Wire Inside a Rectangular Waveguide", Radio Science, vol. 9, pp. 417-420.

SEIDEL, D.B., and WAIT, J.R., 1978, "Transmission Modes in a Braided Coaxial Cable and Coupling to a Tunnel Environment", IEEE Trans., vol. MTT-26, pp 494-499.

WAIT, J.R., and HILL, D.A., 1974, "Guided Electromagnetic Waves Along an Axial Conductor in a Circular Tunnel", IEEE Trans., vol. AP-22; n° 4, pp. 627-630.

WAIT, J.R., and HILL, D.A., 1975, "On the Electromagnetic Field of a Dielectric Coated Coaxial Cable with an Interrupted Shield". IEEE Trans., vol. AP-23, n° 4, pp. 470-479.

WAIT, J.R., and HILL, D.A., 1977, "Influence of Spatial Dispersion of the Shield Transfer Impedance of a Braided Coaxial Cable", IEEE Trans., vol. MTT-25, pp. 72-7[illegible].

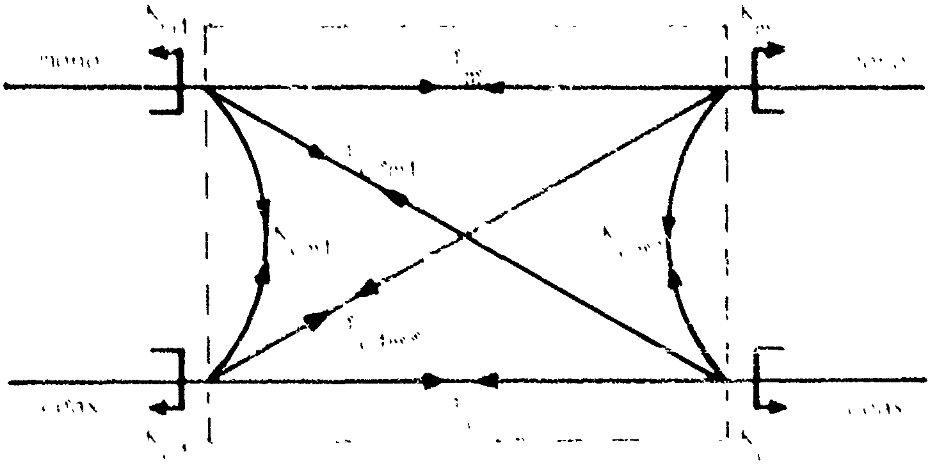


Fig 1 Reflection, transmission and mode conversion factors in the most general case

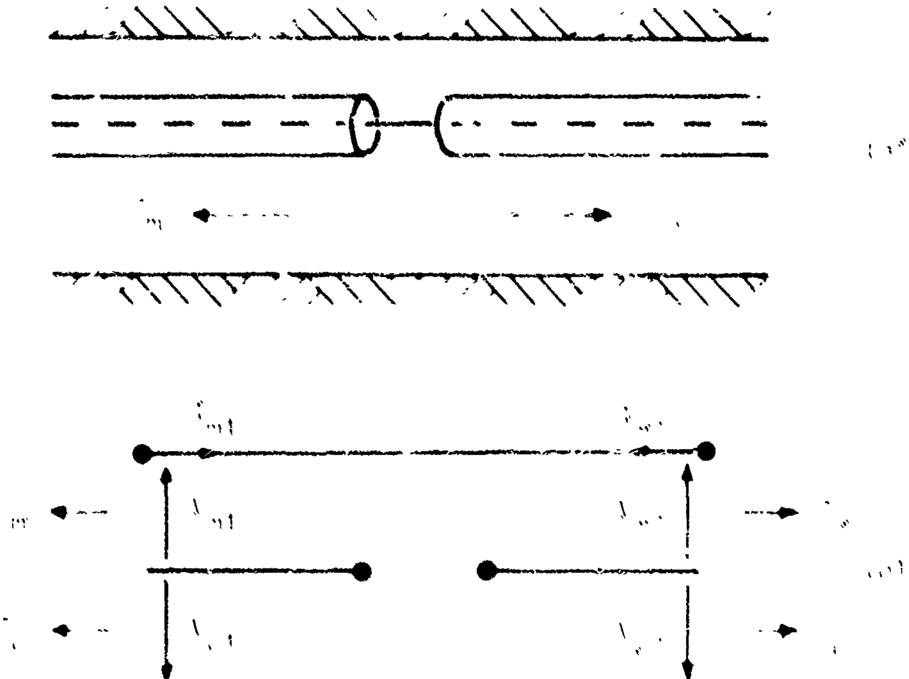
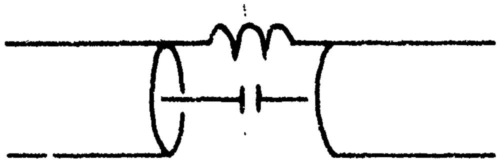
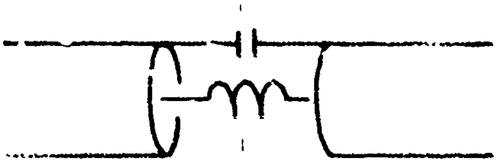


Fig 2 Principle and equivalent circuit for the quasi-static analysis of the annular slot mode converter (without lumped circuit elements)



[illegible]



[illegible]

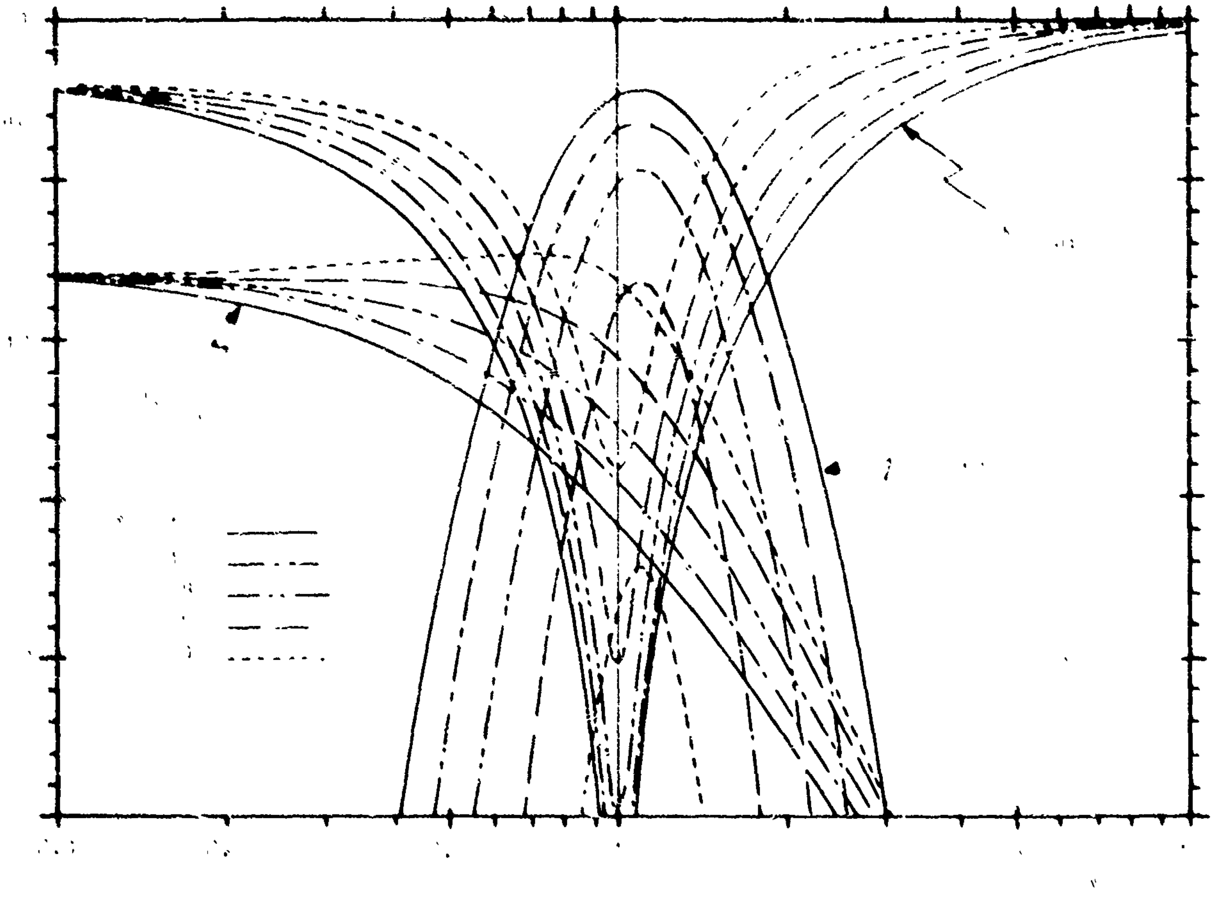


Fig. 3 Selective annular slot converters and performance curves

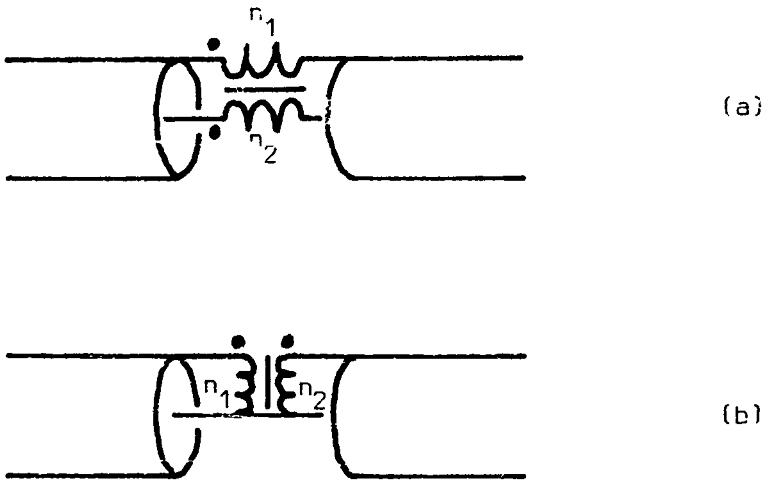

n1
n2
(a)
n1
n2
(b)

Fig.4 Wideband annular-slot converters

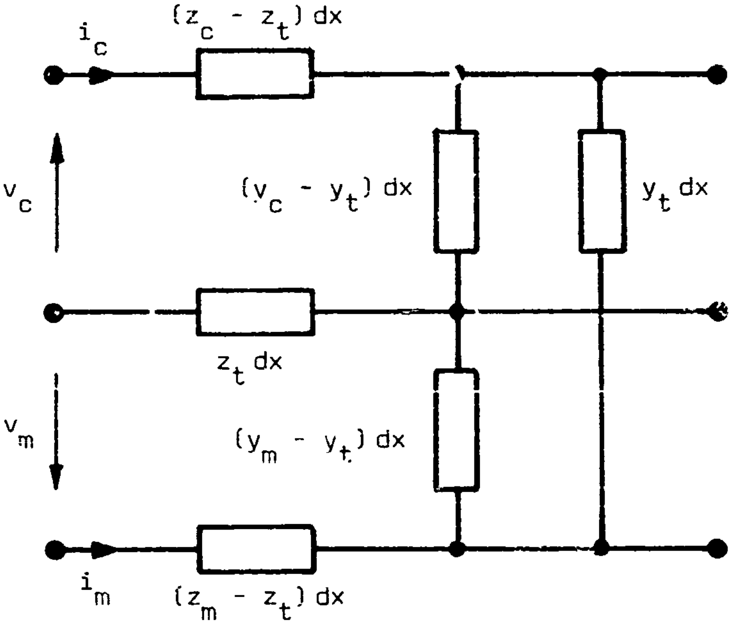

ic
(zc - zt) dx
vc
(yc - yt) dx
yt dx
zt dx
vm
(ym - yt) dx
im
(zm - zt) dx

Fig.5 Equivalent circuit of an elementary length of a leaky cable

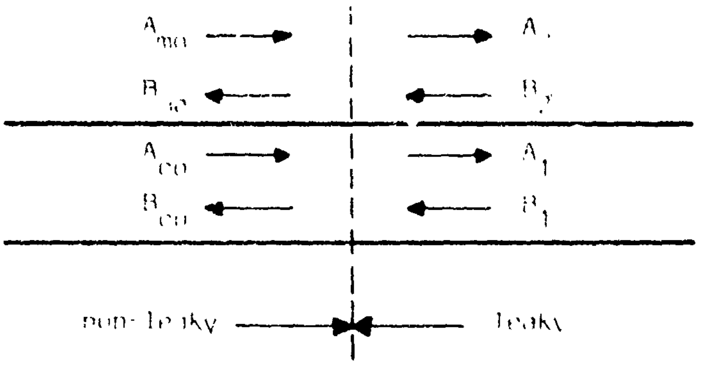

A_{m0}
A
B_{m0}
B
A_{c0}
A_1
B_{c0}
B_1
non-leaky
leaky

Fig 6 Transition from a non-leaky to a leaky cable

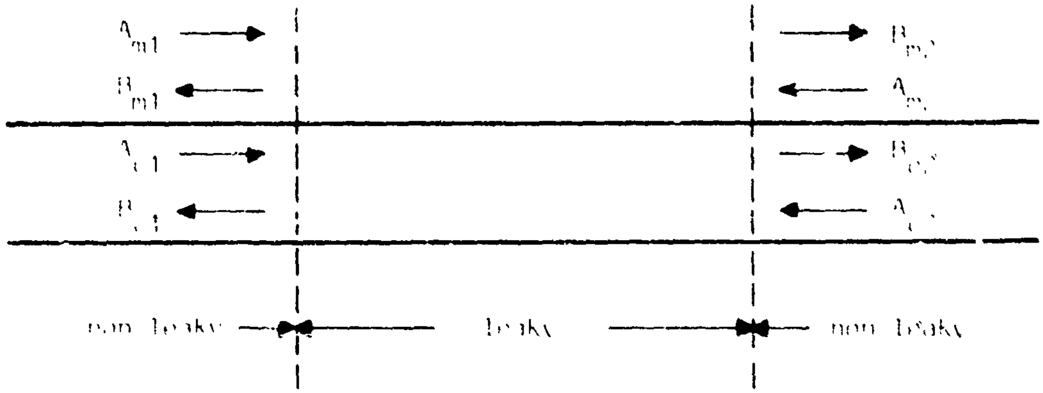

A_{m1}
B_{m2}
B_{m1}
A_{m2}
A_{c1}
B_{c2}
B_{c1}
A_{c2}
non leaky
leaky
non leaky

Fig 7 Leaky section as a mode converter

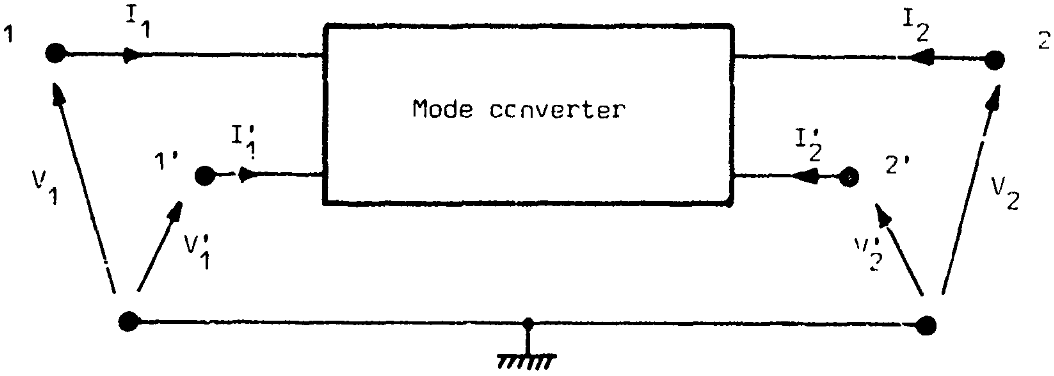

1
I1
Mode converter
I'1
1'
I2
2
I'2
2'
V1
V'1
V2
V'2

Fig.8 Wire currents and voltages in a mode converter for two-wire lines

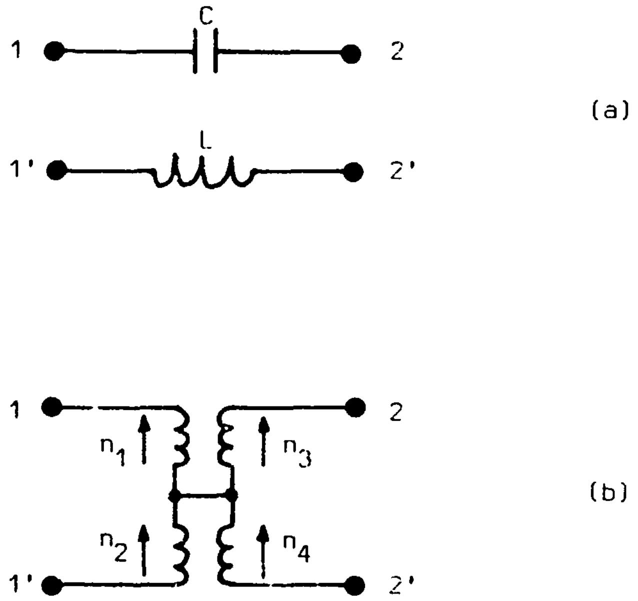

C
1
2
(a)
L
1'
2'
1
2
n1
n3
(b)
n2
n4
1'
2'

Fig.9 Selective (a) and wideband (b) mode converters for two-wire lines

DISCUSSION

D.A. Hill, US

You mentioned the use of mode converters at VHF. How would you modify your transmission line analysis at higher frequencies where waveguide modes can propagate?

Author's Reply

At VHF you can have a very large number of propagating waveguide modes and a modal approach is impractical. A more useful approach is to first study the free-space case wherein the mode converter appears to radiate as a directive antenna and, in the second step, to study the propagation of these radiated waves in tunnels. This approach has been mentioned in one of the referenced papers (Delogne, 1976). A useful method for the study of the propagation of VHF waves into tunnels is the geometrical optical approach which has been used by J.R. Wait.

L.B. Felsen, US

Concerning mode coupling introduced by a section of leaky cable inserted into a non-leaky cable, did you employ a model which relies the properties of the finite leaky section are ascertained from those of an *infinite* leaky cable, without accounting for nonuniformities introduced by end effects in a *finite* periodic structure? If yes, are there results available justifying such an assumption?

Author's Reply

We did not take account of interactions between mode converters. The reason for this is that we space the mode converters in order to have an attenuation of about 15 dB for the monofilar mode between two adjacent mode converters. Interaction through the monofilar mode is thus considerably damped. As the mode converters provide an excellent impedance matching for the coaxial mode, the only important interaction is through the simple transmission in the coaxial mode.

LEAKY COAXIAL CABLES FOR OBSTACLE DETECTION AND CONTINUOUS ACCESS GUIDED COMMUNICATIONS

N.A.M. Mackay, J.C. Beal, D.J. Gale, J.L. Mason
Department of Electrical Engineering
Queen's University
Kingston, Ontario, Canada, K7L 3N6

SUMMARY

Leaky coaxial cables have been introduced commercially with the aim of providing continuous access guided communications in difficult environments such as mines and rail tunnels. They have also been studied as a guiding medium for use in an obstacle detection scheme called guided radar. Recent work at Queen's has been directed towards developing techniques for predicting the performance of leaky cable installations in a variety of configurations. A two-cable cavity resonator test allows the coupling effectiveness of different cable designs to be measured in the laboratory, in free space or in the presence of different surrounding environments. An associated theoretical treatment has been developed, based upon extensions of coupled transmission line theory. This theoretical model of the cavity has been extended to simulate a full scale guided radar system.

The combination of experiment and modelling has produced a more fundamental understanding of the way in which coupled leaky coaxial cables function in various environments.

1. INTRODUCTION

1.1 Preamble

For the past few years so-called leaky coaxial cables have been used as a means of providing continuous access guided communications in environments which are not conducive to normal R.F. propagation, such as mines and tunnels. They have also been used as sensors in a security system for detecting and locating passive objects around a defined perimeter. All types of leaky coaxial cable have in common an incomplete outer conductor which allows a portion of the energy within the cable to "leak out" and interact with the surrounding environment.

At Queen's University in Kingston, Canada we have been interested in the use of leaky cables for one particular application - an obstacle detection system called Guided Radar (Patterson, R.E., 1977). Even though the provision of continuous access communications is not a goal of this work the techniques that have been developed for measuring the important parameters of the cables and the theoretical understanding of how they interact with the environment in which they are situated could have direct application to communications.

1.2 Guided Radar Concept

The basic principle upon which Guided Radar operates is that an object in the vicinity of two coupled leaky cables interacts with the field external to the cables in a manner that can be detected at the end of one of the cables. The system is illustrated in figure 1. Two leaky cables are laid on the ground, buried just below the surface or mounted in the air parallel to each other along a perimeter and terminated at the far end by conventional matched loads. A pulse transmitter is connected to the near end of one cable and a receiver to the near end of the other. As a pulse propagates along the transmit cable energy is coupled into the receive cable and detected, after the appropriate time delay, at the receiver. Discontinuities in the leaky cables, stationary objects and environmental variations produce a time varying return signal at the receiver. However, this return signal or "profile" changes very slowly from one pulse to the next. When a target enters the area where it has a significant effect on the coupled energy, the rapid change in profile from one pulse to the next permits the detection of the target. Obviously, in the pulse system the target can also be located in one dimension by measuring the time delay between transmission of a pulse and the change in the profile.

Of critical importance to the designer of either a guided radar system or a continuous access communications system is the overall system sensitivity and how this is related to the various parameters of cable design and to the environment in which the system must operate.

As shown in figure 2, in a communication system the total loss is made up of transmission loss in the cable itself and a coupling loss between the cable and a portable receive antenna. The guided radar system has not only transmission loss and coupling loss but also a loss due to what might be called the radar cross-section of the target.

It is, however, the coupling that determines the eventual sensitivity of either of these systems and it is the coupling that is most dependent upon the environment and cable design and also the most difficult to predict.

Much of the recent work at Queen's has been directed towards reliably measuring this coupling in a convenient way and towards developing a theoretical model of the coupling process.

1.3 Leaky Coaxial Cables

Coaxial cables can be made leaky in a variety of ways: by milling slots or drilling holes in the solid outer conductor of a conventional coaxial design, by deliberately omitting some of the conductors in a braided cable design or by creating novel designs. Typical designs are illustrated in figure 3. There are a number of cable design parameters that can be varied in each case: the spacing shape and size of the holes; the lossiness of the weave; or in the case of the bottom two types - the use of a circumferential discontinuity only or a combination of circumferential and longitudinal discontinuities. The design of leaky cables seems to have remained an empirical one, based mainly upon ease of manufacture, with very little being finally understood about the way in which the detailed construction of the outer conductor affects their performance in real environments. There are two reasons for this situation: first, because of the complex structure of the cables an exact calculation of the field distribution in their vicinity, particularly when placed in realistic environments, is virtually impossible; second, it has been difficult to measure characteristics of the cables which would indicate reliably their expected performance in a practical sense without resorting to the construction of a full scale system.

To allow various cable designs to be compared with respect to their coupling effectiveness a novel two cable resonant cavity technique has been developed. This test can easily be carried out on short cable lengths within the confines of the laboratory.

2. TWO-CABLE CAVITY RESONATOR

2.1 Description

Two 5 metre lengths of leaky coaxial cable are placed between parallel, metallic end plates as shown in figure 4. The cables are parallel to each other and separated by a fixed distance - either 0.45 m or 0.9 m. Access to the ends of the cables is provided by standard N-type, in-line connectors at each end-plate. Normally both cables are terminated by a standard 50Ω coaxial load at one end, external to the end plate. At the other end of the cavity one cable is connected to a swept-frequency generator and the other cable is connected to a wattmeter or a network analyser used as a ratiometer. The standard test of cable properties consists of a measurement of the power transferred from the source cable to the monitored cable over desired frequency range. The result is a power versus frequency characteristic for the particular cable type referred to as its "coupled power signature".

The cavity resonator can be theoretically modelled as a system of three coupled transmission lines: two coaxial lines and a two wire transmission line formed by the outer conductors of each of the leaky cables. The total model is summarized in figure 5. It is assumed that cable parameters such as phase velocity, attenuation and characteristic impedance apply here to the leaky cables as they do to conventional coaxial TEM transmission lines. The loads at either end of the coaxial cables are assumed to match the nominal characteristic impedance of the cables: 50Ω. The cavity end-plates act as short circuits for the two wire line producing a very low value for Z_{L2} and Z_{L1}. The total voltage on each line is composed of a weighted sum of the voltage contributions from all the lines in the system. When the estimated values for particular parameters and the appropriate boundary conditions for each line are substituted into the coupled equations describing the line voltages the resulting eigenvalue problem can be solved to yield the total voltage seen by a receiver connected to port B under a variety of different experimental conditions. The various estimated parameters can then be adjusted iteratively to produce the closest match to an experimental signature.

2.2 Typical Signatures

Figure 6 shows signatures for one popular type of leaky cable. The experimental signature is shown on top and the corresponding theoretical signature on the bottom. The most prominant feature of these signatures is the appearance of resonant peaks occurring at regularly spaced intervals - namely. 29.2 MHz. This spacing gives a measure of the phase velocity of propagation on the short-circuited two wire line which turns out to be 98% of the free space velocity of light. It is also obvious that the cavity resonances have superimposed upon them an amplitude envelope with minima spaced approximately 225 MHz apart. The spacing of the envelope minima is strongly linked to the difference in phase velocity between a single, isolated leaky cable and the two wire line.

To demonstrate this, a cable was tested that was essentially the same construction as the one previously tested with the exception that low density foam dielectric was used. The signature is shown in figure 7, again with the experimental signature on top and the theoretical signature on the bottom. The resonance spacing remains about 29 MHz and the overall coupling level is similar. However, the envelope minima are now about 460 MHz apart since the coaxial phase velocity - 87% of the speed of light - is closer in value to that of the two wire line.

One figure of merit that results from this cavity test is coupled power level - the maximum power coupled from one leaky cable to the other. By use of a reference cable with known performance in a practical continuous access guided communications or guided radar system the relative usefulness with respect to coupling of one cable type compared to another can be predicted by comparison of the signatures.

Several of the cable types tested in the cavity resonator have also been tested in a full scale guided radar system. In a guided radar system one measure of coupled power is the relative sensitivity of the system to a standard target - in our case, a human being walking along the mid line between the cables.

If the relative coupling from the cavity test is compared to the relative human target sensitivity in a guided radar good correspondence results, particularly for cables of similar construction, as shown in figure 8. For example, the types labelled A1, A2 and A3 differ only in the size of hole milled in the solid outer conductor. Limitations occur however when comparing cables with significantly different construction such as those labelled C and B3. The two measures of "goodness" - the cavity and the field test - rank these cables in opposite order. This effect is most likely due to the large difference in cable spacing in the two tests - 90 cm in the cavity, 4 m in the field - and the difference in rates of radial decay of the field external to the cables that might be expected with the different constructions.

3. MODIFIED TWO-CABLE RESONANT CAVITY

3.1 Description

The resonant cavity appears to be a convenient method for gauging the effect of changes in cable design on performance in a practical system. However, in practical continuous access guided communications or guided radar systems the cables are installed in areas where there are large variations in the properties of the environment surrounding the cables. In a guided radar scheme the cables can be in the air, corresponding to the situation in the cavity, laid on the ground or buried in it. For buried cables, the ground moisture and type of soil could be expected to have considerable effect on the system performance. Therefore, the resonant cavity has been modified to allow the introduction of various materials in the vicinity of the cables to simulate environmental effects.

A trough was built from polyfoam and wood, within the cavity. When the trough is empty it has very little effect on the cable signatures. However, if tap water or wood shavings, moistened with varying amounts of water, are put in the trough significant effects are seen due to the increased attenuation on the two wire line and the altered dielectric constant of the environment.

By filling the trough with wood shavings moistened with various amounts of water controlled changes in the effective dielectric constant of the environment can be effected.

3.2 Typical Results

Figure 9 shows the effect of filling the trough with 30 cm of dry wood shavings. In other words the cables were buried 10 cm below the surface of the shavings. Although the wood shavings were ostensibly dry, they were fresh, and undoubtedly contained some internal moisture. In fact, the effect of the dry wood shavings likely would not be greatly different from that of dry ground which also would always contain some moisture. Several significant changes in the signature have occurred. The nulls in the overall envelope are now spaced approximately 360 MHz apart rather than 225 MHz as they were in free space indicating that the phase velocity on the two wire line has decreased due to the increased dielectric constant of the surrounding medium. There is also a considerable decrease in the coupled power, particularly at high frequencies. When the wood shavings are introduced into the cavity two conflicting influences on the coupled power would be expected. First, since the phase velocities on the coaxial mode and the two wire mode are more closely matched, more power will be coupled from the coaxial line into the two wire line. However, at the same time, more energy is dissipated in the lossy material now surrounding the cables, reducing the output signal. In this particular case the increased attenuation in the wood chips in the predominant effect.

The addition of 2% tap water by volume to the wood chips further increases the dielectric constant, decreasing the two wire line phase velocity. The resulting signature is shown in figure 10. The overall envelope minima spacing increases further. In fact the null spacing of 800 MHz indicates that the coaxial and two wire line phase velocities are very nearly in synchronism.

Another feature that is evident in both of the last signatures is the appearance of a lower envelope for the individual resonances which follows closely the peaks and nulls of the imagined overall envelope. This elevated baseline seems to represent the return signal from the far end plate acting as a large target in a guided radar system. The cavity resonances then magnify this basic target return at the appropriate frequencies by an amount dependent upon the Q of the cavity.

3.5 Theoretical Interpretation

A detailed interpretation of the effects seen in the signatures when a lossy medium is introduced into the cavity has not yet been carried out. However, by introducing the appropriate loss factors into the coupled line model, signature characteristics are produced which are similar to those measured in the cavity. Such a simulated signature is shown in figure 11. The similarity between this plot and the previous one is obvious - both have an elevated baseline with superimposed cavity resonances and very widely spaced envelope minima.

With the theoretical model it is possible to simulate conditions which are very difficult or impossible to achieve in practice. For example, the solid line in figure 11 is the signature when the two-wire line is assumed to be perfectly matched at the near end and has a perfect short circuit at the far end. This curve follows very closely the so-called elevated baseline.

4. GUIDED RADAR SIMULATION

It is possible to simulate the return signal in an actual guided radar system that results from the intrusion of a target by assuming that the target acts as an arbitrary unmatched impedance, say 1000 , connected across the two wire line at some point. The coupled line model can then be used, with the appropriate boundary conditions, to calculate the change in the return signal due to the target. By moving the impedance down the cables from the receiver towards the load a complete return signal can be generated which corresponds to the standard test of performance of a full scale guided radar system - a standard walk - where a human walks at constant velocity down the length of the cables midway between them. Figure 12 shows a typical return signal from an actual system on the bottom and the results from the coupled line model on top. The cables were in the air at a height of 1.5 m. Therefore, the two wire line was assumed to have a phase velocity approaching that of free space and the attenuation coefficient on the two wire line was assumed to be zero. The basic sinusoidal nature of the signal is a result of the fact that synchronous detection is normally used in guided radar and only the in-phase component of the return signal is shown. The envelope results from a beating effect between the energy propagating on the two wire line and the energy propagating on the coaxial lines. Further work is now underway to improve the model of the guided radar in order to be able to predict in advance the influence of the environment in which a particular system might be installed on the sensitivity.

5. SUMMARY

The two cable cavity resonator test produces a coupled power signature which allows the coupling effectiveness of different leaky cable designs to be compared. The cavity has been modified to permit the introduction of various materials into the vicinity of the cables to more closely match the environmental conditions within which the cables are used in a normal guided radar installation. A theoretical coupled transmission line model of the resonant cavity has been developed which produces results which very closely match those obtained in the laboratory. The model can also be extended to simulate a full scale guided radar system. The combination of experiment and modelling has produced a more fundamental understanding of the way in which coupled leaky cables function in various environments.

6. ACKNOWLEDGEMENTS

This work has received support from the Computing Devices Company, Ottawa, the National Research Council of Canada, and the Canadian Institute of Guided Ground Transport at Queen's University.

7. REFERENCE

Patterson, R.F., and Mackay, N.A.M., "A Guided Radar System for Obstacle Detection", IEEE Trans. on Instrumentation and Measurement, IM-26(2), 137, June 1977.

8. BIBLIOGRAPHY

More information about guided radar and the two-cable resonant cavity test can be found in:

Mackay, N.A.M., Beal, J.C., "Guided Radar for Obstacle Detection in Ground Transportation Systems", IEEE AP-S Symposium, Digest, pp. 302, Washington, D.C., May 1978.

Gale, D.J., Beal, J.C., Comparative Testing of Leaky Coaxial Cables by Use of a Two-Cable Cavity Resonator", IEEE MTT-S International Symposium, Digest, pp. 270, Ottawa, Ontario, June 1978.

Two particularly useful special meetings on leaky coaxial cables and tunnel communications are:

Davis, Q.V., (ED.): Special Issue on Leaky Feeders, The Radio and Electronic Engineering Journal, I.E.R.E., 45(5), 201, May 1975.

Wait, J.R., (ED.): Electromagnetic Guided Waves in Mine Environments, Workshop Proceedings, I.T.S., N.T.I.A., U.S. Dept. of Commerce, Boulder, Colorado, March 1978.

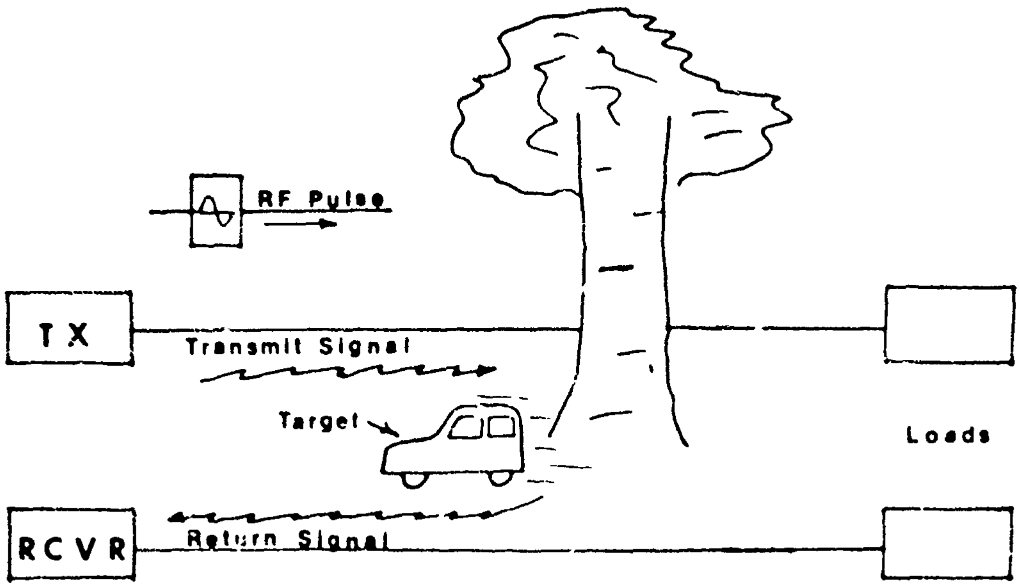
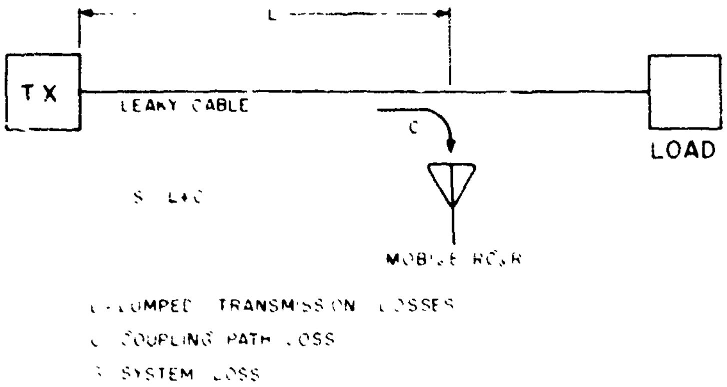
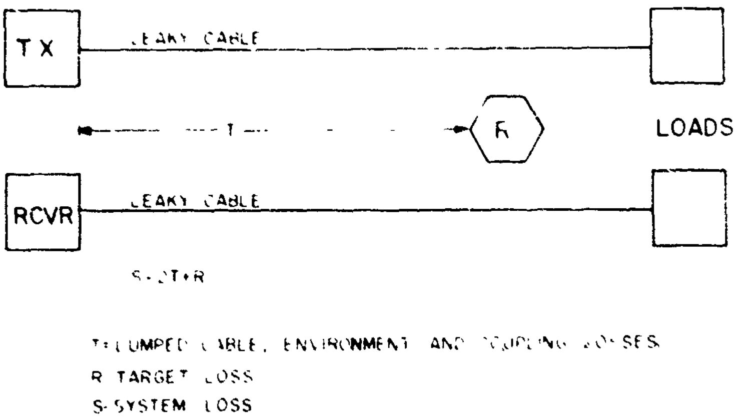

RF Pulse
TX
Transmit Signal
Target
Loads
RCVR
Return Signal

Fig.1 Guided radar concept


L
TX
LEAKY CABLE
C
LOAD
S = L+C
MOBILE RCVR
L = LUMPED TRANSMISSION LOSSES
C = COUPLING PATH LOSS
S = SYSTEM LOSS

(a) CAGC SYSTEM LOSS


TX
LEAKY CABLE
T
R
LOADS
RCVR
LEAKY CABLE
S=2T+R
T=LUMPED CABLE, ENVIRONMENT AND COUPLING LOSSES
R=TARGET LOSS
S=SYSTEM LOSS

(b) SYSTEM LOSS IN A GUIDED RADAR TARGET DETECTION SCHEME

Fig.2 Signal losses in guided radar and continuous access guided communications systems

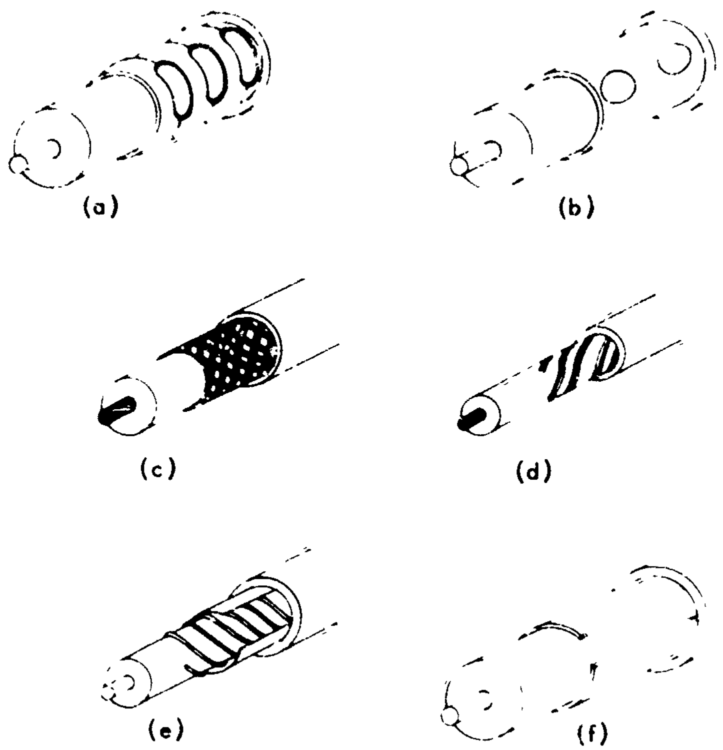


Fig.3 Leaky coaxial cable constructions

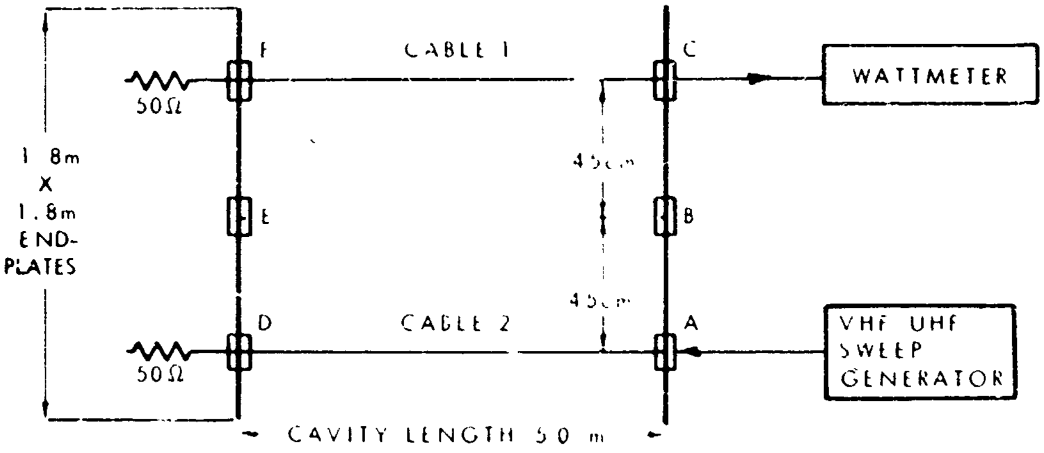
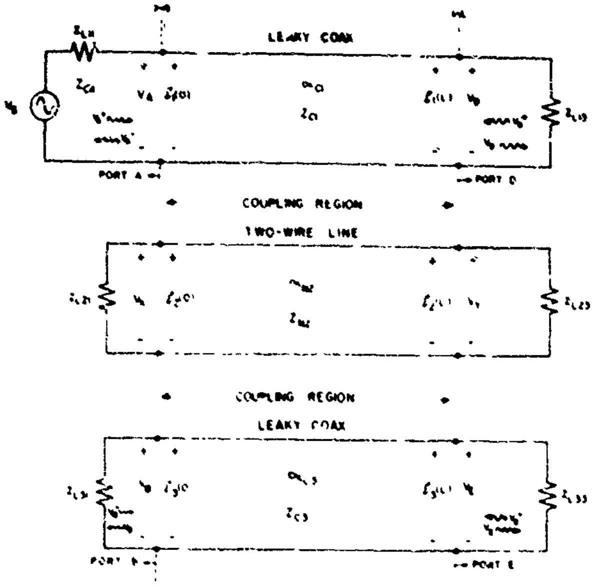

F
CABLE 1
C
WATTMETER
50Ω
1.8m
X
1.8m
END-
PLATES
45cm
E
B
45cm
D
CABLE 2
A
VHF UHF
SWEEP
GENERATOR
50Ω
CAVITY LENGTH 5.0 m

Fig.4 Two cable cavity resonator


LEAKY COAX
PORT A
PORT D
COUPLING REGION
TWO-WIRE LINE
COUPLING REGION
LEAKY COAX
PORT B
PORT E

Z DISTANCE ALONG THE COUPLING REGION

V_s SOURCE VOLTAGE

V_n = $V_n^+ + V_n^-$, THE TOTAL PORT n VOLTAGE COMPOSED OF INCIDENT AND REFLECTED COMPONENTS

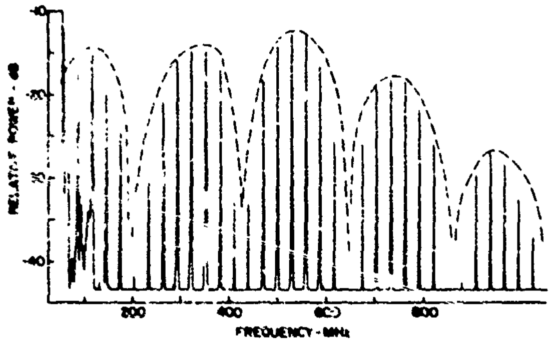
$\mathcal{E}_i(Z)$ = THE TOTAL VOLTAGE ON THE i^{th} LINE

Z_{Lim} = THE IMPEDANCES THAT TERMINATE THE i^{th} LINE

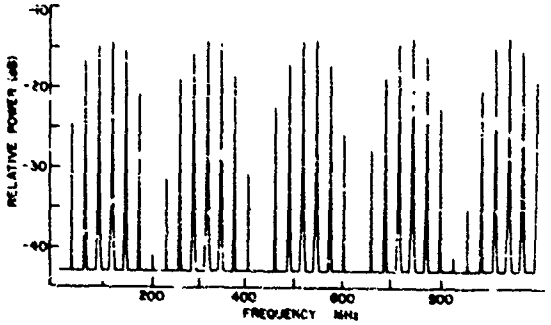
Z_{C11}, Z_{C1}, Z_{N2}, Z_{C3} = CHARACTERISTIC IMPEDANCES

α_{C1}, α_{N2}, α_{C3} = ATTENUATION COEFFICIENTS

Fig.5 Two cable cavity resonator transmission line model

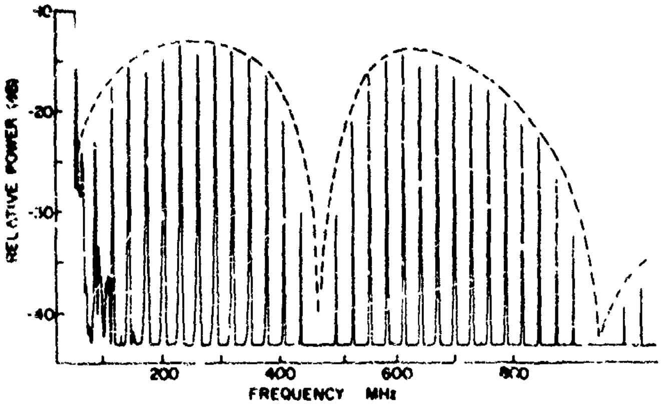


CABLE B TEST SIGNATURE CABLES SPACING: 90 cm

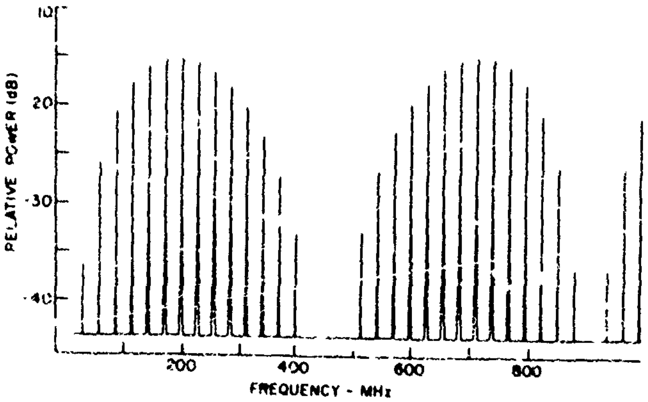


CABLE B THEORETICAL SIGNATURE

Fig 6 Coupled power signatures - experimental and theoretical



CABLE A TEST SIGNATURE CABLES SPACING:
90 cm



CABLE A THEORETICAL SIGNATURE

Fig.7 Coupled power signatures - experimental and theoretical

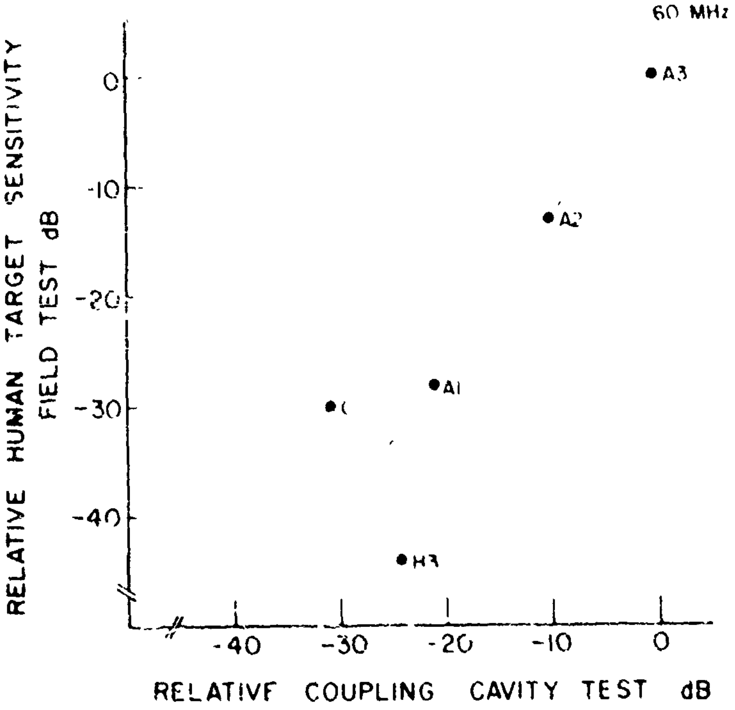
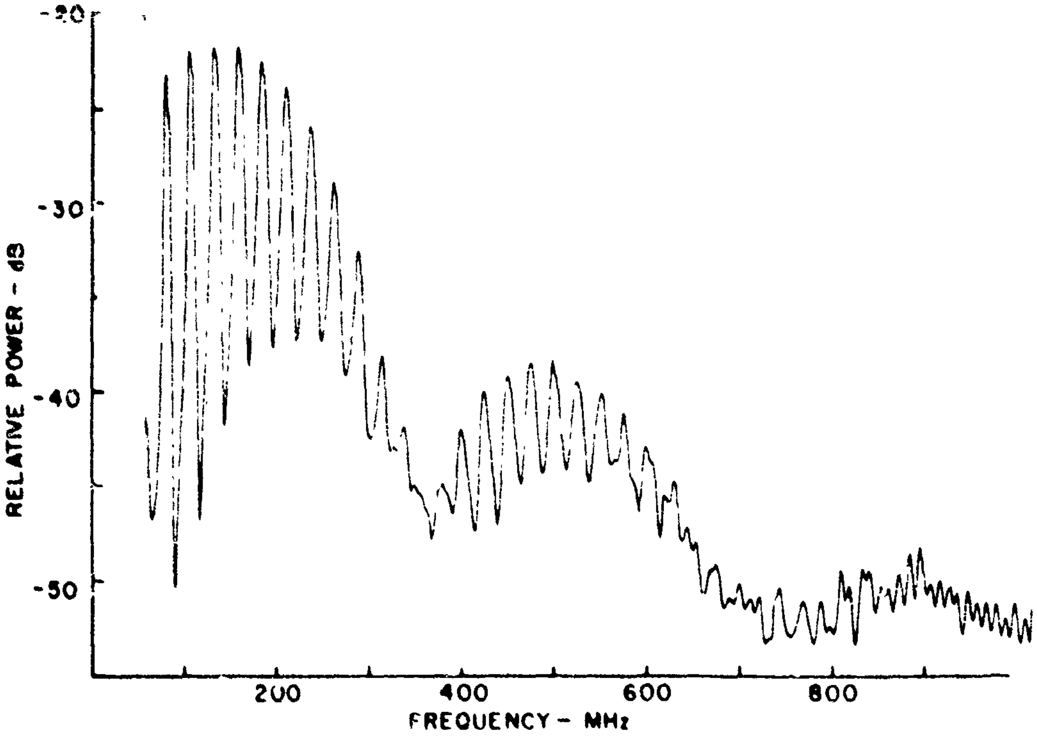
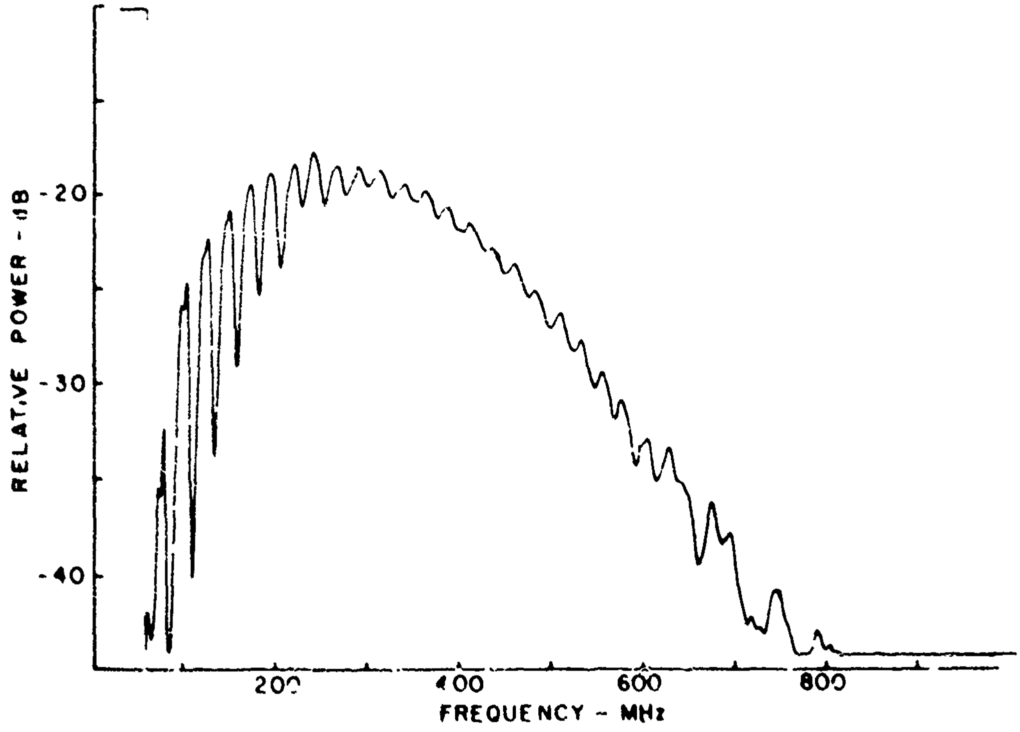


Fig 8 Comparison of field and laboratory test results



30cm WOOD SHAVINGS.

Fig.9 Signature in lossy environment



30cm WOOD SHAVINGS, 2% TAP WATER.

Fig.10 Signature in lossy environment

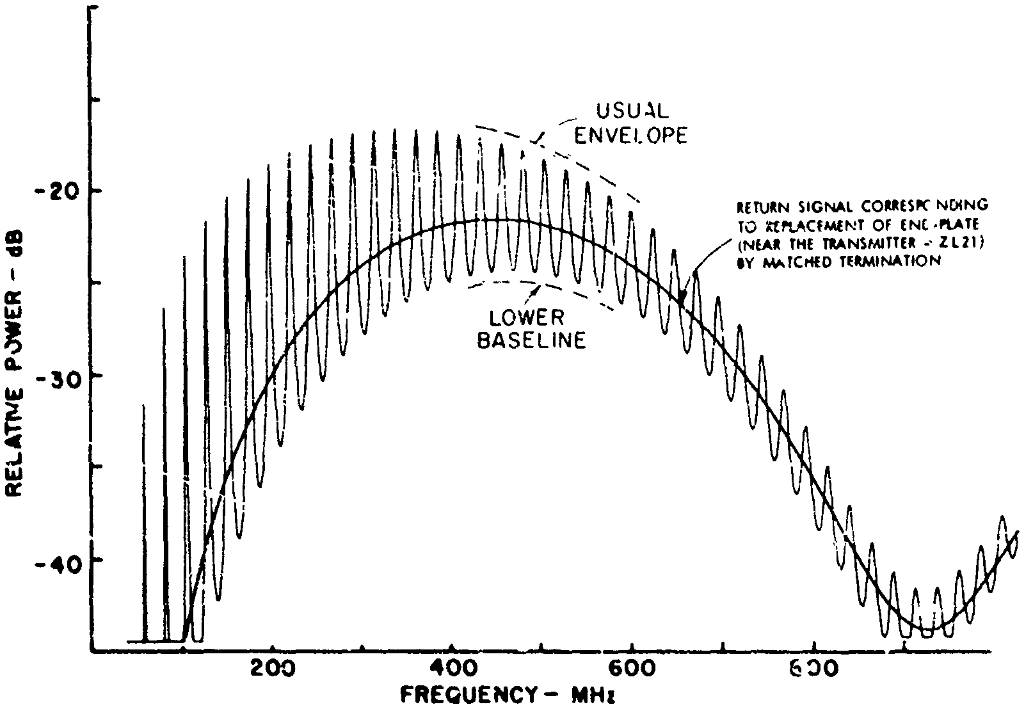
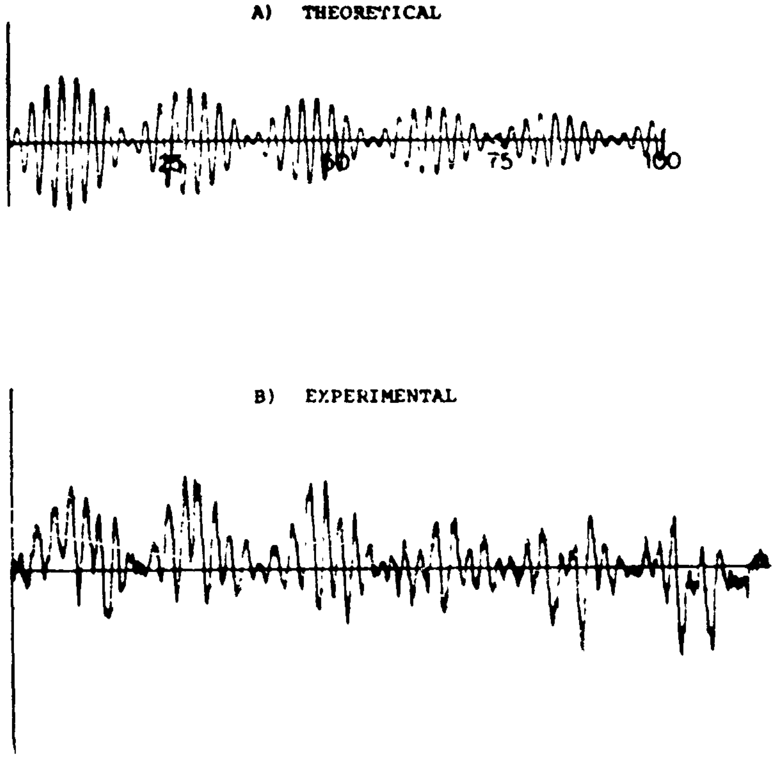


Fig.11 Theoretical signature in a lossy environment



COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS OBTAINED FROM A HUMAN TARGET IN A GUIDED RADAR SYSTEM

Fig.12 Theoretical and experimental guided radar return signal

DISCUSSION

E.R. Schmerling, US

Did you terminate the cables in the characteristic impedance of the unmodified cables, or in the impedance needed to obtain unity standing wave ratio of the modified coupled system?

Author's Reply

The impedance needed to obtain the lowest possible standing wave ratios was used

DEFINITION OF SUBSURFACE FEATURES BY GEOPHYSICAL PROBING*

R. J. Lytle
Lawrence Livermore Laboratory
Livermore, California 94550

ABSTRACT

This report provides an overview of recent advances in near surface geophysical probing. Work progresses in four areas. One of these areas is fundamental studies, which involves not only mathematical studies but also laboratory and controlled in-situ experiments. Empirical curves can be generated that relate remotely sensed observable to the parameters important for construction or other near surface applications. Another area is instrument development. For a sensing method to be effective and routinely used, the functioning device must be quick, easy, and cheap to use in the field. Examples in this report show the recent instrumentation development results. Data processing development is another progessing area, particularly for the field environment. Another area is data interpretation. Once the instrument system is functioning and empirical curves have been generated in fundamental studies, the capability exists to interpret the data. Data interpretation uses all the known information that applies to the site. Advances in systematic ways of routinely analyzing this data are presented.

1. INTRODUCTION

Electromagnetic methods are increasingly being used in remote sensing of the ground. Measurements are performed in the air, on the surface, in boreholes, and in tunnels and caverns. As this is a broad field, we consider herein only techniques probing to the distances of a few centimeters to a kilometer. This excludes techniques sensing only the immediate ground surface (i.e., the first few centimeters).

Applications of electromagnetic probing include geophysical prospecting (e.g., exploration for minerals, petroleum, water), geotechnical engineering (e.g., structural and foundation studies related to potential landslides, dam safety), geophysical diagnostics (e.g., monitoring energy processes in the earth such as in-situ coal gasification and in-situ oil shape retorting, geothermal reservoir assessment) and communication system studies (e.g., environmental effects, ground screen considerations, communication efficiency).

A number of books, overview articles and collected papers on the state-of-the-art of electromagnetic probing have been published in recent years (DAVIS, D. T., et al, 1979; DINES, K. A., et al, 1979; CALVERT, T. J., et al, 1977; FELLER, G. V., 1976; LYTLE, R. J., 1974; MORRISON, H. F., et al, 1976; NITSAN, U., et al, 1977; WAIT, J. R., 1971). Other papers have presented equipment innovations, improved measurement methods, and new data analysis methods. These aspects are briefly reviewed below.

2. MEASUREMENT TECHNIQUES

One can infer the subsurface constitutive parameters by using techniques such as (among others): resistivity, mutual impedance, wave tilt, surface impedance, field variation with distance, input impedance, borehole induction methods, borehole tunnel or surface transmission and reflection methods, transient probing with pulse transmission and reflection, parametric sounding, and geometric sounding. These methods are now classical. No basically new methods have been recently proposed. The major activities of those working of electromagnetic probing of the ground have been directed towards new applications of these methods, improving the modeling and interpretation capability, and refinements of the basic experiments.

3. RECENT ADVANCES IN FUNDAMENTAL STUDIES

With the interest in alternative energy means, there has been study of retorting oil in-situ from oil shale and gasifying coal in-situ, and then using the oil or gas as the energy source. Laboratory work has indicated that: for in-situ oil shale retorting, electromagnetic measurements may provide information such as grade of oil shale prior to retorting, amount of oil retorted, and nature and position of oil being produced; for in-situ coal gasification, electromagnetic measurements are useful in monitoring the progress and location of the coal burnfront (DAVIS, D. T., et al, 1979).

It has long been observed that rocks undergoing fracturing emit acoustical energy. An interesting physical phenomena has been the observance in the laboratory that quartz-bearing rocks also emit radio frequency energy as they are being fractured (NITSAN, U., 1977). This fracture-related emission has various potential geophysical applications ranging from monitoring cracks in rock deformation experiments, to studies of electromagnetic effects associated with earthquakes.

---

*This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore Laboratory under contract number W-7405-ENG-48.

4. RECENT ADVANCES IN INSTRUMENT DEVELOPMENT

Equipment development can permit more accurate and sensitive measurements and thus enhance the ability to detect surface anomalies. An an example, theoretical studies have indicated that it is feasible to probe to unprecedented depths of exploration by monitoring the low frequency (10 to 2,000 Hz) input resistance of a single superconducting coil (MORRISON, H. F., et al, 1976). To see whether this is feasible, research has determined that the super-conducting wires, large scale nonmetallic cryostats, the required measurement circuitry, and a data acquisition system for a superconducting coil are realizable in an airborne system. When deployed, this system will provide maps of absolute ground conductivity. These results will be useful in mineral exploration and other applications.

In borehole logging, it is important to distinguish between oil and water within the medium external to the borehole. By using a sufficiently high frequency so that propagation in the region surrounding a borehole is governed by displacement currents rather than conduction currents, this has been achieved. This approach measures the difference in time of arrival of a signal sent from a transmitter to two collinear receivers at the side of the borehole (CALVERT, T. J., et al, 1977). The dielectric constant of the fluid filling the pore space within the rock can then be inferred by using an independent measurement of the rock porosity. As water and oil respectively have relative dielectric constants of approximately 80 and 2, this permits a discernment of whether oil or water is present.

5. RECENT ADVANCES IN DATA PROCESSING

Before fielding an experiment, it is prudent to perform an experimental design and decide which experimental format is needed and what data sampling is required to adequately resolve the expected range of model parameters. This approach helps minimize experimental costs but still achieve the required resolution. A tutorial example of how to perform such an experimental design has recently been provided for a loop excitation of a layered earth (WARD, S. H., 1976). Parametric, geometric, and combined sounding formats were considered in the experimental design.

Geophysical techniques can be used to infer the structure between boreholes. The principle behind the techniques is that the properties of a medium are revealed by the effect of the medium on electromagnetic waves passing through it. The properties revealed by wave propagation from one borehole to another are the attenuation and the travel time (or velocity) of the waves. The basic data-collection procedure with two boreholes is shown in Figure 1. Many ray paths link the source locations with the receiver locations. With this combination of many source and receiver locations, one can effectively sample the region between the boreholes. Unfortunately, the many source and receiver locations generate much data. To interpret this data and present it meaningfully, we need a picture similar to the attenuation picture in Figure 2. Fortunately, the data-processing technology that has been developed for the medical industry (see Figure 3) can be adopted (DINES, K. A., et al, 1979).

The basic idea of the procedure used to infer the detailed picture of the region between boreholes is shown in Figure 4. A remotely sensed variable, say total attenuation between source and receiver, is the sum of the differential attenuation along the ray path linking source and receiver. For simplicity, three representative ray paths linking three different combinations of source and receiver are shown in Figure 4. The cell representation indicated has only four cells, denoted as cells 11, 12, 21, and 22. The local attentuation rate in cell ij is denoted as α_{ij}. The length of the kth ray path through cell ij is designated as ${}_kD_{ij}$. Using the equation format indicated in Figure 4, we can easily construct and solve a linear system of equations in terms of the local attenuation.

This format for data collection and processing was tried in a experiment titled Forest Glen, a site for a future Washington, D.C., Metro station. We collected enough data to give 3000 cells in the picture. The picture resulting from the data taken is shown in Figure 5, where the shading represents the cell attenuations in the sampled region. The darker region near the top left signifies a region of high attenuation rate. The lighter region near the bottom signifies a region of less attenuation. The dark region correlated well with a region in the borehole logs that indicated heavy fracturing and the light region correlated with a region of competent rock. These results indicate that one can detect lateral and vertical inhomogeneities in the region between boreholes.

A question that commonly arises in cross-borehole data interpretation is, "What is the effect of ray-bending?" The effect of ray bending can properly be accounted for by using ray optics. This enables one to still solve the inverse problem of finding a model that explains the experimental data. By using the computer to study the synthetic wave-velocity profile in Figure 6, an example has been constructed that helps illustrate the effect of considering ray bending. The velocities V_1 and V_2 in Figure 6 differ by 60%. By representing the region between boreholes with many cells, one can use an iterative data-inverse procedure to infer the bending ray paths and the velocity in each cell.

In Figure 7 are shown the presentations of the ideal profile, the interpreted profile ignoring the bending rays and assuming in the interpretation that the rays were straight-line paths linking source and receiver, and the interpreted profile accounting and solving for the bending rays. Note that the straight-line assumption gives a rough idea of the structure, and the bending-ray approach gives much better resolution. Thus, for some problems, it is possible to solve the inverse problem even with significant ray bending.

6. RECENT ADVANCES IN DATA INTERPRETATION

By combining data from different types of geophysical measurements, it is possible to obtain better resolution from the joint data set than is feasible from data obtained with only one method (VOZOFF, K., et al, 1975). For example, resistivity data provides accurate information for the resistivity-thickness product of a planar layer ground. For the same situation, magnetotelluric data provides good thickness information but limited resistivity information. By combining the resistivity and magnetotelluric data for a site adequately described by a planar layered model, it has been shown possible to obtain good resolution for both the resistivity and thickness of the individual planar layers.

7. SUMMARY

Electromagnetic measurements are meeting a wide variety of applications where knowledge of the near surface geophysical structure is required.

8. REFERENCES

DAVIS, D. T., LYTLE, R. J., and LAINE, E. F., 1979 (in press), Use of High Frequency Electromagnetic Waves for Mapping an In-Situ Coal Gasification Burn Front, In-Situ, Vol. 3, No. 2.

DINES, K A., LYTLE, R. J., August 1979 (in press), Computer Geophysical Tomography, Proceedings of the IEEE, VOl. 67, No. 8.

CALVERT, T. J., RAU, R. N., April 13-15, 1977, Electromagnetic Propagation - A New Dimension in Logging, paper presented at the 47th Annual California Regional Meeting of the Society of Petroleum Engineers of AIME, Bakersfield, California.

KELLER, G. V., Guest Editor, October 1976, "Special Issue on Electromagnetic Probing in Solid Earth Geophysics," IEEE Trans. Geoscience Electronics, Vol. GE-14, pp. 218-269.

LYTLE, R. J., July 1974, "Measurement of Earth Medium Electrical Characteristics: Techniques, Results, and Applications," IEEE Trans. Geoscience Electronics, Vol. GE-12, pp. 81-101.

MORRISON, H. F., DOLAN, W. M., and DEY, A., December 1976, "Earth Conductivity Determinations Employing a Single Superconducting Coil," Geophysics, Vol. 41, No. 6A, pp. 1184-1206.

NITSAN, U., August 1977, Electromagnetic Emission Accompanying Fracture of Quartz-Bearing Rocks, Geophysical Research Letters, Vol. 4, No. 8, pp. 333-336.

WAIT, J. R., 1971, Electromagnetic Probing in Geophysics, Golem Press, Boulder, Colorado.

WAIT, J. R. Guest Editor, April 1976, "Special Issue: Subsurface Telecommunications and Geophysical Probing" Radio Science, Vol. 11, pp. 233-418.

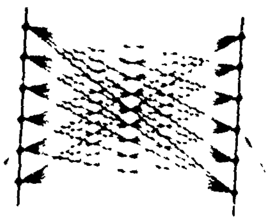
WAIT, J. R., 1978, guest editor of a special issue of the Proceedings of the IEEE, topic: Applications of Electromagnetic Theory to Geophysical Exploration, to appear in Volume 67, No. 8.

WARD, S. H., Editor, December 1976, "Special Electromagnetics Supplement," Geophysics, Vol. 41, No. 6A.

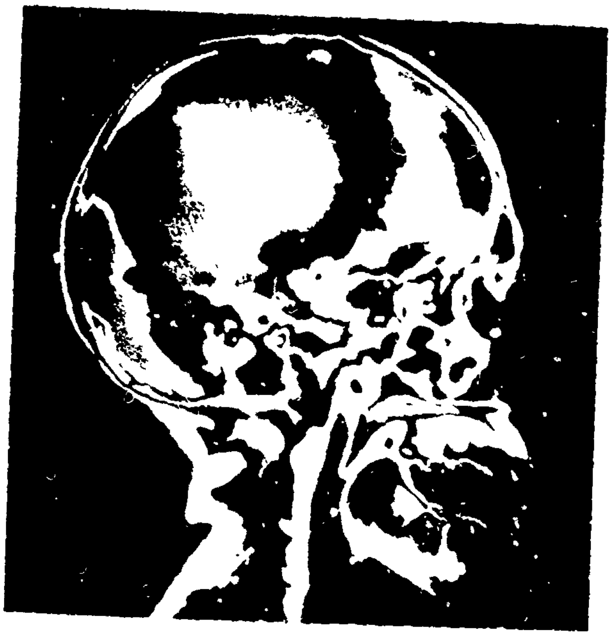
WARD, S. H., editor, January 1977, "Workshop on Electrical Methods in Geothermal Exploration," a report on a workshop financed by the U.S. Geological Survey under a grant to the University of Utah, Contract No. 14-08-00[illegible]-G-359, 182 pages, available from the Dept. of Geology and Geophysics, University of Utah, Salt Lake City, Utah.

VOZOFF, K., JUPP, D. L. B., 1975, "Joint Inversion of Geophysical Data, Geophy. J. R. Astro. Soc., 43, pp. 977-991.

Proceedings of the Workshop on Modeling of Electric and Electromagnetic Methods, May 17-19, 1978, Lawrence Berkeley Laboratory report LBL-7053, UC-66b, 203 pages, available from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161.



[illegible]



[illegible]

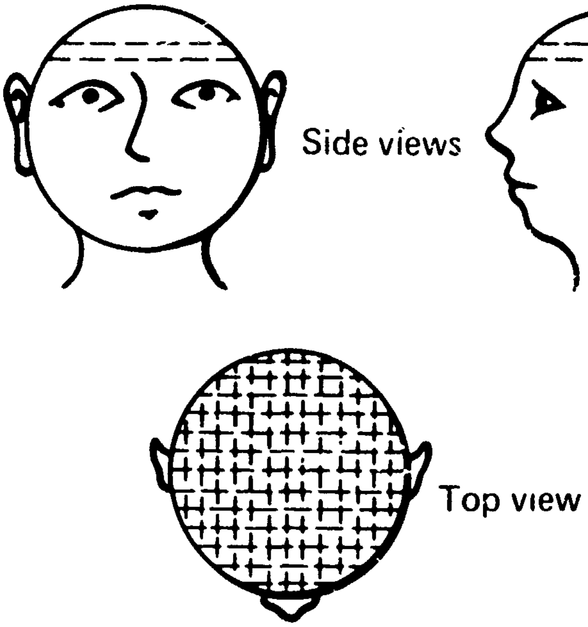
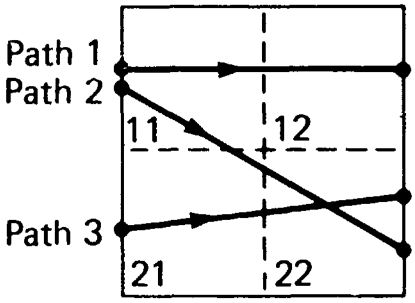


Fig 3 Geophysical probing can use data-reduction techniques developed for medical tomography, giving both side and top views of the region being probed



$$\mathrm{Atten}_1 = {}_1D_{11}\alpha_{11} + {}_1D_{12}\alpha_{12} + {}_1D_{21}\alpha_{21} + {}_1D_{22}\alpha_{22}$$

$$\mathrm{Atten}_2 = {}_2D_{11}\alpha_{11} + {}_2D_{12}\alpha_{12} + {}_2D_{21}\alpha_{21} + {}_2D_{22}\alpha_{22}$$

$$\mathrm{Atten}_3 = {}_3D_{11}\alpha_{11} + {}_3D_{12}\alpha_{12} + {}_3D_{21}\alpha_{21} + {}_3D_{22}\alpha_{22}$$

Fig 4 The total attenuation is the sum of the differential attenuations

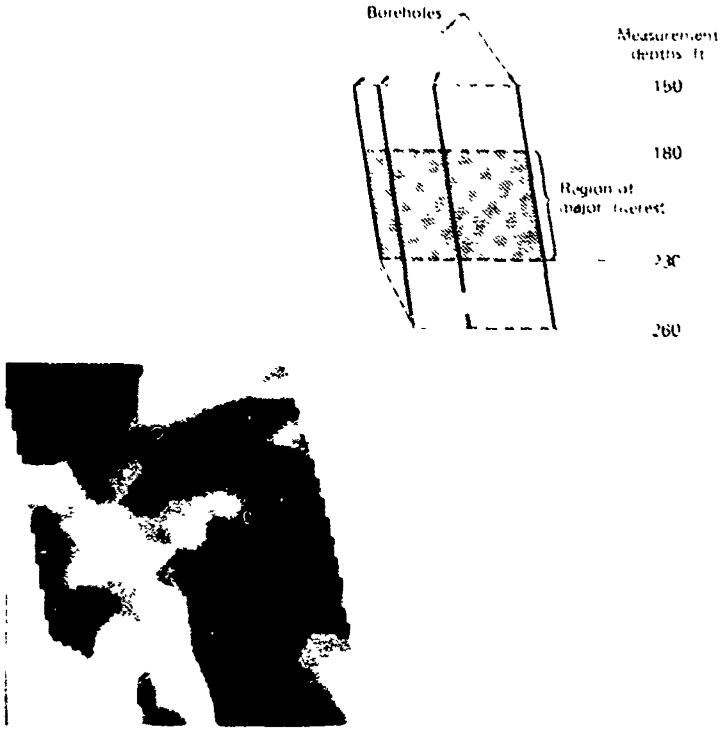

Boreholes
Measurement depths ft
150
180
Region of major interest
230
260

Fig. 5 The Forest Glen site of a future Washington Metro station was probed from four boreholes extending into the prospective station area. The four boreholes were all in a plane.

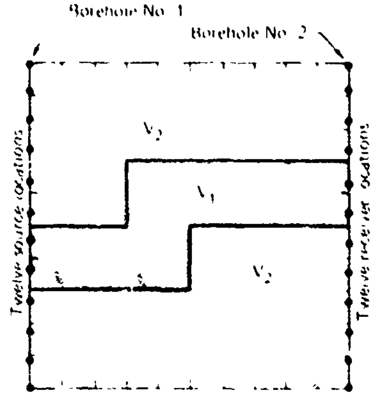

Borehole No. 1
Borehole No. 2
V_2
V_1
V_2
Twelve source locations
Twelve receiver locations

Fig. 6 Synthetic profile of regions with differing velocities V_1 and V_2. We use this profile to show the eff of considering ray bending in our algorithmic data interpretation.

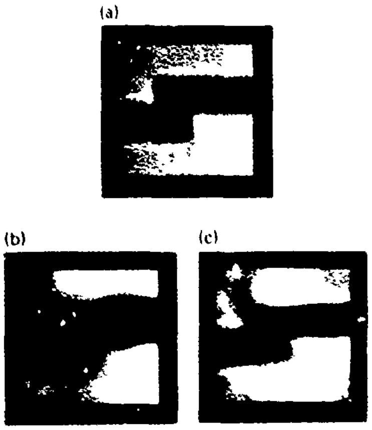


Fig. 7 Graphic representation (originally in color) of (a) the ideal model (same as Figure 6), (b) the straightline

DISCUSSION

I. Kaufman, US

(1) What frequencies have you used, and what determines the optimum frequency?

(2) How does your scheme compare with data obtained and cost of seismic methods of probing?

(3) Have you found any difficulties that are due to refraction of your e m. waves?

Author's Reply

(1) Frequencies of 1-500MHz have been used. Generally, the highest frequency that will propagate when high signal to noise ratio is used. However, exceptions occur such as in tunnel detection, where as low a frequency that will produce two significant minima is usually used

(2) To my knowledge, seismic methods have to date not been successfully used for the range of comparable applications. In terms of different observed effects, seismic methods are commonly pulse signals, whereas this electromagnetic method uses continuous-wave transmission

(3) For the situations we have considered, refraction has not been found to create insurmountable difficulties It can be appropriately accounted for theoretically using iterative ray bending methods A report describing the results obtained using this approach is forthcoming In addition, below the water table the variation of dielectric constant is usually not dramatic and has not presented undue problems as yet

RESULTATS EXPERIMENTAUX CONCERNANT LA PROPAGATION LIBRE DES ONDES U.H.F. DANS LES TUNNELS

P. DEGAUQUE, P. MANGEZ, B.DEMOULIN et R. GABILLARD

Universite de Lille, Departement Electronique, B.P. 36, 59650 VILLENEUVE D'ASCQ
FRANCE

RESUME

Dans des galeries d'exploitation minière ou en general dans certains tunnels, il peut être utile d'etablir des liaisons radio sur quelques centaines de metres, si possible à l'aide d'appareils portatifs. Pour cela une radiocommunication dans la gamme haute frequence (400 MHz - 1 GHz) semble interessante. En effet, dans cette bande de frequences, la longueur d'onde est assez petite par rapport aux dimensions de la galerie pour que les modes dominants qui vont se propager correspondent à des rayons qui se reflechissent sous incidence rasante. Dans ce cas, seule une faible partie de l'energie est perdue par refraction sur les parois. Dans cet article, nous donnons les affaiblissements lineiques obtenus pour differents types de tunnel. Nous envisageons tout d'abord le cas des tunnels rectilignes et nous mettons experimentalement en evidence l'influence de la geometrie transversale de la galerie et de la rugosite des parois. Nous donnons ensuite les affaiblissements supplementaires subis par l'onde lorsque le tunnel comporte une ou plusieurs courbes.

1. INTRODUCTION

Pour realiser une telecommunication en tunnels à l'aide d'emetteurs recepteurs portatifs, plusieurs solutions sont possibles dans differentes gammes de frequences. Tout d'abord la zone des tres basses frequences (quelques kHz) est souvent utilisee pour transmettre des informations directement à travers le sol, les antennes d'emission et de reception etant des boucles magnetiques. Les telecommunications par courant porteur à des frequences inferieures à quelques MHz ou au dela du MHz à l'aide de coaxiaux à pertes, se sont egalement developpees tant du point de vue theorique qu'experimental, (HILL, D.A. and WAIT, J.R., 1976 ; DELOGNE, P., 1976 ; DEGAUQUE, et al., 1976.....) Cependant la pose d'un cable est, dans certains cas, soit impossible soit ne se justifie pas si les liaisons radio à assurer sont occasionnelles. Si la portee desiree est de quelques centaines de metres on peut envisager une emission à des frequences suffisamment elevees pour que les ondes soient guidees par les parois du tunnel. Les premiers travaux ont montre que la frequence de transmission devait être superieure à 400MHz pour eviter des pertes trop importantes qui se produisent lorsque la frequence d'emission est proche de la frequence de coupure du guide d'ondes.

Après un bref rappel des approches theoriques possibles, nous donnons dans cet article les resultats experimentaux concernant les affaiblissements lineiques obtenus dans des tunnels rectilignes. En effet, le champ électrique decroit sous une forme quasiment exponentielle en fonction de la distance emetteur-recepteur et il est donc possible d'exprimer le coefficient d'attenuation en dB/100m. Nous envisageons successivement differents types de tunnel, betonnes ou non afin de mettre en evidence l'influence de la section de la galerie et de la rugosite des parois. Nous donnons ensuite l'attenuation supplementaire subie par l'onde lorsque le tunnel comporte une ou plusieurs courbes.

2. PRELIMINAIRE THEORIQUE

Dans la gamme U.H.F. que nous envisageons (400MHz - 1GHz), la plupart des roches se comportent comme un dielectrique de constante dielectrique relative de 5 à 10 et de conductivite de l'ordre de 10^{-2} mho/m. Ceci suggère donc de modeliser les galeries de mine par un guide dielectrique creux de section de forme simple se rapprochant de celle d'une galerie reelle. Cette modelisation permet d'etablir les expressions approchees donnant les pertes de propagation et de mettre en evidence les parametres importants dont dependent ces pertes dans une galerie de mine. Nous envisageons successivement le cas d'un guide circulaire et d'un guide rectangulaire.

2.1) Propagation dans un guide circulaire diélectrique creux

Soit un cylindre circulaire de rayon a constitue par un dielectrique de caracteristique $\varepsilon_1, \mu_1, \sigma_1$ entoure d'un milieu dielectrique infini de caracteristiques $\varepsilon_0, \mu_0, \sigma_0$, les differents modes de propagation sont solutions d'une équation modale implicite (STRATTON, J.A., 1961). Dans le cas des galeries de mine, le milieu 1 est de l'air (ε_0 μ_0) et le milieu 2 est un dielectrique non magnetique ($\varepsilon^* = \varepsilon_r + i\sigma/\omega, \mu_0$). MARCATILI et SCHMELTZER (1964) ont propose une solution approchée de l'equation modale valable pour les modes inferieurs qui ont une constante de propagation h proche de celle de la propagation libre dans l'air et pour un rayon a de la galerie beaucoup plus grand que la longueur d'onde λ. En effet l'équation modale s'ecrit dans ce cas :

$$J_{n-1}(\lambda_1 a) = i\,\nu_n \frac{\lambda_1}{k_0} J_n(\lambda_1 a) \qquad (1) \quad \text{avec } \lambda_1 = \sqrt{k_0^2 - h^2} \text{ et } k_0 = \omega\sqrt{\varepsilon_0\mu_0}$$

De plus $\nu_n = (\varepsilon^*/\varepsilon_0 - 1)^{-1/2}$ pour les modes TE_{om}, $\nu_n = \varepsilon^*/\varepsilon_0\,(\varepsilon^*/\varepsilon_0 - 1)^{-1/2}$ pour les modes TM_{on}

et $\nu_n = \frac{1}{2}(\varepsilon^*/\varepsilon_0 + 1)(\varepsilon^*/\varepsilon_0 - 1)^{-1/2}$ pour les modes EH_{nm}.

La resolution approchee de l'equation (1) permet de deduire l'expression donnant le taux d'attenuation du mode (n,m) :

$$\alpha = I_m(h) = \left(\frac{u_{nm}}{2\pi}\right)^2 \frac{\lambda^2}{a^3} \, Re(\nu_n) \qquad (2)$$

où u_{nm} est le m^{ieme} racine de la fonction de Bessel d'ordre n-1. Les modes à faible perte donc les modes inferieurs.

On remarque de plus que le taux d'attenuation est proportionnel au carre de la longueur d'onde λ et inversement proportionnel au cube de la dimension de la galerie. Il apparait donc qu'en augmentant la fréquence de travail nous pouvons minimiser les pertes de propagation et que les galeries de grande dimension sont favorables à ce type de propagation. Enfin, les applications numeriques montrent que les pertes dues à la conductivité des parois sont negligeables par rapport aux pertes par refraction dans le dielectrique, fonction de la permittivite relative ε_r (GLASER, J.I., 1969).
A 1 GHz, les pertes sont pratiquement proportionnelles à $1/\sqrt{\varepsilon_r}$ pour les modes TE_{om} et à $\sqrt{\varepsilon_r}$ pour les modes TM_{om} et EH_{nm}. Les modes TE_{01} et TM_{01} exigeant une symetrie de revolution, le mode EH_{11} dont les lignes de champ électrique sont paralleles entre elles est un mode à faibles pertes susceptible d'être excite dans la galerie.

2.2. Propagation dans un guide rectangulaire creux

Le modele de guide qui s'approche le plus d'une galerie de mine reelle est probablement le guide dielectrique rectangulaire creux. Cependant, il n'est pas possible d'etablir analytiquement l'equation modale de ce type de guide à cause de la difficulte presentee par les conditions aux limites sur les parois. Une première simplification consiste à envisager une galerie possedant deux parois opposees constituees par un conducteur parfait (ADAM, B. et KNEUBUHL, F., 1975) les modes LSE et LSM (longitudinal section electric et magnetic) susceptibles de s'y propager presentent des coefficients d'affaiblissement voisins de ceux trouves dans l'approximation du guide circulaire.

D'autres auteurs tels que LAAKMANN et STEIER (1976) ont etudie le guide rectangulaire dans l'hypothèse des faibles pertes et pour des hauteur et largeur, notees d_1 et d_2, beaucoup plus grandes que la longueur d'onde. Les modes EH_{nm} qui peuvent se propager sont reperes avec un indice supplementaire v ou h suivant que le champ électrique a sa composante dominante verticale ou horizontale. L'analyse modale faite en satisfaisant uniquement les conditions aux limites sur les quatres parois sans tenir compte des discontinuites aux coins donne une solution d'une précision satisfaisante lorsque les pertes sont faibles. Les pertes de propagation pour les modes EH^v et EH^h donnees respectivement par :

$$\alpha^v = \frac{m^2}{2} \frac{\lambda^2}{d_2^3} Re\left\{\frac{\varepsilon_2^*/\varepsilon_0}{\sqrt{\varepsilon_2^*/\varepsilon_0 - 1}}\right\} + \frac{n^2}{2} \frac{\lambda^2}{d_1^3} Re\left\{\frac{1}{\sqrt{\varepsilon_1^*/\varepsilon_0 - 1}}\right\} \qquad (3)$$

$$\alpha^h = \frac{m^2}{2} \frac{\lambda^2}{d_2^3} Re\left\{\frac{1}{\sqrt{\varepsilon_2^*/\varepsilon_0 - 1}}\right\} + \frac{n^2}{2} \frac{\lambda^2}{d_1^3} Re\left\{\frac{\varepsilon_1^*/\varepsilon_0}{\sqrt{\varepsilon_1^*/\varepsilon_0 - 1}}\right\} \qquad (4)$$

ε_1^* est la permittivité des parois laterales tandis que ε_2^* correspond à la permittivite du toit et du plancher de la galerie. Ces resultats concordent avec ceux etablis par EMSLIE et LAGACE (1975). Ces auteurs montrent également que les pertes ohmiques sont negligeables devant les pertes par refraction tant que l'angle de perte reste faible (< 10^{-1}). Si $\varepsilon_1^* = \varepsilon_2^*$ c'est des deux modes [illegible] le mode de perte est le mode qui a son champ électrique parallèle au plus grand [illegible] de la galerie. [illegible] *présente un maximum de champ électrique au centre de la galerie. Comme dans le cas des guides creux les pertes sont proportionnelles à $\lambda^2 d^3$.*

Enfin une approche utilisant la théorie des raies a ete faite par S.F. MAHMOUD et J.R. WAIT (1974). Elle est très intéressante en haute fréquence car elle offre l'avantage d'éviter le calcul des couplages entre les nombreux modes qui peuvent se propager. Une comparaison avec les resultats obtenus par la théorie modale lorsque deux parois sont parfaitement conductrices a d'ailleurs ete faite par ces auteurs. Les courbes theoriques font apparaître des phénomènes de battement mais la decroissance moyenne de l'amplitude du champ en fonction de la distance émetteur-recepteur est sensiblement exponentielle. L'effet de la rugosite et de l'inclinaison des murs est beaucoup plus difficile à prendre en compte. En modifiant les coefficients de reflexion de Fresnel du à la rugosite, S.F MAHMOUD et J.R. WAIT ont montré que l attenuation supplementaire due à la rugosité augmentait en fonction de la fréquence mais que l'affaiblissement total restait une fonction decroissante de la frequence. L'inclinaison des parois provoque par contre une attenuation supplementaire importante en haute frequence (LAGACE,R.L. et al., 1975).

Compte tenu de la difficulte theorique rencontree dans la modelisation d'une galerie de mine, de nombreux essais experimentaux sont necessaires à la comprehension des phenomènes. Les premières etudes ont ete entreprises par GODDARD (1973), CRARY (1972) et recemment par J. CHIBA et al.(1978), L. DERYCK (1978).

Dans cet article nous completons les travaux precedents en donnant les resultats experimentaux obtenus pour differents types de galeries dans la bande de fréquences comprise entre 400 MHz et 1000 MHz. Nous envisageons successivement la propagation dans une galerie rectiligne et l'influence des courbures.

3. PROPAGATION DANS UNE GALERIE RECTILIGNE

L'antenne dipole d'emission fixe est alimentée par un generateur UHF délivrant une puissance de 10 Watts. L'antenne de reception mobile est branchée à un mesureur de champ par l'intermediaire d'un cable coaxial de 10 m de long pour ne pas perturber la mesure par la présence des opérateurs. L'approche théorique ayant montré que le champ electrique est maximum au centre de la galerie, les dipoles d'emission et de reception sont toujours placés dans cette condition. Nous envisageons d'abord la propagation dans une galerie betonnee et nous étudierons ensuite l'influence des rugosités et de la géometrie de la section droite du tunnel.

3.1. Propagation dans une galerie betonnee

La première galerie envisagee, representee sur la figure 1a, est de section rectangulaire. Elle est caracterisée par une largeur de 3.4m et une hauteur de 2.3m. Seul le plancher de la galerie n'est pas betonne mais le sol y est relativement lisse. Les courbes de la figure 1b représentent la variation du niveau reçu N en db au dessus du µV en fonction de la distance x emetteur-recepteur. Comme nous comparons uniquement des affaiblissements nous n'avons pas tenu compte du gain de l'antenne de reception. La fréquence d'emission etant de 1000 MHz, differentes orientations d'antennes ont ete essayees. Elles sont reperees par deux lettres H (horizontal) et V (vertical), la première lettre correspondant à la position de l'antenne d'emission et la deuxième à la position de l'antenne de reception. On remarque d'apres les courbes que l'affaiblissement en polarisation VV ou HH est pratiquement le même, 14 db/100m dans ce cas. De plus, même après 150m, l'onde reste polarisee puisque, en orientation croisee VH et HV, le niveau reçu est d'environ 20 db inferieur à celui obtenu lorque les antennes d'emission et de reception sont parallèles. Sur la figure 2, nous avons trace les courbes N(x) pour trois frequences à 450 MHz, 750 MHz et 1000MHz et pour les configurations HH et VV. Le tableau ci dessous resume les affaiblissements obtenus :

Frequence / Orientation	450 MHz	750 MHz	1000 MHz
HH	33 db/100m	20 db/100m	14 db/100m
VV	52 db/100m	23 db/100m	14 db/100m

Tableau 1

A 1GHz, l'affaiblissement mesure est superieur à celui que prevoit la theorie modale approchee pour le mode a plus faible perte (5 à 10 db/100m, suivant l'orientation de l'antenne d'emission). De plus cet affaiblissement devient pratiquement independant du type de polarisation, horizontale ou verticale. A 450MHz, on remarque que l'affaiblissement en polarisation verticale est bien superieur à celui obtenu en polarisation horizontale. Il est donc interessant d'examiner l'amplitude du niveau reçu pour des configurations d'antennes croisees VH ou HV.

Sur la figure 3a, correspondant à une antenne d'emission verticale, nous avons represente l'affaiblissement supplementaire du à une reception en polarisation croisée. Cette courbe $N_{VH} - N_{VV} = f(x)$ montre que la composante diffusée excite progressivement un mode de polarisation horizontale. Au bout d'une centaine de metres, ce mode devient d'ailleurs preponderant, le niveau reçu en configuration croisee VH etant superieur à celui obtenu en configuration parallele VV. La courbe de la figure 3b, $N_{HV} - N_{HH} = f(x)$, obtenue donc lorsque l'antenne d'emission est horizontale, montre que l'onde reste polarisee horizontalement, l'affaiblissement supplementaire en configuration croisée etant de 12 à 20 db.

A une distance de 100m de l'émetteur et pour une fréquence d'émission de 750 MHz, l'amplitude du champ en différents points de la galerie a été relevée (figure 4). Nous constatons que le champ est maximum au centre du tunnel et que l'écart d'amplitude entre le champ au centre et celui mesuré à 30 cm des parois [illegible] 11db pour l'orientation HH et de 16 db pour l'orientation VV. Nous remarquons également que les tuyaux et [illegible]bles électriques n'introduisent qu'une légère modification de la carte de champ dans leur proche voisi[illegible]. variation de l'ordre de 3db.

Une deuxième série de mesures a été effectuée dans un couloir souterrain dont la section droite rectangulaire est plus faible que dans le cas précédent (figure 5), sa largeur étant de 2m et sa hauteur de 2.8m. Les courbes de la figure 5 donnent les variations de l'amplitude du signal reçu en fonction de la distance émetteur-récepteur. Les affaiblissements obtenus sont les suivants :

Fréquence / Orientation	450 MHz	750 MHz	1000 MHz
HH	67 db/100m	40 db/100m	34 db/100m
VV	66 db/100m	38 db/100m	31 db/100m

Tableau 2

On remarque que même à la fréquence la plus élevée, l'atténuation est beaucoup plus importante que celle trouvée dans la première galerie. Ceci peut s'expliquer par les dimensions plus faibles de la section droite du couloir souterrain, les formules approchées montrant un affaiblissement inversement proportionnel au cube des dimensions transversales.

Un troisième type de galerie botonnée où des mesures ont été faites se caractérise par un plafond vouté (figure 6). La largeur de la galerie est de 3.1m, sa hauteur au centre de 2.7m et au niveau des parois latérales elle n'est plus que de 2m. Contrairement aux deux galeries précédentes, le sol est très humide et parfois même recouvert d'eau. Les courbes d'amplitude du champ reçu sont reportées sur la figure 6 et les taux moyens d'attenuation sont donnés dans le tableau 3 :

Fréquence / Orientation	450 MHz	750 MHz	1000 MHz
HH	25db/100m	16db/100m	14db/100m
VV	40db/100m	19db/100m	16db/100m

Tableau 3

A 1GHz, on remarque que les attenuations sont du même ordre de grandeur que celles obtenues dans la première galerie ayant une surface de section droite voisine. La forme géométrique de la section ne semble donc pas être un paramètre très critique sur la valeur du coefficient d'atténuation en haute fréquence.

3.2. Propagation dans une galerie dont les murs présentent une certaine rugosité.

Deux séries de mesures ont été effectuées dans des galeries de mine de charbon. Dans le premier cas envisagé, le tunnel, ayant une section moyenne de forme rectangulaire, est pratiquement à l'etat brut c'est-à-dire qu'il comporte très peu de soutènements ni cadres métalliques ou rondins de bois. Sa largeur moyenne est de 4.2m et sa hauteur de 2.8m. Les irrégularités des parois laterales et du plafond sont très variables le long de la galerie, la valeur moyenne de la rugosité a pour ordre de grandeur ≃ 15cm. Les variations du signal reçu en fonction de la distance émetteur-recepteur sont représentées sur la figure 7 pour les fréquences de 450MHz et 1000MHz. Différents obstacles tels que présence de wagonnets...n'ont pas provoqué de fluctuations très importantes du signal dans leur voisinage car leurs dimensions transversales étaient beaucoup plus petites que celles de la galerie. De plus ces obstacles localisés ne dépassaient pas 5m de longueur. Les atténuations moyennes obtenues ont été reportées dans le tableau 4 :

Fréquence / Orientation	450 MHz	1000 MHz
HH	23 db/100m	16 db/100m
VV	45 db/100m	2[illegible] db/100m

Tableau 4

Bien que la surface de la section droite de la galerie ($11 m^2$) soit plus importante que celle de la première galerie ($7,8 m^2$), son coefficient d'affaiblissement à 1 GHz est plus important. Ceci peut s'expliquer par la rugosité importante des parois. La différence de permittivité relative entre le charbon et le béton peut aussi jouer un rôle. Il serait évidemment souhaitable de disposer de deux galeries tout à fait semblables mais dont les parois soient de natures différentes. Pour une fréquence d'émission de 1GHz, nous donnons sur la figure 8 l'affaiblissement supplémentaire lorsque les antennes d'émission et de réception sont croisées. Comme dans le cas de la première galerie étudiée, on remarque qu'avec une excitation horizontale, l'onde reste polarisée horizontalement ($N_{HV} - N_{HH} \simeq -15$ db). Par contre, si l'antenne d'émission est verticale, la polarisation de l'onde se modifie et après une distance de propagation de l'ordre de 100m, l'onde devient dépolarisée ($N_{VH} \simeq N_{VV}$)

Une deuxième configuration correspond à celle fréquemment rencontrée en mine. Il s'agit d'un tunnel sensiblement semi-circulaire, soutenu par des cadres métalliques placés tous les 50cm. L'épaisseur de ces cadres métalliques suivant la section droite de la galerie est d'environ 10cm. Entre chaque cadre, un réseau de fils métalliques parallèles espacés de 10cm évite les éboulements des blocs de petite taille. L'irrégularité moyenne des parois derrière ces grillages est de l'ordre de ~ 5cm. L'humidité qui règne à l'intérieur du tunnel est très importante car elle est voisine de la saturation en eau.

Nous avons tout d'abord reporté sur la figure 9 l'amplitude du signal reçu à 1000MHz pour différentes configurations d'antennes : HH, HV, VH. On remarque que pour ce type de galerie, l'onde est presque totalement dépolarisée, l'orientation du dipole récepteur dans un plan de section droite ayant peu d'importance. Pour ne pas surcharger la figure, nous n'avons pas reporté les points correspondant à la configuration VV car ils se situent également dans le nuage formé par les points précédents. Une autre représentation possible est donnée sur la figure 10. Tout d'abord la courbe de la figure 10 a) montre que le niveau reçu en HH est en moyenne supérieure de 3db à celui reçu en configuration VV. On note sur les figures 10b et 10c que l'onde émise en polarisation verticale est dépolarisée après une distance de 100m, tandis que celle émise en polarisation horizontale ne reste que très faiblement polarisée ($N_{HV} - N_{HH} \simeq -5$db). Nous avons tracé sur la figure 11 les variations du signal reçu pour les configurations HH et VV avec la fréquence comme paramètre. Les affaiblissements sont les suivants:

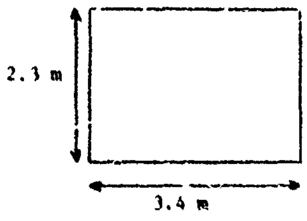
Fréquence Configuration	450 MHz	750 MHz	1000 MHz
HH	56 db/100m	35 db/100m	20 db/100m
VV	55 db/100m	37 db/100m	20 db/100m

Tableau n° 5

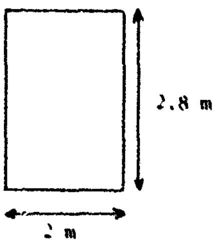
Quelque soit la fréquence, il n'y a pas de polarisation privilégiée donnant un coefficient d'atténuation minimum. Enfin une carte de champ électrique dans un plan transversal de la galerie a été faite à une distance d'environ 170m de l'émetteur. L'amplitude des champs en différents points est indiquée sur la figure 12. Les mesures ont d'abord été faites en maintenant l'antenne de réception dans le plan transversal de la galerie et en l'orientant dans ce plan de manière à capter le maximum de champ électrique. Les résultats montrent que l'amplitude du champ reste relativement constante, le minimum de champ se produisant au voisinage des parois avec une atténuation maximum d'environ 10 db par rapport au champ existant au centre de la galerie. Sur cette figure 12, l'amplitude de la composante longitudinale du champ électrique est également indiquée. Pour cela l'antenne de réception a donc été maintenue parallèle à l'axe de la galerie. Cette composante longitudinale est, bien sur, plus faible que la composante transversale.

3.3. Conclusion

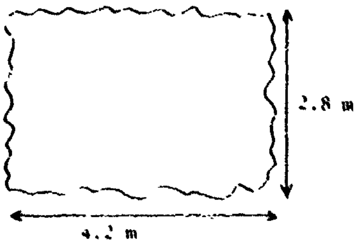
En conclusion de cette première partie, nous avons reporté les affaiblissements obtenus à 1000 MHz pour les différents types de galeries étudiés (tableau 6)



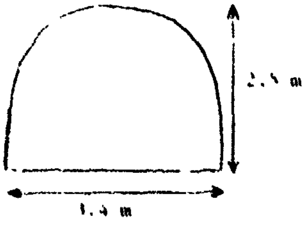
(1)



(2)



(3)



(4)

Types de tunnel / Position des antennes	(1)	(2)	(3)	(4)
HH	14 db/100m	34db/100m	16 db/100m	20 db/100m
VV	14 db/100m	31 db/100m	21 db/100m	20 db/100m

Tableau 6

Nous remarquons que , pour les galeries envisagées, le taux d'atténuation à 1 GHz dépend peu du plan de polarisation de l'onde, le facteur le plus important étant la surface de section droite de la galerie. Au cours de l'étude, nous avons vu également que si l'antenne d'émission est parallèle au plus petit coté de la galerie, l'onde se dépolarise rapidement et même dans certains cas le niveau reçu est plus important en configuration croisée qu'en configuration parallèle. Bien que pour les sections de tunnels étudiés jusqu'à présent, l'affaiblissement à 1000MHz soit encore important, il est possible d'établir une liaison sur quelques centaines de mètres, notamment pour les types de galeries (1), (3) et (4). Enfin un avantage de ce type de propagation réside dans le fait que le champ est maximum au centre de la galerie, c'est à dire là où est généralement effectuée la réception et que, de toute façon, l'affaiblissement supplémentaire du à une réception en un point quelconque de la section ne dépasse pas 10db, sauf à proximité immédiate des parois.

4. INFLUENCE D'UNE COURBE

Le premier tunnel que nous avons décrit lors de l'étude de la propagation en ligne droite change d'orientation à son extrémité, le rayon de courbure étant d'environ 20 m (Figure 13a). Dans cette zone, le tunnel a une section droite en forme de demi-cercle et comporte de nombreuses armatures métalliques. Le reste de la galerie est bétonnée, comme nous l'avons déjà indiqué, et de forme rectangulaire. L'émetteur a été placé à une distance de 25 m du début de la courbe et la variation d'amplitude du niveau reçu, à la fréquence de 1000 MHz et pour les 2 orientations VV et HH est représentée sur la figure 13a. On remarque tout d'abord que cette courbure, de 30 m de long, provoque une atténuation de 10 db pour l'orientation VV et de 15 db pour l'orientation HH. Puis, dans la partie rectiligne de la galerie, l'onde se propage avec un taux d'atténuation très voisin de celui mesuré lors de l'étude de la propagation en ligne droite (14 db/100m).

Le deuxième type de tunnel dans lequel des mesures ont été faites est celui dont la section droite a approximativement la forme d'un demi-cercle et qui comporte des soutènements métalliques sur toute sa longueur. On avait vu dans ce cas, qu'à 1000MHz, l'affaiblissement est de 20 db/100m en ligne droite. Le coude fait par la galerie a un rayon de courbure de 50 m (Figure 13b). On voit sur cette figure que l'atténuation due à cette courbe de 60 m de longueur est de 27 db pour la configuration HH et de 40 db pour la configuration VV. Le coefficient d'affaiblissement linéique après la courbe est de 20 db/100m, donc identique à celui mesuré initialement en propagation rectiligne. La comparaison de ces résultats montre que l'orientation optimum de l'antenne d'émission dépend du type de tunnel et de courbe envisagée. *De plus, l'atténuation supplémentaire provoquée par un changement d'orientation du tunnel d'environ 80° est de 25 à 30 db.*

Enfin, il est intéressant de savoir si l'onde reste polarisée après le passage dans une partie courbe de tunnel. Pour cela, nous avons repris la première configuration décrite (galerie rectangulaire bétonnée) car l'étude faite en propagation rectiligne nous avait montré qu'à 1000 MHz, l'onde restait polarisée sur une grande distance (cf. Figure 1). Les courbes de la figure 14 représentent donc respectivement la différence entre les niveaux reçus en polarisation parallèle : $N_{VV}-N_{HH}$ et en polarisation croisée : $N_{VH}-N_{VV}$ et $N_{HV}-N_{HH}$. Ces courbes montrent d'a qu'après le virage, le champ reçu en configuration VV devient supé- supérieur de 10 db à celui reçu e. figuration HH. Si l'émission se fait avec une antenne verticale, la polarisation reste marquée ($N_{VH}-N_{VV}$ - 10 db). Par contre, si l'antenne d'émission est horizontale, l'onde se dépolarise totalement. Ceci rejoint une conclusion faite lors de l'étude de la propagation en galerie droite : si le niveau reçu dans une configuration parallèle (HH dans ce cas) est inférieur à celui reçu dans l'autre configuration (VV), l'onde émise avec l'antenne horizontale se dépolarise, la réception en configuration croisée HV pouvant même devenir préférable.

5. PROPAGATION EN GALERIE NON RECTILIGNE

Le tracé et la coupe de la galerie sont représentés sur la figure 15, ce tunnel présente donc des coudes plus ou moins importants et l'antenne d'émission a été placée à une distance de 40 m de la première courbe. On voit sur la figure 15 qu'il est possible de donner une valeur moyenne de l'affaiblissement linéique. A 450 MHz, cet affaiblissement est de 70 db/100m contre 50 db/100m à 1000 MHz. A titre de comparaison, une galerie rectiligne de section à peu près identique provoque une atténuation de 56 db/100m à 450 MHz et de 20 db/100m à 1000MHz. L'étude des niveaux reçus en fonction de l'orientation des antennes montre que la polarisation horizontale est un peu plus favorable que la polarisation verticale.

6. CONCLUSION

Nous avons étudié dans cet article la propagation libre des ondes électromagnétiques dans des tunnels et dans une gamme de fréquences comprise entre 450 MHz et 1000MHz. Nous avons vu que dans cette bande de fréquences le coefficient d'affaiblissement diminue lorsque la fréquence augmente. Pour une galerie ayant une section droite d'environ 10 m^2, l'atténuation à 450 MHz est très importante, de 30 à 40 db/100m. Par contre, à 1000 MHz, l'établissement d'une liaison radio sur quelques centaines de mètres est envisageable, l'atténuation moyenne étant comprise entre 15 et 20 db/100 m. Si la section droite est plus faible, l'affaiblissement augmente rapidement. Pour une section d'environ 5 m^2 par exemple, l'atténuation à 1000 MHz est comprise entre 30 et 35 db/100 m. Nous avons vu également qu'un changement d'orientation du tunnel de 90°, se faisant avec un rayon de courbure de 20 m, produit une attenuation supplémentaire à 1 GHz de 20 à 30 db. Enfin, pour un type de galerie non rectiligne, nous avons montré que l'atténuation peut atteindre 50 db/100m.

Nous poursuivons actuellement ces expérimentations dans d'autres types de tunnels, tunnels routiers notamment, présentant une section beaucoup plus importante. Nous étudions également l'influence d'obstacles localisés tels que wagonnets, engins ... et le couplage qui se produit à l'intersection de deux galeries.

7. REMERCIEMENTS

La plupart des études expérimentales ont été faites sous contrat avec le Centre d'Etudes et de Recherches des Charbonnages de France. Nous remercions les différentes personnes ayant participé à ces travaux ainsi que le personnel des Bassins des Houillères du Nord-Pas-de-Calais et de Provence pour leur coopération.

REFERENCES

ADAM, B. et KNEUBUHL, F., 1975,"Transversaly excited 337 µm HCN waveguide laser", Applied Physics, vol. 8, pp. 281-291.

CHIBA, J., INADA, T., et KUWAMOTO, Y., 1978, "Radio Communication in tunnels", I.E.E.E. Trans. on MTT, vol. 26, 8, pp. 1-5.

CRARY, J.H., 1972, "Determination of the electromagnetic environment in coal mines", Final Tech. Rept., ITS, Boulder, Colorado.

DEGAUQUE, P., DEMOULIN, B., FONTAINE, J., et GABILLARD, R., 1976, "Theory and experiment of a mobile radio communication in tunnels by means of a leaky braided coaxial cable", Radio Science, vol. 11, 4, pp. 305-314.

DELOGNE, P., 1976, "Basic mechanisms of tunnel propagation", Radio Science, vol. 11, 4, pp. 295-304.

DERYCK, L., 1978, "Natural propagation of electromagnetic waves in tunnels", I.E.E.E. Trans. on Veh. Tech., vol. 27, 3, pp. 145-150.

EMSLIE, A.G., LAGACE, R.L., et STRONG, P.F., 1975, "Theory of the propagation of UHF radio waves in coal mine tunnels", I.E.E.E. Trans. on Ant. Prop., vol. 23, 2, pp. 192-205.

GLASER, J.I., 1969, "Attenuation and guidance of modes in hollow dielectric waveguides", I.E.E.E. Trans. on M.T.T., pp. 173-174.

GODDARD, A.E., 1973, "Radio propagation measurements in coal mines at UHF and VLF", Thru the earth E.M. workshop, Golden, Colorado.

HILL, D.A. et WAIT, J.R., 1976, "Calculated transmission loss for a leaky feeder communication system in a circular tunnel", Radio Science, vol. 11, 4, pp. 315-322.

LAAKMAN, K.D. et STEIER, W.H., 1976, "Waveguides : characteristic modes of hollow rectangular dielectric waveguides", Applied optics, vol. 15, 5, pp. 1334-1341.

MAHMOUD, S.F., et WAIT, J.R., 1974, "Geometrical approach for electromagnetic wave propagation in rectangular mine tunnels", Radio Science, vol. 9, 12, pp. 1147-1158.

MARCATILLI, E.A.J. et SCHMELTZER, R.A., 1964, "Hollow metallic and dielectric waveguide for long distance optical transmission and lasers", Bell Syst. Tech. J., pp. 1783-1809.

STRATTON, J.A., 1961, "Théorie de l'Electromagnétisme", Dunod Ed.

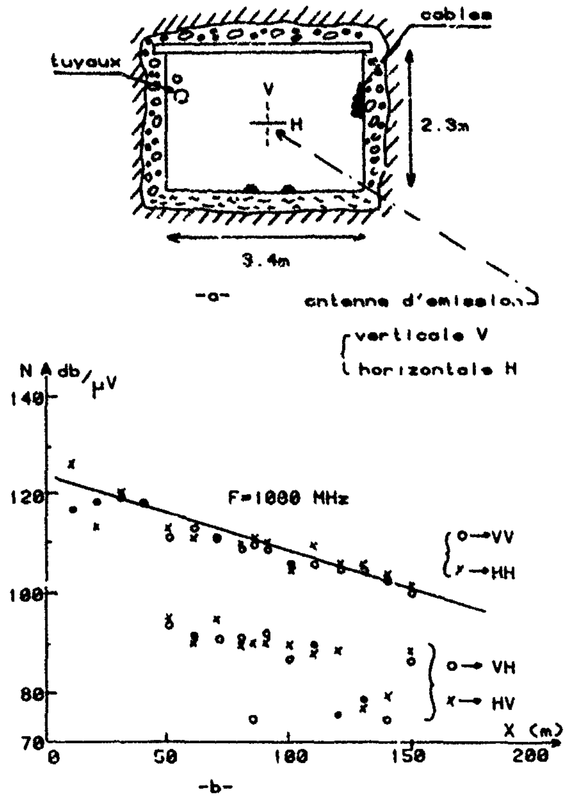


Figure 1 - a) Coupe de la galerie bétonnée - b) Amplitude du signal reçu en fonction de la distance émetteur-récepteur pour différentes configurations d'antennes.

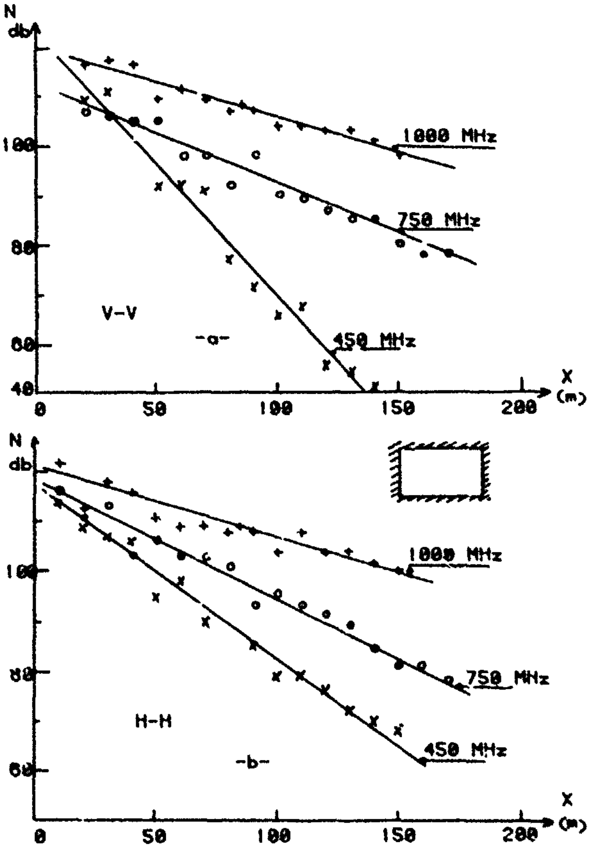


Figure 2 - Galerie rectangulaire dont les murs sont bétonnés : amplitude du champ pour différentes fréquences.

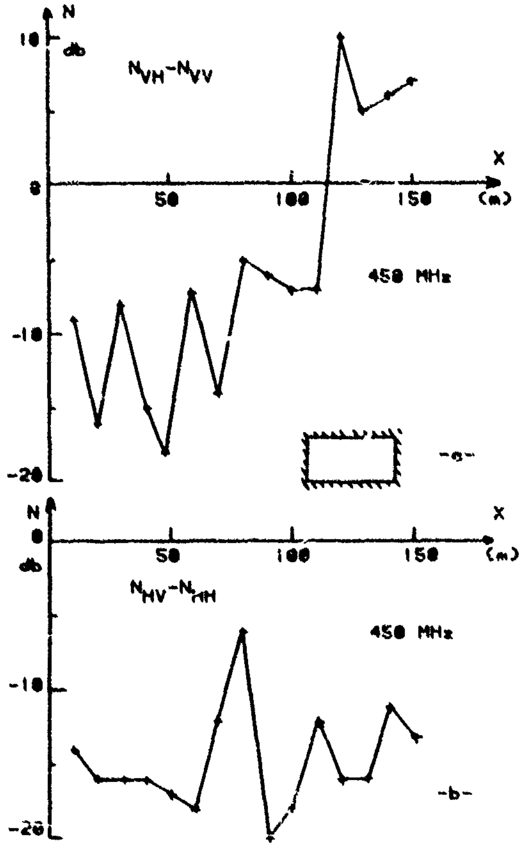
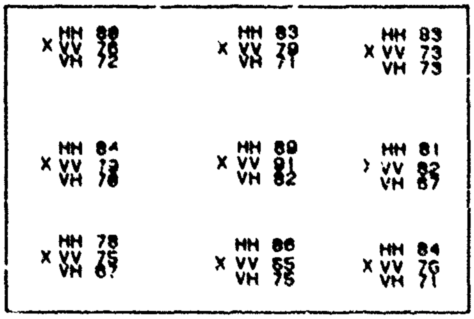


Figure 3 - Atténuation supplémentaire due à une configuration d'antennes croisées.

Carte de champ



X HH 80 VV 76 VH 72	X HH 83 VV 79 VH 71	X HH 83 VV 73 VH 73
X HH 84 VV 73 VH 78	X HH 89 VV 91 VH 82	X HH 81 VV 82 VH 87
X HH 78 VV 75 VH 67	X HH 86 VV 85 VH 75	X HH 84 VV 76 VH 71

Niveaux en " db/μV "

Figure 4 - Amplitude du champ dans le plan de section droite du tunnel pour différentes positions d'antennes.

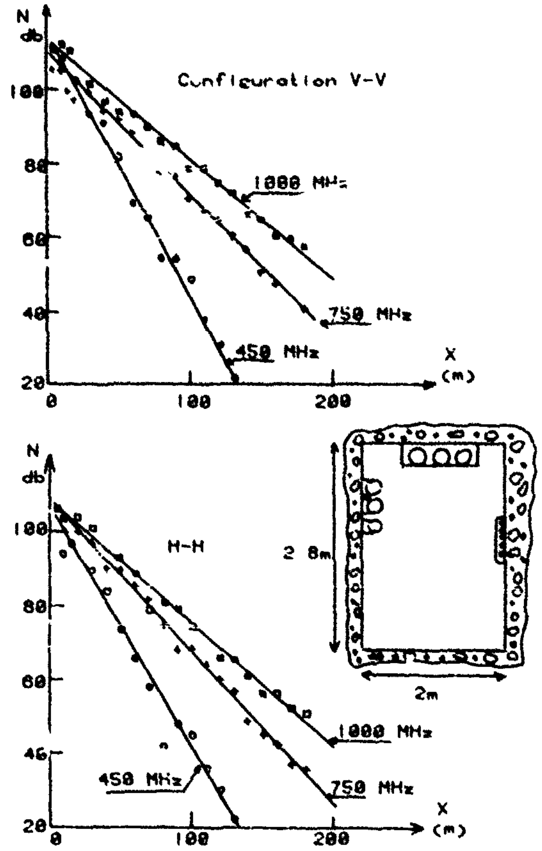


Figure 5 - Coupe d'un couloir souterrain et amplitude des champs en fonction de la distance émetteur-récepteur.

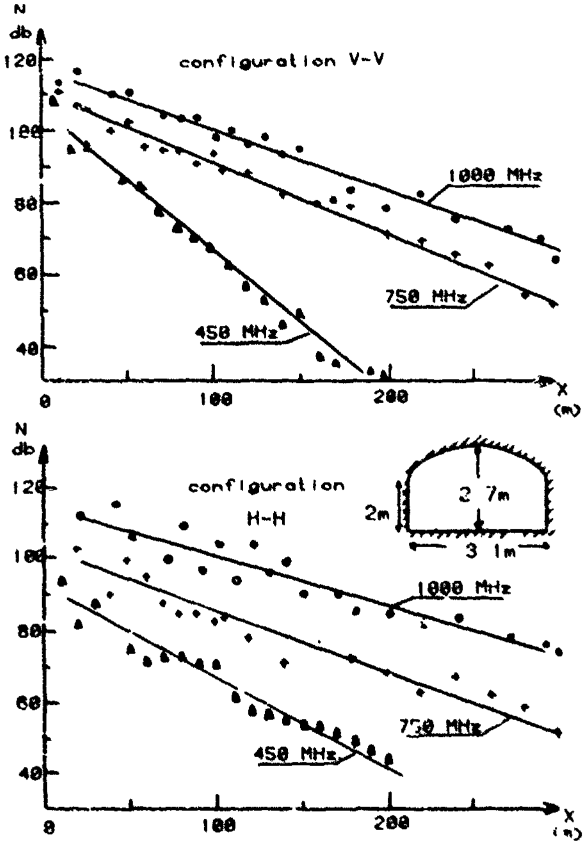


Figure 6 -

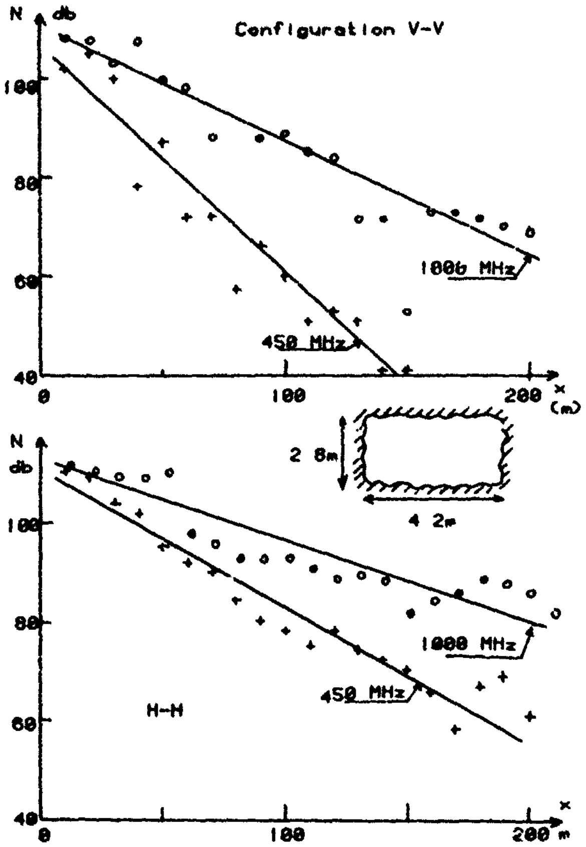


Figure 7 - Propagation dans une galerie de forme approximativement rectangulaire creusée dans une zone carbonifère.

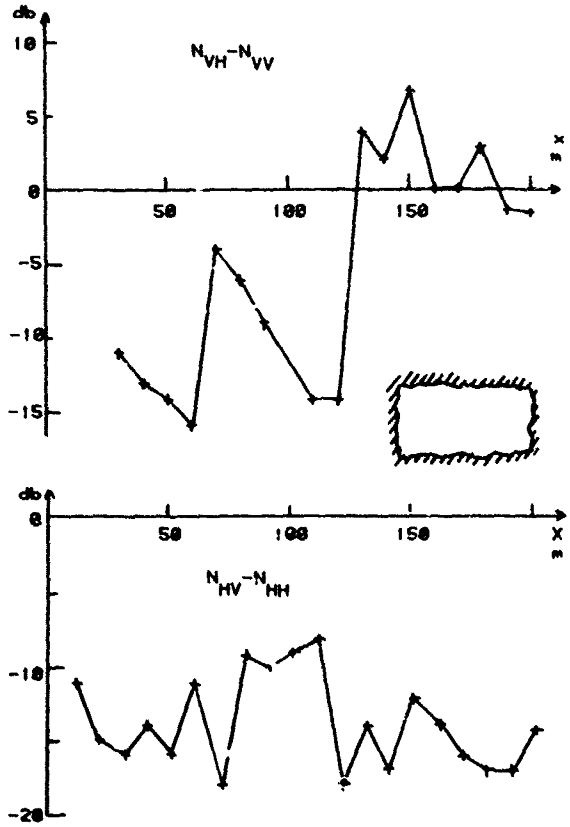


Figure 8 - Atténuation supplémentaire due à une configuration d'antennes croisées.

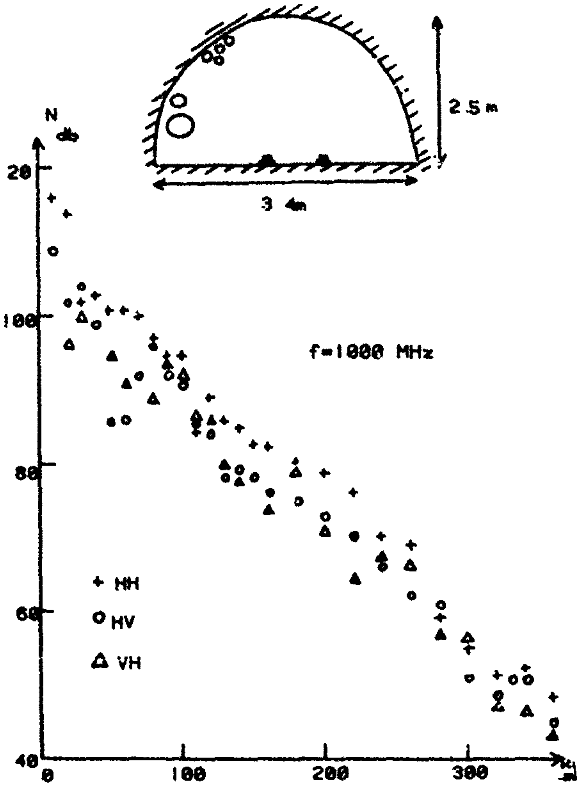


Figure 9 - Propagation à 1000 MHz dans une galerie de mine en présence de cadres métalliques de soutènement.

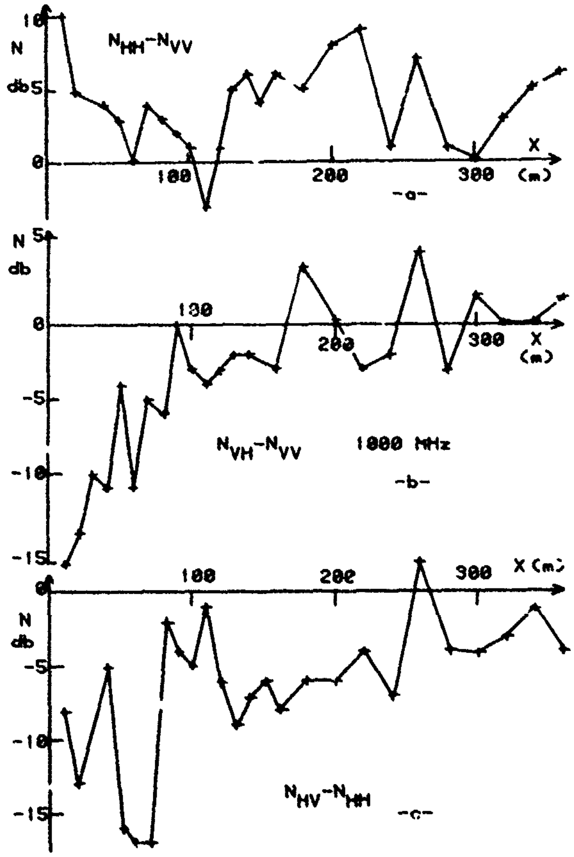


Figure 10 - Variation des niveaux reçus suivant les orientations des antennes d'émission et de réception. F = 1000 MHz.

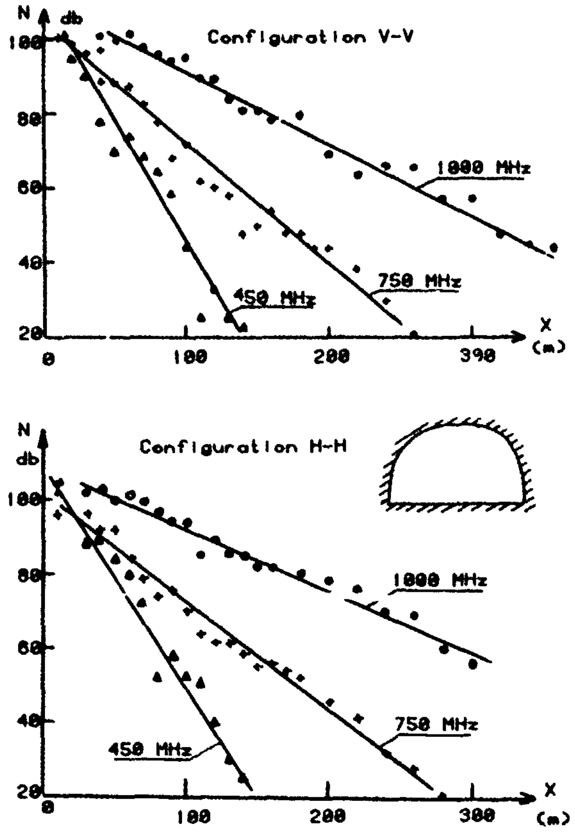
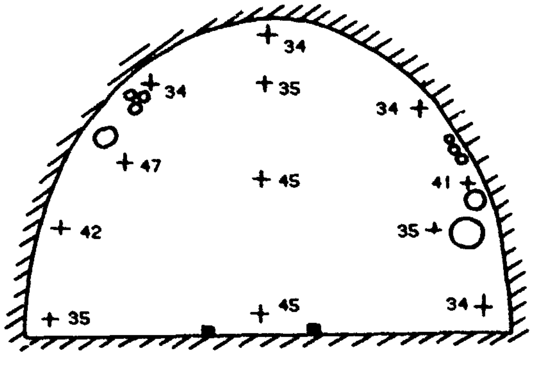
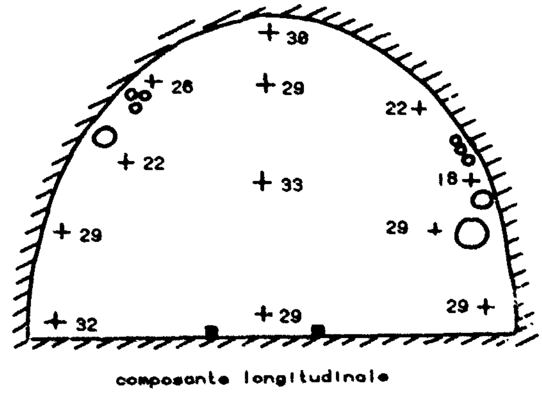


Figure 11 - Amplitude du signal pour différentes fréquences d'émission.


34
34
35
34
47
45
41
42
35
35
45
34

composante transversale


30
26
29
22
22
33
18
29
29
32
29
29

composante longitudinale

Figure 12 - Cartes de champ pour les composantes transversale et longitudinale.

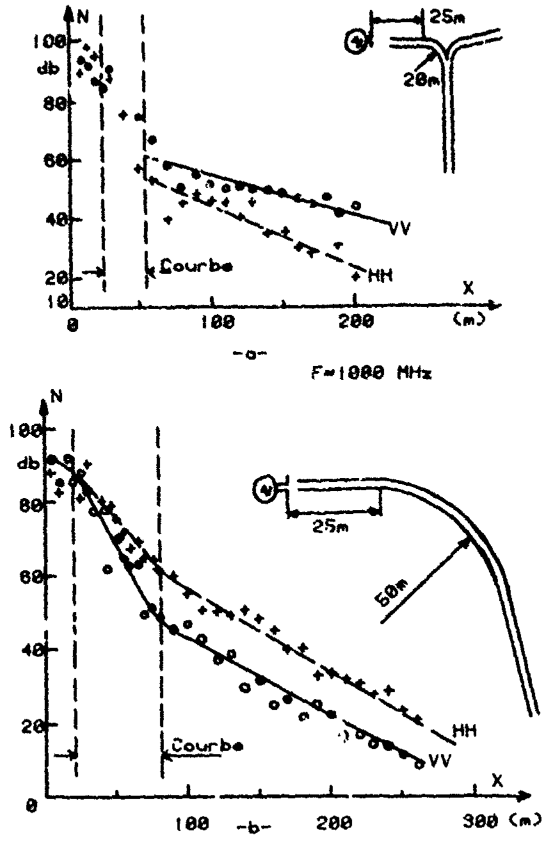


Figure 13 - Influence d'une courbe sur la propagation de l'onde.

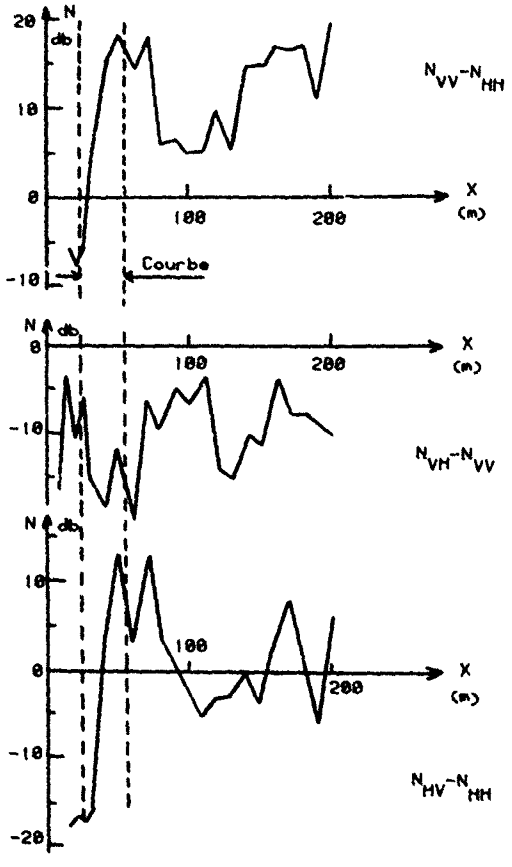


Figure 14 - Polarisation de l'onde après son passage dans une partie courbe du tunnel.

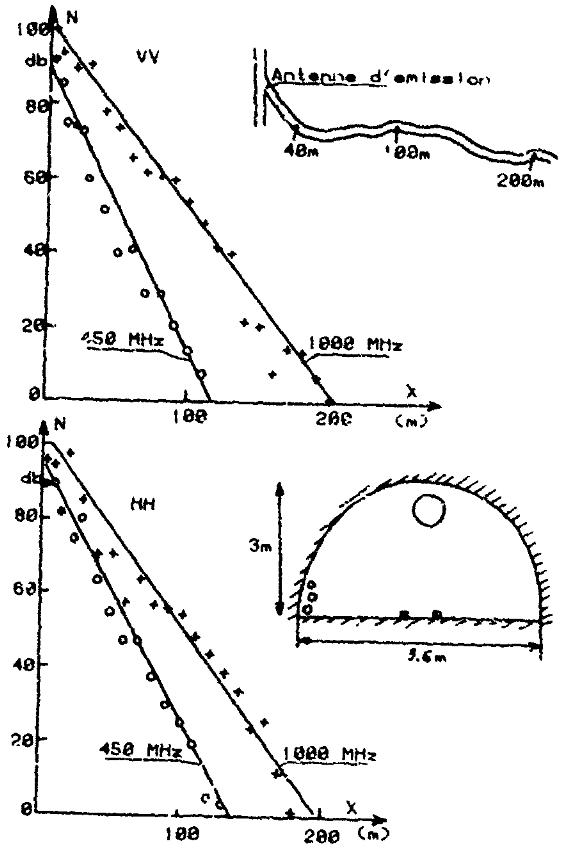


Figure 15 - Propagation en tunnel non rectiligne.

DISCUSSION

E.D.R. Shearman, UK

There has so far been no mention of the natural and man-made noise environment, yet it is clearly an essential part of the assessment of a complete system. Has the author any information on the levels of man-made and natural noise over the relevant frequency band?

Author's Reply

In the high frequency range, there is no natural noise, since the electromagnetic waves are highly attenuated during their path from the earth-surface to a mine tunnel – the man-made noise is due essentially to the multiple harmonics of the electric current (50 Hz) and to the motors. In this last case you can measure the noise for frequencies smaller than 50 MHz, but in the HF or UHF range it becomes negligible, and the noise amplitude is smaller than the noise of the field-strength apparatus.

EXCITATION OF THE H.F. SURFACE WAVE BY VERTICAL AND HORIZONTAL APERTURES

James R. Wait and David A. Hill
Institute for Telecommunication Sciences/NTIA
U.S. Dept. of Commerce, Boulder, Colo. 80303, U.S.A.

ABSTRACT

Excitation of the Zenneck wave over a flat homogeneous earth model was examined earlier [Hill and Wait, 1978]. The infinite vertical aperture with a Zenneck wave variation was found to excite only the Zenneck wave with no radiation field. The finite vertical aperture was found to excite a field which is similar to the Zenneck wave near the aperture but resembles the usual ground wave at large distances. Numerical results for the various aperture heights were given for frequencies of 1 and 10 MHz. Both homogeneous ground and sea paths were considered. Here we consider the generalization of the theory to allow fully for earth curvature. Also both vertical and horizontal apertures are considered. These formulas are the basis for more field strength calculations for various aperture distributions including the Zenneck wave taper. The conclusions reached earlier [Hill and Wait, 1978] are not modified.

1. INTRODUCTION

Ever since the pioneering work by Zenneck [1907] and Sommerfeld [1909, 1926], the propagation of radio waves along the earth's surface has captured the fancy of theoreticians. The subject reached a state of some maturity in the 1930-1940 period when Van der Pol and Bremmer [1937, 1938, 1939] exposed a general theory for the diffraction of a dipole field by a finitely conducting spherical earth. Their approach also reached a pinnacle of complexity because they wished to retain mathematical generality. But they did clearly identify the nature of a number of approximations that led to simplified working formulas (i.e., the residue series) that have been used by many for numerical studies. This subject is very clearly and exhaustively described by Bremmer [1949] who also made important individual contributions.

Benefiting by hindsight, a series of papers by Fock [1945, 1965] and his Soviet colleagues appeared about this same time that essentially rederived the Van der Pol-Bremmer theory by introducing various physical approximations at the outset rather than at the end of the analysis. In a parallel development based on wartime research in the early 1940s, Booker and Walkinshaw [1946] showed that the residue series representations were really nothing more than sums of normal modes. When the problem was formulated in this fashion for a perfectly conducting earth, the modal spectrum was discrete and no need arose to introduce a continuous spectrum of modes that was an inherent feature of the earlier studies. Sommerfeld was apparently very conversant with this approach as can be evidenced by his most eloquent discussion of a related problem in his famous Munich lectures on theoretical physics [e.g., see Sommerfeld, 1949].

Here we present an exposition of the normal mode approach to ground wave propagation; this formulation is well adapted to the solution of the surface wave excitation problem. The principal end result is the residue-series representation for the vertical electric field over a spherical earth when excited by either a vertical or a horizontal antenna. The derivation incorporates the essential features of prior analyses without becoming embroiled in the mathematical niceties of branch cuts and intricate and tricky deformations of integration contours in the complex wave number planes. Previous papers that deal generally with the topic are as follows: Spies and Wait [1966], Wait [1967], Zucker [1969], Bahar [1970], Wait [1970], Wait [1971], Cho and King [1972], King, et al, [1974], Wait [1974], King and Wait [1976]. Many of these papers also contain additional references.

We will deal exclusively with an airless earth since the physics of the diffraction process is not changed when a smooth gravitationally stratified atmosphere is allowed for [Fock, 1965, Wait, 1972]. Furthermore, the earth's lower boundary is taken to be a homogeneous dissipative medium characterized by a specified conductivity σ and permittivity ε. Any irregularities of the surface profile of the earth's lower boundary are also ignored. We neglect any ferromagnetic effects and assume that the whole space is characterized by the free-space permeability μ_o.

2. FORMULATION

Our specific propagation model is illustrated in Fig. 1. In terms of spherical coordinates (r,θ,ϕ) the earth's surface is at $r = a$ where a is the actual earth's radius. A conically shaped source region is defined by $r > a$ and $\theta < \theta_o$; later we let θ_o become vanishingly small.

An essential point of the present derivation is that complete azimuthal symmetry prevails in the sense that $\partial/\partial\phi = 0$. Also the source itself is assumed to excite only T.M. (transverse magnetic) waves in the sense that the magnetic field has only an azimuthal or ϕ component H_ϕ. Simple physics tells us that the electric field has only components E_r and E_θ. It is not surprising to learn that all three field components can actually be derived from a single scalar function U. For example, in the region $r > a$ and $\theta > \theta_o$, if we write

$$H_\phi = -i\varepsilon_o\omega\partial U/\partial\theta \tag{1}$$

then U is what is known as a Debye potential. Here ε_o is the permittivity of free space and ω is the angular frequency. The assumed time factor is $\exp(i\omega t)$. Then, from Maxwell,

$$E_r = (k^2 + \partial^2/\partial r^2)(rU) \tag{2}$$

$$E_\theta = (1/r)(\partial^2/\partial r\partial\theta)(rU)$$

As a further consequence, for $r > a$ and $\theta > \theta_o$,

$$(\nabla^2 + k^2)U = 0 \tag{4}$$

Solutions can be obtained by straight-forward separation of variables; however, we do not need to insist on singlevaluedness in the infinitely extended θ domain since we restrict attention to $\pi > \theta > \theta_o$. The radial functions are spherical Hankel functions $h_\nu(kr)$ of order ν that behave as $\exp(-ikr)$ as the argument $kr \to \infty$. The corresponding angular functions are Legendre functions $P_\nu(-\cos\theta)$ also of order ν; the argument $-\cos\theta$ or $\cos(\pi - \theta)$ is chosen such that the fields remain finite at the antipode $\theta = \pi$. To be explicit we write

$$U = \sum_\nu A_\nu \frac{h_\nu(kr)}{h_\nu(ka)} P_\nu(-\cos\theta) \tag{5}$$

where the summation encompasses all required values of ν (to be specified below). Here A_ν is a coefficient yet to be determined.

3. THE SURFACE IMPEDANCE CONDITION

We now invoke the surface impedance boundary condition that greatly simplifies the derivation. Simply stated, this requires that

$$E_\theta = -ZH_\phi(\text{at } r = a) \tag{6}$$

When Z is the surface impedance. A good value to choose for this parameter is

$$Z = (ik/\gamma)[1 + (k^2/\gamma^2)]^{\frac{1}{2}}\eta_o \tag{7}$$

when $\gamma = [i\mu_o\omega(\sigma + i\varepsilon\omega)]^{\frac{1}{2}}$ is the propagation constant for the earth and $\eta_o = (\mu_o/\varepsilon_o)^{\frac{1}{2}} = 120$ is the characteristic impedance of free space. Actually this form for Z would be exact for a vertically polarized wave at grazing incidence on a planar earth model. The so-called Leontovich boundary condition [e.g., see Fock 1965] is nothing more than the simpler version $Z = (ik/\gamma)\eta_o$ that has the additional constraint $|k^2/\gamma^2| << 1$. Actually the surface impedance formula can be made "exact" for the present problem by using

$$E_{\nu,\theta} = -Z_\nu H_{\nu,\phi} \tag{8}$$

where

$$Z_\nu = \frac{1}{\sigma + i\varepsilon\omega} \left[\frac{(\partial/\partial r)[rj_\nu(-i\gamma r)]}{rj_\nu(-i\gamma r)}\right]_{r = a} \tag{9}$$

is the "impedance' for spherical waves of order ν. Here j_ν is the spherical Bessel function that remains finite at $r = 0$, the center of the earth! The limiting case (7) is obtained from (9) when the Debye approximation is made and noting that the important values of ν are near k (i.e., grazing waves). The reader can delve into this question by reading the author's textbook [Wait, 1972] and several review papers on the subject.† Here we will be content with the form given by (7) which has the great merit that Z the effective surface impedance of the earth does not depend on the characteristic value ν. As a consequence our boundary condition becomes

$$\left\{\frac{d}{dx} [\chi h_\nu(\chi)] - i \frac{Z}{\eta_o} \chi h_\nu(\chi) \right\}_{\chi = ka} = 0 \tag{10}$$

Another revealing form of this equation is obtained by rewriting it as follows

$$Z + Z_{e,\nu} = 0 \tag{11}$$

where

$$Z_{e,\nu} = -\frac{1}{i\varepsilon_o\omega} \left[\frac{(\partial/\partial r)[rh_\nu(kr)]}{rh_\nu(kr)}\right]_{r = a} \tag{12}$$

is an "external" wave impedance. Equation (11) can be termed the transverse resonance relation.

4. AIRY FUNCTION APPROXIMATION

In spite of the compactness of (10) the determination of roots is a formidable task. In the early work on the subject such as by Vvedensky [1935, 1936, 1937] and Millington [1939] in the early 30s, the so-called tangent approximation was made which in effect replaced the spherical wave functions $h_\nu(kr)$ in (10) by their second order or Debye representation. Unfortunately this approximation breaks down just where it is needed most, i.e., for the lowest order roots where the attenuation is the least. It was primarily Van der Pol and Bremmer who surmounted this difficulty by utilizing a third order representation for the spherical wave functions in terms of Hankel functions of order 1/3. This process is described in a very comprehensive fashion by Bremmer [1949] in his book. A similar and, in many respects, a parallel development was carried out by Fock [1951] who used a very compact Airy function representation that was equivalent to the more cumbersome Hankel functions of order 1/3. We will adopt the latter form here; it amounts to making the substitution $\nu + 1/2 = (ka/2)^{1/3}t + ka$ and to writing

$$\chi h_\nu(\chi) = i(ka/2)^{1/6}w(t - y) \tag{13}$$

where

$$y = (2/ka)^{1/3}(\chi - ka),\ \chi = kr$$

and where the Airy function $w(t)$ satisfies the Stokes' differential equation

---

† e.g., see list of references, notably King and Wait [1976].

$$(d^2/dt^2)w(t) - tw(t) = 0 \tag{14}$$

In terms of contemporary Airy functions

$$w(t) = \pi^{1/2}[Bi(t) - iAi(t)]$$

Using (13) and the corresponding form for the derivative, we easily deduce that (10) or (11) is approximated by

$$[dw(t)/dt] - qw(t) = 0 \tag{15}$$

where $q = -i(ka/2)^{1/3} 2/\eta_c$.

Roots of this equation are denoted t_s where the subscript s assumes the values 1, 2, 3... . The corresponding values of the roots ν_s are then obtained from

$$\nu_s + (1/2) \simeq ka + (ka/2)^{1/3}\, t_s$$

which is a highly accurate approximation for the lower order roots where $|\nu_s - ka| \ll (ka/2)^{2/3}$.

5. THE EXCITATION PROBLEM

A further key step in the analysis is to enforce the source condition which we can state as follows

$$I(r) = 2\pi r\theta_o H_\phi(r,\theta_o) \tag{16}$$

for $\theta_o \ll 1$ and for $a < r < \infty$. Here $I(r)$ can be interpreted as the total axial current. Now, following Sommerfeld [1949], we note that for $\theta \ll 1$

$$P_\nu(-\cos\theta) \simeq \pi^{-1}\sin(\nu\pi)\ln\theta^2 \tag{17}$$

and

$$\partial P_\nu(-\cos\theta)/\partial\theta \simeq \pi^{-1}\sin(\nu\pi)(2/\theta) \tag{18}$$

Then using (1) and (5), the condition (15) becomes

$$I(r) = -4i\varepsilon_o\omega\sum_\nu A_\nu \frac{rh_\nu(kr)}{h_\nu(ka)}\sin\nu\pi \tag{19}$$

where the summation is to include all roots ν or ν_s. An important aspect of the present development is that the radial wave functions for these discrete modes (i.e., $\nu = \nu_s$) are orthogonal in the sense that

$$\int_a^\infty h_\nu(kr)h_{\nu'}(kr)dr = 0 \tag{20}$$

if $\nu \neq \nu'$. This relationship holds exactly under the assumption that the surface impedance boundary condition given by (6) is valid for all discrete modes. Further discussion of this point appears elsewhere [Wait, 1968] where it is also shown that the normalizing integral can be approximated by

$$\int_{ka}^\infty \left[\frac{h_\nu(kr)}{h_\nu(ka)}\right]^2 d(kr) \simeq \left(\frac{ka}{2}\right)^{1/3}(t_s - q^2) \tag{21}$$

for the roots $\nu = \nu_s$. It is now a simple matter to deduce that

$$A_\nu \simeq \frac{i\eta_o}{4\sin\nu\pi}\left(\frac{2}{ka}\right)^{1/3}\frac{1}{t_s - q^2}\int_a^\infty I(r)\frac{h_\nu(kr)}{h_\nu(ka)}\frac{1}{r}\,dr \tag{22}$$

If we deal with a vertical (i.e., radially oriented) electric dipole source at $r = r_o$, this amounts to saying that

$$I(r) = Ids\,\delta(r - r_o) \tag{23}$$

where $\delta(r - r_o)$ is the unit impulse function of $r = r_o$. Then (22) reduces to the approximate form

$$A_\nu \simeq \frac{i\eta_o}{4\sin\nu\pi}\left(\frac{2}{ka}\right)^{1/3}\frac{1}{t_s - q^2}\frac{Ids}{r_o}\frac{w(t_s - y_o)}{w(t_s)} \tag{24}$$

where $y_o = k(2/ka)^{1/3}(r_o - a)$ and where the Airy function representations for the spherical Hankel functions is valid when $|kr_o - \nu| \ll (ka/2)^{2/3}$.

6. EXPRESSIONS FOR THE FIELD

The field quantity of most physical interest is the radial or vertical electric field E_r. Clearly from (2) this is given by

$$E_r = \frac{1}{r}\sum_{\nu=\nu_s} A_\nu\nu(\nu + 1)\frac{h_\nu(kr)}{h_\nu(ka)}P_\nu(-\cos\theta) \tag{25}$$

Again here the height-gain function can be represented in terms of Airy functions by

$$\frac{h_\nu(kr)}{h_\nu(ka)} \simeq \frac{w(t_s - y)}{w(t_s)} \tag{26}$$

where $y = (2/ka)^{1/3}k(r - a)$ provided

$$|kr - \nu| << (ka/2)^{2/3}$$

for the applicable range of the heights $r - a$. Also in most cases we approximate the multiplicative factor

$$\nu(\nu + 1) \simeq (\nu + 1/2)^2 \simeq [ka + (ka/2)^{1/3}\ t_s]^2 \simeq (ka)^2 \tag{27}$$

for the most important terms in (25). Also in this same vein

$$P_\nu(-\cos\theta) \simeq \left(\frac{2}{\pi\nu\sin\theta}\right)^{1/2} \cos[(\nu + 1/2)(\pi - \theta) - \frac{\pi}{4}] \tag{28}$$

provided $|\nu| >> 1$ and θ is not near π. In fact, the further simplification is possible in most cases:

$$P_\nu(-\cos\theta) \simeq \left(\frac{2}{\pi ka\theta}\right)^{1/2} \frac{1}{2} \exp[+i(\nu + 1/2)(\pi - \theta) - i\frac{\pi}{4}] \tag{29}$$

since $[-\mathrm{Im}.\nu(2\pi - \theta)] >> 1$. Here we can identify $a\theta = d$ as the great circle distance between source dipole and observer.

Having made all the above approximations it is possible to write (25) in the following form

$$E_r \simeq E_o\left(\frac{\pi x}{i}\right)^{1/2} \sum_s \frac{e^{-ixt_s}}{t_s - q^2} \frac{w(t_s - y_o)}{w(t_s)} \frac{w(t_s - y)}{w(t_s)} \tag{30}$$

where $x = (ka/2)^{1/3}\ \theta = (ka/2)^{1/3}\ d/a$ and where

$$E_o = -i\mu_o\omega Ids/(2\pi d)$$

can be identified as the corresponding vertical field of the electric dipole located on the surface of a flat perfectly conducting plane. The ratio E_r/E_o is sometimes called the attenuation function and denoted W.

7. DISCUSSION OF HEIGHT GAIN FUNCTIONS

We now backtrack a bit and consider that we have a ground based vertical antenna of physical height h_o with a specified current distribution $I(h)$. Then it easily follows that

$$E_r = -\left[\frac{i\mu_o\omega}{2\pi d}\int_0^{h_o} I(h)dh\right] W \tag{31}$$

where the corresponding attenuation function is written

$$W = \left(\frac{\pi x}{i}\right)^{1/2} \sum_s \frac{e^{-ixt_s}}{t_s - q^2}\, G_s(y)\ H_s \tag{32}$$

where

$$G_s(y) = w(t_s - y)/w(t_s) \tag{33}$$

is the "height-gain" function for the observer and H_s is the "source excitation" function defined by

$$H_s = \int_0^{h_o} I(h)G_s(y)dh \Big/ \int_0^{h_o} I(h)dh \tag{34}$$

where

$$y = (2/ka)^{1/3}\ kh$$

It is now useful to note that the function $G(y) = w(t - y)/w(t)$ in general satisfies

$$d^2G/dy^2 = (t - y)G \tag{35}$$

Also, of course, $G(0) = 1$ and because of the boundary condition (15)

$$[dG_s/dy]_{y = o} = -q \tag{36}$$

Then it is a simple matter to show that

$$G_s(y) = 1 - qy + \frac{t_s y^2}{2} - \frac{1 + t_s q}{6}\ y^3 + \ldots \tag{37}$$

or, more generally that

$$G_s(y) = \sum_{n=0,1,2\ldots} A_{s,n} y^n \tag{38}$$

where the coefficients $A_{s,n}$ satisfy the recurrence relation

$$A_{s,n+2}(n+2)(n+1) = A_{s,n}t_s - A_{s,n-1} \tag{39}$$

when, by definition

$$A_o = 1 \text{ and } A_{-1} = A_{-2} = A_{-3} = 0$$

of course we can also write

$$G_s(y) = \sum_{n=0} a_{s,n}\, h^n \tag{40}$$

where

$$a_{s,n} = (y/h)^n A_{s,n} = [(2/ka)^{1/3}k]^n A_{s,n} \tag{41}$$

The "source excitation" function is now written

$$H_s = \sum_{n=0,1,2} a_{s,n} \int_o^{h_o} I(h)h^n dh \Big/ \int^{h_o} I(h)dh \tag{42}$$

This immediately illustrates that the higher order moments of the current distribution may contribute significantly to the total field.

We now represent the source current distribution as a sum of a finite number of exponentials. That is, we let

$$I(h) = \sum_{m=1,2,\ldots} p_m \exp(-\alpha_m h) \tag{43}$$

for $h_o > h > 0$ and specify that $I(h) = 0$ for $h > h_o$. This is an obvious form for the source if we are attempting to optimize the launching of some type of surface wave. Thus using (43) and (43), it follows readily that

$$H_s = \sum_m \sum_n a_n p_m \int_o^{h_o} e^{-\alpha_m h}\, h^n dh \Big/ \left[\sum_m \int_o^{h_o} e^{-\alpha_m h}\, dh\right] \tag{44}$$

or equivalently

$$H_s = \sum_m \sum_n a_n p_m \Lambda_{m,n} \Big/ \sum_m \Lambda_{m,o} \tag{45}$$

where

$$\Lambda_{m,n} = \int_o^{h_o} e^{-\alpha_m h}\, h^n dh \tag{46}$$

$$= \frac{n!}{\alpha_m^{n+1}} - e^{-\alpha_m h_o}\left[\frac{h_o}{\alpha_m} + \frac{nh_o^{n-1}}{\alpha_m^2} + \frac{n(n-1)h_o^{n-2}}{\alpha_m^3} \cdots + \frac{n!h_o}{\alpha_m^n} + \frac{n!}{\alpha_m^{n+1}}\right] \tag{47}$$

and

$$\Lambda_{m,o} = \int_o^{h_o} e^{-\alpha_m h}\, dh = \frac{1 - e^{-\alpha_m h_o}}{\alpha_m} \tag{48}$$

8. EXTENSION TO HORIZONTAL ANTENNA EXCITATION

In formulating the problem we had assumed an azimuthally symmetric source that was specialized to a radially oriented vertical antenna carrying a filamental current. This is the most obvious choice of a transmitting antenna that is to launch a vertically polarized (or T.M.) ground wave over the spherical earth. But another strong contender is a horizontal antenna that carries a specified filamental current throughout it's length. The natural building block here is the horizontal or tangentially-oriented electric dipole. Thus we could begin with the formal exact solution of the horizontal electric dipole in the vicinity of a homogeneous sphere [Wait, 1956]. This formalism, however, is not needed for present purposes if we make use of the reciprocity theorem and the prior solution for the vertical electric dipole.

For the vertical electric dipole an individual mode or term in (5) can be written

$$U = a_\nu h_\nu(kr)P_\nu(-\cos\theta) \tag{49}$$

where the radial function, of argument $X = kr$, satisfies

$$(d^2/dX^2)(Xh_\nu) + [X^2 - \nu(\nu+1)]h_\nu = 0 \tag{50}$$

Also for the ranges of interest,

$$P_\nu(-\cos\theta) \simeq \text{const.} \times (\sin\theta)^{-1/2} \exp[-i(\nu + 1/2)\theta] \tag{51}$$

$$E_{\nu,r} = \nu(\nu + 1)r^{-1} U_\nu \tag{52}$$

$$H_{\nu,\phi} \approx -\varepsilon_o \omega(\nu + 1/2)U_\nu \tag{53}$$

$$E_{\nu,\theta} \approx -i(\nu + 1/2)\partial(rU_\nu)/\partial r \tag{54}$$

Now the "wave tilt" parameter W_ν for a given mode is defined and given by

$$W_\nu = \frac{E_{\nu,\theta}}{E_{\nu,r}} \approx -\frac{i(\nu + 1/2)}{\nu(\nu + 1)} \frac{1}{U} \frac{\partial}{\partial r} (rU) \tag{55}$$

Using (52) this can also be written

$$W_\nu \approx -\frac{i(\nu + 1/2)}{\nu(\nu + 1)rE_{\nu,r}} \frac{\partial}{\partial r} (r^2 E_{\nu,r}) \tag{56}$$

The corresponding "wave impedance" parameter is

$$Z_\nu = -\frac{E_{\nu,r}}{H_{\nu,\phi}} \approx \frac{\nu(\nu + 1)}{\varepsilon_o \omega(\nu + 1/2)r} \tag{57}$$

Now using the Airy function approximation given by (13) it follows without difficulty that

$$W_\nu \approx \frac{\Delta}{Sq} \frac{w'(t_s - y)}{w(t_s - y)} \tag{58}$$

where Δ and S are defined by $q = -i(ka/2)^{1/3}\Delta$ or $\Delta = Z/\eta_0$ and $(\nu + 1/2)^2 \approx \nu(\nu + 1) \approx (kaS)^2$. Here, as before, t_s for mode s is related to ν(or ν_s) by $(\nu + 1/2) \approx ka + (ka/2)^{1/3}t_s$. In most cases S can be replaced by 1. Thus we see that $W_\nu \to \Delta$ as $y \to 0$ since $w'(t_s) = qw(t_s)$.

An equivalent representation for the wave-tilt parameter for a given mode follows from (33), thus

$$W_\nu \approx -\frac{\Delta}{S_q G_s(y)} G_s'(y) \tag{59}$$

On using the series for the height function $G_s(y)$ given by (37), it follows that

$$W_\nu \approx \frac{\Delta}{S} \quad \frac{1 - t_s y/q + (1/2)(1 + t_s q)y^2/q + \ldots}{1 - qy + t_s y/2 + \ldots} \tag{60}$$

Now if we write $S = (1 - C^2)^{1/2} \approx 1 - (C^2/2)$ then for a given mode $t_s \approx -C^2(ka/2)^{2/3}$. Keeping just first order terms in y we see that

$$W_\nu \approx \frac{1}{S} \frac{\Delta + ik(r - a)C}{1 + ik(r - a)} \tag{61}$$

This result can be shown to be quite consistent with the corresponding wave tilt for a vertically polarized wave with an angle of incidence of arc cosC or arc sinS (i.e., for a grazing angle of approximately C radians).

We now may apply the reciprocity theorem to determine the vertically polarized fields produced by a horizontal electric antenna. First of all we rewrite (30) specifically for the vertical field of a height h_2 and a range d vertical electric dipole VED of moment $I(h_1)dh_1$ at height h_1:

$$E_r = -\frac{i\mu_o \omega I(h_1)dh_1}{2\pi d} \left(\frac{\pi x}{i}\right)^{1/2} \sum_s \frac{e^{-ixt_s}}{t_s - q^2} G_s(h_1)G_s(h_2)e^{-ikd} \tag{62}$$

where the height-gain functions are defined by

$$G_s(h_i) = w(t_s - y_i)/w(t_s); \; i = 1,2 \tag{63}$$

and where $y_i = (2/ka)^{1/3}kh_i$. As before $x = (ka/2)^{1/3}$ d/a and t_s are the roots of $w'(t_s) - qw(t_s) = 0$ where $q = -i(ka/2)^{1/3}\Delta$ where Δ is the normalized surface impedance. Using the basic definition for the "wave tilt parameter" given by (55) it is evident that the corresponding expression for the total horizontal field E_θ for the VED is

$$E_\theta = -\frac{i\mu_o \omega I(h_1)dh_1}{2\pi d} \left(\frac{\pi x}{i}\right)^{1/2} \sum_s \frac{e^{-ixt_s}}{t_s - q^2} W_s G_s(h_1)G_s(h_2)e^{-ikd} \tag{64}$$

where

$$W_s = \Delta w'(t_s - y)/[qw(t_s - y)] \tag{65}$$

It is useful here to note that

$$W_s G_s(h_2) = \Delta g_s(h_2) \tag{66}$$

where

$$g_s(h_2) = \frac{w'(t_s - y_2)}{qw(t_s)} = \frac{w'(t_s - y_2)}{w'(t_s)} \tag{67}$$

We now can immediately invoke the reciprocity theorem to write down the expression for the vertical electric field at height h_2 of a horizontal electric dipole HED of moment $Id\ell_1$ at height h_1 for the same great circle range d. The result [after interchanging the indices 1 and 2 in (64)] is seen to be

$$E_r = -\frac{i\mu_0\omega I(h_1)d\ell_1}{2\pi d}\left(\frac{\pi x}{i}\right)^{1/2}\Delta\sum_s \frac{e^{-ixt_s}}{t_s - q^2}G_s(h_2)g_s(h_1)\cos\phi\ e^{-ikd} \tag{68}$$

where

$$g_s(h_1) = w'(t_s - y_1)/w'(t_s) \tag{69}$$

is the height-gain function for the HED. Here, of course, ϕ is the azimuthal angle at the receiver for the source HED oriented in the $\phi = 0$ direction. In addition to the E_r field and the E_θ and H_ϕ derived therefrom there will be an H_r field and H_θ and E_ϕ derived therefrom. Here we will only be concerned with the TM (transverse magnetic) or vertically polarized fields that are fully characterized by (62).

We now consider a horizontal antenna at height h_1 of total length L with a current distribution $I(\ell)$. Then, clearly, the moment $I(h_1)d\ell_1$ in (1) is to be replaced by

$$\int_{-L/2}^{L/2} I(\ell)e^{ik_s\ell\cos\phi}\,d\ell \tag{70}$$

and moved inside the summation. Here k_s is the wave number ν/a for the mode in question; that is

$$k_s a = ka + (ka/2)^{1/3}\,t_s$$

In an analogy to the case of the vertical antenna we can define an antenna gain function in the following fashion

$$F_s = \int_{-L/2}^{L/2} I(\ell)e^{ik_s\ell\cos\phi}\,d\ell \Bigg/ \int_{-L/2}^{L/2} I(\ell)d\ell \tag{71}$$

The working expression for the vertical electric field is thus

$$E_r = \left[-\frac{i\mu_0\omega}{2\pi d}\int_{-L/2}^{L/2} I(\ell)d\ell\right]\Delta\,\hat{W}\cos\phi \tag{72}$$

where

$$\hat{W} = \left(\frac{\pi x}{i}\right)^{1/2}\sum_s \frac{e^{-ixt_s}}{t_s - q^2}G_s(y)F_s g_s(h_1) \tag{73}$$

where

$$y(= y_2) = (2/ka)^{1/3}\,k(r - a)$$

An immediate example would be to assume a single travelling wave of current on the structure i.e., $I(\ell) = I_0\exp(-\Gamma\ell)$ for $-L/2 < \ell < +L/2$. Then

$$\int_{-L/2}^{+L/2} I(\ell)d\ell = \frac{2I_0}{\Gamma}\sinh\frac{\Gamma L}{2} \tag{74}$$

and

$$F_s = \frac{\sinh[(\Gamma - ik_s\cos\phi)(L/2)]}{(\Gamma - ik_s\cos\phi)(L/2)}\cdot\frac{(\Gamma L/2)}{\sinh(\Gamma L/2)}$$

9. SOME NUMERICAL RESULTS AND FINAL REMARKS

The preceding formulation has been used to obtain extensive numerical results for various kinds of aperture distributions. We give a few examples here for a frequency of 10 MHz with propagation over land with a conductivity $\sigma_e = 10^{-2}$ mhos/m and a relative permittivity $\varepsilon_e/\varepsilon_0 = 6$. In each case the effective height of the receiving (vertical whip) antenna is zero and the great circle range is denoted d. Normal atmospheric refraction is accounted for by using an effective earth radius equal to four thirds times the actual earth's radius.

In Fig. 2 the attenuation function W as defined by (32), is plotted as a function of the range d for the Zenneck wave distribution [Hill and Wait, 1978] for the current on the ground-based vertical antenna of height h_0. In the present case this means that $I(h) = I_0\exp(-\alpha h)$ where $\alpha = 0.02693 - i\,0.03946\ \mathrm{m}^{-1}$. The values of h_0 shown range from 0 to 200m as indicated. The case $h_0 = 0$, of course, corresponds to the ground-based vertical electric dipole. These results show that the Zenneck wave distribution is not very good for launching ground waves over a spherical earth. The reduction of the field strength with increasing h_0 is due to phase cancellation of the contributions in the various current elements. This effect would not take place for the planar model (e.g., as in Hill and Wait, 1978).

In Fig. 3 results are shown for the corresponding attenuation function W_d when we have simply a vertical electric dipole source at height h. Here the results illustrate the expected great advantage of raising up the centroid of the current distribution so that full advantage can be taken of the height-gain function.

Finally in Fig. 4 we show the attenuation function $\hat{W}$ as defined by (73) for excitation by a horizontal antenna of length L. The effective height h_1 of this structure is taken to be zero and we also set $\phi = 0$ corresponding to the end-fire direction. The lengths L vary from 0 to 195m. Here the current distribution is a travelling wave with a propagation constant $\Gamma = ik(1 - \Delta^2)^{1/2} = 0.0512 + i\,0.2076\text{m}^{-1}$ where k is the free space wave number. Again this corresponds to a Zenneck wave, a plane earth with the same electrical constants. The case $\Gamma = ik = 0.20958\text{m}^{-1}$ of course corresponds to a travelling wave on the antenna structure with an assumed free space propagation constant. In the latter case the field strength at the receiver is actually higher than for the Zenneck wave distribution. Again this is not too surprising because the propagation constant of the dominant creeping waves over the spherical earth are nearer ik than to the corresponding wave value.

In general we may conclude that the Zenneck wave illumination is a poor choice for enhancing the ground wave field strength over a spherical earth.

REFERENCES

BAHAR, E., 1970, "Propagation of radio waves over a non-uniform layered medium," Radio Science, Vol. 5, No. 7, pp. 1069-1076.

BOOKER, H.G., and WALKINSHAW, W., 1946, "The mode theory of tropospheric refraction and its relation to wave-guides and diffraction," Meteorological Factors in Radio-Wave Propagation, Physical Society, London.

BREMMER, H., 1949, Terrestrial Radio Waves, Elsevier, New York.

CHO, S.H., and KING, R.J., 1972, "EM ground-wave propagation over the curved lunar surface," IEEE Trans. Geosci. Electronics, Vol. GE-10, pp. 96-105.

FOCK, V.A., 1945, "Diffraction of radio waves around the earth's surface," Jour. Theo. and Exp. Physics, Vol. 15, pp. 480-490.

FOCK, V.A., 1965, Propagation and Diffraction of Electromagnetic Waves, Pergamon Press, New York and Oxford.

HILL, D.A., and WAIT, J.R., 1978, "Excitation of the Zenneck surface wave by a vertical aperture," Radio Science, Vol. 13, No. 6, pp. 967-977.

KING, R.J., CHO, S.H., JAGGARD, D.L., BRUCKNER, G.E., and HUSTIG, C.H., 1974, "Experimental data for ground wave propagation over cylindrical surfaces," IEEE Trans. on Antennas and Propagation, Vol. AP-22, No. 4, 551-556.

KING, R.J., and WAIT, J.R., 1976, "Electromagnetic groundwave propagation - theory and experiment, Symposia Mathematica, Vol. 18, pp. 107-208, Academic Press.

MILLINGTON, G., 1939, "The diffraction of wireless waves around the earth," Phil. Mag., Series 7, Vol. 27, No. 184, pp. 517-542.

SOMMERFELD, A.N., 1909, 1926, "The propagation of waves in wireless telegraphy," Ann. Phy., Series 4, Vol. 28, pp. 665-737 and Vol. 31, 1135-1153.

SOMMERFELD, A.N., 1949, Partial Differential Equations, Academic Press, New York.

SPIES, K.P., and WAIT, J.R., 1966, "On the calculation of the ground wave attenuation factor at low frequencies," IEEE Trans. on Antennas and Propagation, Vol. AP-14, No. 4, pp. 515-517.

VAN DER POL, B., and BREMMER, H., 1937, 1938, 1939, "The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere," Phil. Mag., Series 7, Vol. 24, pp. 141-176, pp. 825-864; Vol. 25, pp. 817-834; Vol. 26, pp. 261-275.

VVEDENSKY, B., 1935, 1936, 1937, "The diffractive propagation of radio waves," Tech. Physics (USSR), Vol 2, pp. 624-639; Vol. 3, pp. 913-325; Vol. 4, pp. 579-591.

WAIT, J.R., 1956, "Low frequency radiation from a horizontal antenna over a spherical earth," Can. Jour. Phys., Vol. 34, pp. 586-595.

WAIT, J.R., 1965, "Theory of diffraction by a curved inhomogeneous body," Jour. Math. Phys., Vol. 8, No. 4, pp. 920-925.

WAIT, J.R., 1968, "Diffraction and scattering of the electromagnetic ground wave by terrain features," Radio Science, Vol. 3, No. 10, pp. 995-1003.

WAIT, J.R., 1970, "Propagation of electromagnetic waves over a smooth multi-section curved earth - an exact theory," Jour. Math. Phys., Vol. 11, No. 9, pp. 2850-2860.

WAIT, J.R., 1971, "Theory of ground wave propagation," Electromagnetic Probing in Geophysics, Chap. 5, pp. 163-207, Golem Press, Boulder, Colo.

WAIT, J.R., 1972, Electromagnetic Waves in Stratified Media, 2nd Edition, Pergamon Press, New York and Oxford.

WAIT, J.R., 1974, "Recent analytical investigations of electromagnetic ground wave propagation over inhomogeneous earth models," Proc. IEEE, Vol. 62, No. 8, pp. 1061-1072.

ZENNECK, J., 1907, Uber die Fortpflanzung ebener electromagnetische Wellen einer ebener Leiterflache und ihre Beziehung zur drahtlosen Telegraphie, Ann. Phys., Series 4, Vol 23, pp. 846-866.

ZUCKER, F.J., 1969, "Surface Wave Antennas," Antenna Theory, (ed. by R.E. Collin, and F.J. Zucker), Chap. 21, Part II, McGraw Hill New York.

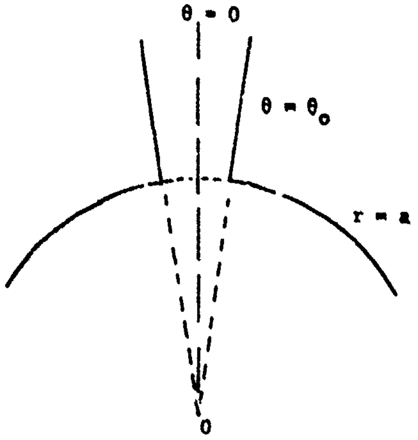


Fig.1 Conical source region and a portion of earth's surface

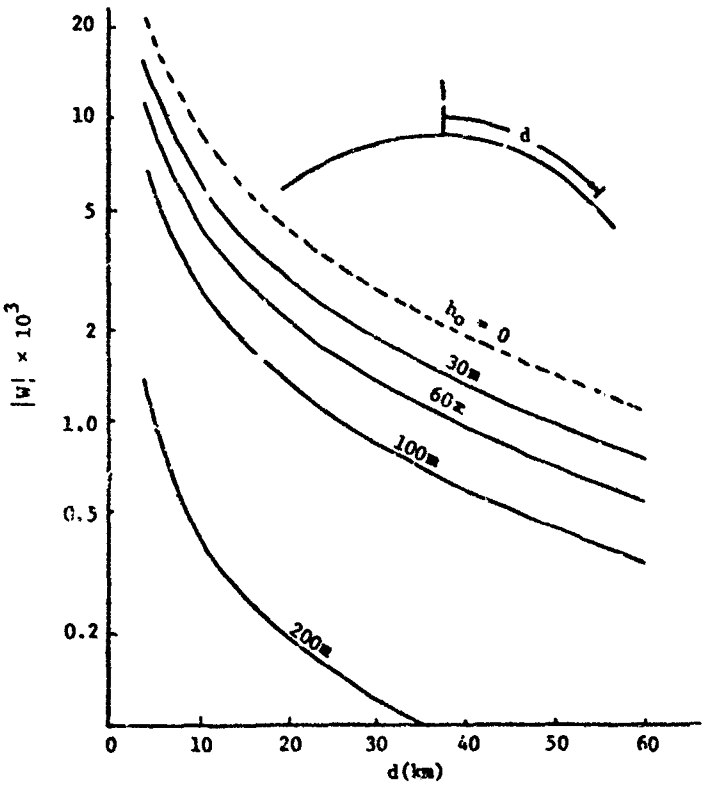


Figure 2

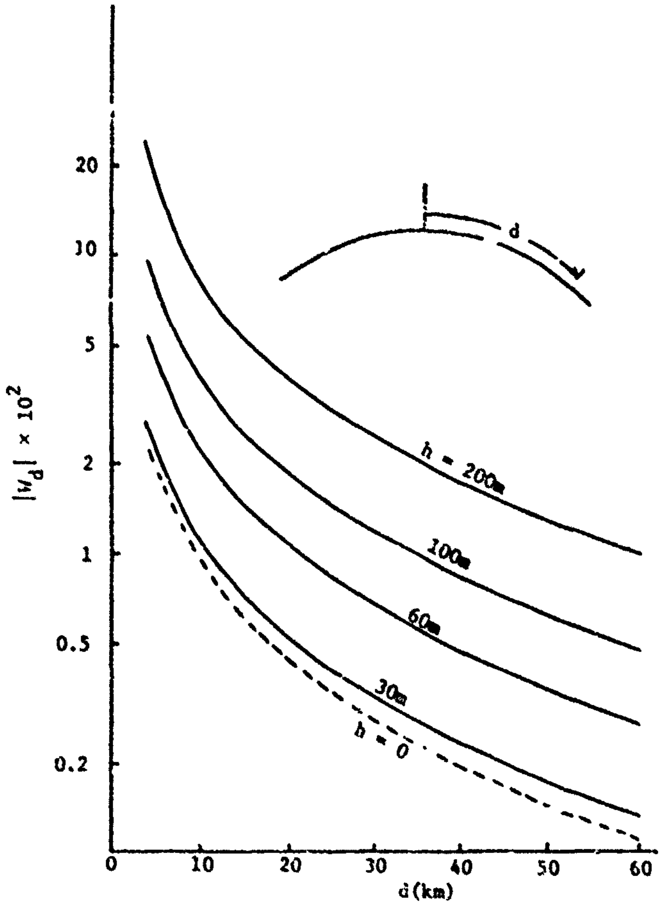
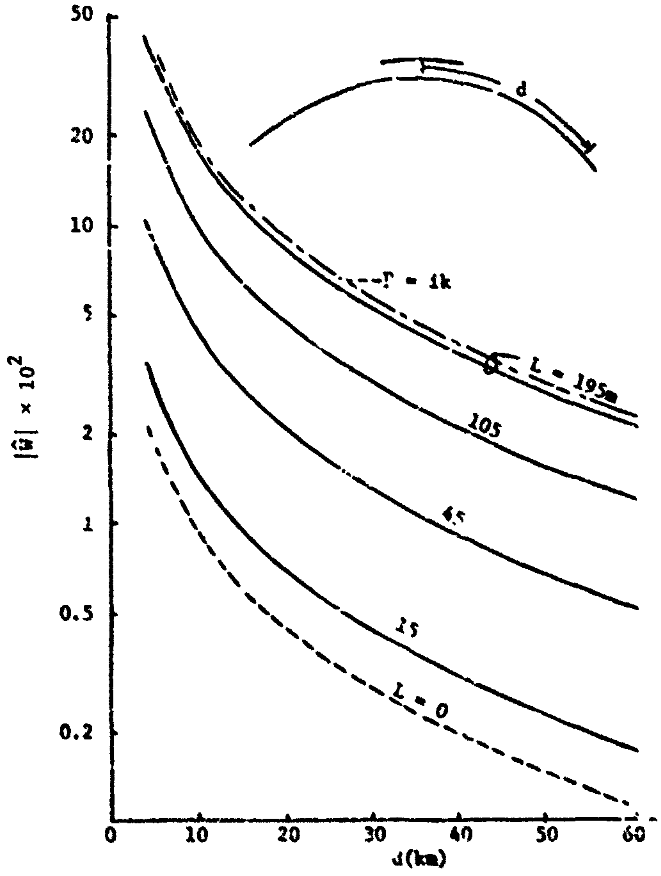

20
10
5
2
1
0.5
0.2
$|W_d| \times 10^2$
d
h = 200m
100m
60m
30m
h = 0
0
10
20
30
40
50
60
d(km)

Figure 3


50
20
10
5
2
1
0.5
0.2
$|\hat{W}| \times 10^2$
d
Γ = ik
L = 195m
105
45
15
L = 0
0
10
20
30
40
50
60
d(km)

DISCUSSION

H. Poeverlein, Ge

Do you still think that an aperture of infinite height would allow an improvement corresponding to the Zenneck wave? Is it the limitation to a realistic height extension that prevents a better result?

Author's Reply

Even an infinite aperture could not excite the planar earth form of the Zenneck wave *on a spherical earth* But, of course, such an aperture would excite the usual creeping wave modes. Thus, I do not think the height limitation of the aperture is the crucial factor. Rather, one should minimize the phase cancellation of the fields radiated from the individual elements in the aperture. The Zenneck wave distribution is particularly bad in this respect.

EFFECTIVE USE OF NATURAL MODES IN VHF AND UHF TUNNEL PROPAGATION

L. DERYCK

Université de Liège
Institut de Physique
B-4000 SART TILMAN
BELGIUM

SUMMARY

An experimental study of the electromagnetic wave propagation in various tunnels, at frequencies between 1 MHz and 1,000 MHz is presented.

The results obtained provide a better understanding of natural propagation mechanisms in underground galleries containing no transmission line. The analysis of these results reveals the existence of a cutoff frequency and a strong correlation between the propagation characteristics and the transverse dimensions of the tunnel.

Further experiments are made in tunnels where a transmission line is used as a waveguide; they show under what conditions the natural modes can be useful in this case. These modes can advantageously replace the so-called monofilar mode which is generally used for coupling between the transmission line and the mobile transceivers. Experimental checking shows that when using a well adapted transmission line, exploitation of natural propagation can greatly reduce the cost of the line installation without reducing its performances.

1. INTRODUCTION

The most effective way, known up to the present, of solving the problem of radio propagation in tunnels, consists in using transmission lines as supports for propagation of transverse electromagnetic modes (Fontaine et al., 1973, Delogne, 1974; Cree and Giles, 1975; Delogne and Safak, 1975; Deryck, 1975; Martin, 1975; Wait, 1975). The two modes mostly used are the so-called coaxial mode and the monofilar mode. The coaxial mode ensures the energy propagation whereas the monofilar mode makes possible the coupling between the line and the transceivers.

However, even when containing no transmission line, the tunnel itself acts as a waveguide and makes the propagation of natural modes possible. These modes appear when frequency is high enough, higher than a limit value, which is called the cutoff frequency of the tunnel. These modes are generally greatly attenuated so that they could only be useful to ensure communications in short tunnels. However, for distances shorter than 200 m, these modes can be more suitable than the transmission-line-supported modes (Damoso and De Padova, 1976).

The theory of natural mode propagation in tunnels has been extensively studied (Comstock, 1971; Goddard, 1973; Mahmoud and Wait, 1974; Emslie et al., 1975; Anderson et al., 1975). When frequency is not too high, the tunnel can be assimilated to a metallic waveguide with imperfectly conducting walls. The various theories proposed give similar results, except in the cases of degenerated modes which present some discrepancy (Anderson et al, 1975). At higher frequencies, the walls of the tunnel can be assimilated to a lossy dielectric. Theoretical results show that in this case, the attenuation falls off like the frequency square (Mahmoud and Wait, 1974; Emslie et al, 1975). When experimental checkings are made, they are generally limited to 2 or 3 frequencies.

One of the objects of this paper is to present experimental measurements of natural propagation in actual tunnels of various shapes and sizes and for frequencies lower and higher than the cutoff frequency. The results obtained obviously show that there are three different propagation processes. Moreover, by showing the predominant influence of the transverse size of the gallery, they will lead to a better estimation of the value of the attenuation of natural propagation in a tunnel.

Another object of this paper, is to show under what conditions the natural modes can advantageously replace the monofilar mode when a transmission line is used.

Experiments are presented which make conspicuous that using natural modes instead of the monofilar mode, with a view to ensuring the coupling between the line and the transceivers can often increase the line performances and reduce the cost of its installation.

2. NATURAL PROPAGATION THEORY

Assimilating the tunnel to a hollow conducting waveguide makes possible a very attractive explanation of the various propagation mechanisms which occur around its cutoff frequency when this tunnel contains no transmission line.

Such a guide makes possible propagation of transverse electric and transverse magnetic modes when the frequency is higher than a limiting value which is the cutoff frequency of a particular mode. The values of these cutoff frequencies depend on the given mode, and are also determined by the shape and the transverse dimension of the gallery. The cutoff wavelengths are given for a rectangular waveguide by (Marcuvitz, 1951).

$$\lambda_{mn} = \frac{2\sqrt{ab}}{\sqrt{m^2 \frac{b}{a} + n^2 \frac{a}{b}}} \quad (1)$$

where a is the width of the guide and b its height, and m, n are equal to 1, 2, 3, ..., for the TM_{mn} modes, and equal to 0, 1, 2, 3, ..., for the TE_{mn} modes. For a circular waveguide, they are given for the TM_{mn} modes by :

$$\lambda_{mn} = \frac{2\pi}{\lambda_{mn}} r \quad (2)$$

where r is the radius of the waveguide and λ_{mn} is the nth nonvanishing root of the mth order Bessel function. For the TE_{mn} modes, the cutoff wavelengths are given by :

$$\lambda_{mn} = \frac{2\pi}{\chi_{mn}'} r \quad (3)$$

where χ_{mn}' is the nth nonvanishing root of the first derivative of the mth order Bessel function.

Below the lowest cutoff frequency, propagation is not possible. There, the field decrease is only function of the transverse dimensions of the tunnel. The attenuation α is independent of the electrical properties of the wall. It increases towards a limiting value with decreasing frequency in accordance with the following equation (Terman, 1947) :

$$\alpha = \frac{54.6}{\lambda_c} \sqrt{1 - \left(\frac{\lambda_c}{\lambda}\right)^2} \quad \text{dB/unit length} \quad (4)$$

where λ_c is the longest cutoff wavelength of the waveguide.

Above its cutoff frequency, attenuation of each mode depends on the frequency, shape, transverse dimensions and electrical properties of the waveguide.

For rectangular waveguides, this attenuation calculated by a classical method, is given for TM_{mn} modes by (Marcuvitz, 1951) :

$$\alpha = 8.69 \frac{2R}{\zeta a} \left(\frac{m^2 + n^2 \frac{a^3}{b^3}}{m^2 + n^2 \frac{a^2}{b^2}} \right) \frac{1}{\sqrt{1 - \left(\frac{\lambda}{\lambda_{mn}}\right)^2}} \quad \text{dB/m} \quad (5)$$

where

ζ : intrinsic impedance of the propagation medium = $\sqrt{\mu/\varepsilon}$

R : $10.88 \,.\, 10^{-3} \sqrt{(10^7/\sigma)\,(1/\lambda)}$ ohms

σ : conductivity of the guide walls in mho/m

μ : permeability of the propagation medium in henry/m

ε : permittivity of the propagation medium in farad/m.

For TE_{mn} modes, this attenuation α is given by :

$$\alpha = 8.69 \frac{R}{\zeta b} \left[\frac{\varepsilon_n m^2 \frac{b}{a} + \varepsilon_m n^2}{m^2 \frac{b}{a} + n^2 \frac{a}{b}} \sqrt{1 - \left(\frac{\lambda}{\lambda_{mn}}\right)^2} + \frac{\left(\varepsilon_n + \varepsilon_m \frac{b}{a}\right)\left(\frac{\lambda}{\lambda_{mn}}\right)^2}{\sqrt{1 - \left(\frac{\lambda}{\lambda_{mn}}\right)^2}} \right] \quad \text{dB/m} \quad (6)$$

where $\varepsilon_m = 1$ if $m = 0$ and $\varepsilon_m = 2$ if $m \neq 0$.

For circular waveguides, the attenuation α for the TM_{mn} modes is given by (Marcuvitz, 1951) :

$$\alpha = 8.69 \frac{R}{\zeta r} \frac{1}{\sqrt{1 - \left(\frac{\lambda}{\lambda_{mn}}\right)^2}} \quad \text{dB/m} \quad (7)$$

and for the TE_{mn} modes by :

$$\alpha = 8.69 \frac{R}{\zeta r} \left[\frac{m^2}{\chi_{mn}'^2 - m^2} + \left(\frac{\lambda}{\lambda_{mn}}\right)^2 \right] \cdot \frac{1}{\sqrt{1 - \left(\frac{\lambda}{\lambda_{mn}}\right)^2}} \quad \text{dB/m} \quad (8)$$

However, the walls of actual tunnels are often far from being well-conducting. As a result, there are two limits to this waveguide model. First, below the cutoff frequency of the tunnel, if frequency is low enough, the electromagnetic wave can propagate through the rock, in the same manner as if there were no tunnel. Attenuation α in the tunnel is then the same as through the rock. It is given by (Gabillard, 1970) :

$$\alpha = 8.69 \sqrt{(2\pi f \mu \sigma / 2)} \quad \text{dB/m} \tag{9}$$

where f is the frequency in hertz, μ the permeability in henry/m and σ the conductivity of the rock in mho/m.

When frequency is low enough, attenuation as given by (9), can be lower than attenuation resulting from (4).

Secondly, at frequencies high enough, the walls of the tunnel act as a dielectric medium rather than as a conducting medium. Then, the loss is essentially due to the fact that any part of the wave that impinges on a wall of the tunnel is partially refracted into the wall and partially reflected back into the tunnel. This problem has been studied theoretically for circular and rectangular tunnels (Glaser, 1969; Mahmoud and Wait, 1974; Emslie et al., 1975).

On condition that wavelength is short as compared to the waveguide transverse dimensions and that we have :

$$\frac{\sigma}{2\pi f \varepsilon} >> 1 \tag{10}$$

The attenuation of a horizontally polarized wave in a rectangular tunnel is given by (Emslie et al., 1975) :

$$\alpha_h \ (m, n) = 4.343 \ \lambda^2 \left(\frac{m^2 \varepsilon_r}{a^3 \ (\varepsilon_r - 1)^{1/2}} + \frac{n^2}{b^3 \ (\varepsilon_r - 1)^{1/2}} \right) \ \text{dB/m} \tag{11}$$

The attenuation of the vertically polarized wave is then :

$$\alpha_v \ (m, n) = 4.343 \ \lambda^2 \left(\frac{m^2}{a^3 \ (\varepsilon_r - 1)^{1/2}} + \frac{n^2 \varepsilon_r}{b^3 \ (\varepsilon_r - 1)^{1/2}} \right) \ \text{dB/m} \tag{12}$$

where ε_r is the relative permittivity of the wall. Let us notice that these attenuations are generally increased, as a result of wall roughness (Mahmoud and Wait, 1974; Emslie et al., 1975).

3. EXPERIMENTAL RESULTS IN ACTUAL TUNNELS

The most difficult problem to solve, with a view to correlating the experimental results with the theoretical investigations, are that TEM modes may be present simultaneously with waveguide modes, and also that there is generally no information about the electrical characteristics of the tunnel walls. TEM modes are due to the presence in most galleries of longitudinal conductors such as pipes, railway tracks, trolley wires, etc...

Moreover, the rock electrical characteristics which are found in various tables are useless for the present application. These characteristics vary generally with frequency (Fritsch, 1963). As an example, rock conductivity generally increases with frequency (Debye effect) and with the voltage applied (Wien effect). This dependence is very complex, both Wien effect and Debye effect depending themselves on temperature and pressure (Onsayer effect). This attenuation dependence on frequency is more complex than it appears in equations (5) to (9).

As a result, it is not expected to obtain a perfect correlation between experiments and theory. However, we shall see that this comparison is useful as it makes possible a qualitative explanation of the various propagation mechanisms observed.

The first experiments were conducted in a tunnel without any conductor except the rock and includes conductivity measurements of its wall in situ. This tunnel is dug in calcareous tufa. It is 1,600 m long, 5 to 6 m high, and 4 to 5 m wide. The rock overburden is about 50 m thick. The value of its conductivity was obtained by measuring the attenuation of the wave propagating through the rock, and found equal to 10^{-2} mho/m at 30 MHz. This tunnel is located in Lanaye, near Liège in Belgium. Its shape is approximately similar to a circular cylinder.

The attenuation of electromagnetic waves was measured in a straight section of this tunnel for various frequencies between 2 MHz and 1 GHz. The results obtained are represented by crosses on fig. 1. Curve 1 of this diagram represents the theoretical attenuation of a wave propagating through a medium having a conductivity of 10^{-2} mho/m ((9)). Curve 2 shows the attenuation of a hollow circular waveguide below its cutoff frequency ((4)). The cutoff frequency was 35 MHz corresponding to approximating the tunnel by a circular cylinder with a radius of 2.5 m with propagation in the TE_{11} mode. Taking account of the conductivity of the walls, we drew curves 3, 4 and 5, which give the theoretical attenuation of TE_{11}, TE_{01}, and TM_{11} modes for such a guide ((7)) and ((8)).

It can be seen in fig. 1 that, for each frequency, experimental results are in agreement with the theroretical curve which predicts the lowest attenuation.

Between 40 MHz and about 90 MHz, the TE_{11} mode is dominant, the attenuation of the other modes being greater by at least 100 dB/100 m.

As a result, the field decrease is uniform in the gallery. At higher frequencies, the field becomes more irregular as a result of interference phenomena between various modes. Moreover, at some frequencies, we observe two gradients in the field decrease. As an example, at 200 MHz, fig. 2 shows that near the transmitter attenuation is about 108 dB/100 m, but at a greater distance from the transmitter, an attenuation of only 28 dB/100 m is observed. These values agree with the theoretical attenuation of the TE_{11} and of the TE_{01} modes respectively.

Experiments were also performed in a rectangular road tunnel before it was open to traffic. It is 17 m wide, 4.9 m high and about 600 m long. Its walls are made of concrete. This tunnel is located in Brussels, Belgium.

In this tunnel, the attenuation of electromagnetic waves was measured at various frequencies for vertically and horizontally polarized antennas. The results are shown in fig. 3, where the crosses represent horizontal polarization and the circles vertical polarization. Theoretical attenuation has been calculated below cutoff frequency using (4), and above cutoff frequency for the TE_{01} and TE_{10} modes using (6) under the assumption that the conductivity of the walls was 0.1 mho/m, which is a probable value for concrete. It is obvious that a vertical antenna excites essentially the TE_{10} mode, and a horizontal antenna the TE_{01} mode.

Cutoff frequencies of these modes are, respectively, 8.8 MHz and 30.6 MHz. It can be seen in fig. 3 that below 40 MHz, polarization has practically no influence on propagation. In fact, the only mode which can propagate with low attenuation at these frequencies is the TE_{10} mode. When both modes can propagate with low attenuation, differences appear in attenuation according to polarization. These differences are no longer observed at the highest frequencies as a result of the high attenuation of the TE_{10} mode.

Comments on these results are in order.

1) Around the cutoff frequency, just below it, relation (4) is always verified with perfect precision. As a result, propagation is never possible below the cutoff frequency, except if wall conductivity is low enough, so that propagation through rock is possible.
2) Attenuation, as given by (11) or (12) leads to a value much lower than the experimental results. This cannot be attributed to wall roughness. Indeed, let us remember that walls are made of concrete and so are very smooth. However, let us note that the highest frequency considered (1,000 MHz)

   $$\frac{\sigma}{2\pi f\varepsilon_o}$$

   is equal to 1.8, a value which is surely not very much lower than ε_r.
3) Equations (5) to (8) show that if attenuation increases with the square root of resistivity, it is more strongly dependent on the transverse dimensions of the tunnel. Thus, the feasibility of a radio link in a tunnel is more dependent on the transverse size of the tunnel than on its wall conductivity. One can see in fig. 4 how attenuation increases when the transverse dimensions of the tunnel are reduced by a half, and when conductivity is reduced by a half. This figure shows clearly that the influence of the transverse dimensions on attenuation is dominant. This dependence is still higher when relations (11) and (12) are available. In this case, attenuation is inversely proportional to the cube of the transverse dimensions.

Two important conclusions can be drawn from these experiments. First, propagation is rarely possible below the cutoff frequency of the tunnel. Secondly, attenuation of natural modes will be low in wide-sized tunnels.

These conclusions can be verified by any car driver having a radio receiver on board. They can easily observe that VHF can be caught in wide road tunnels, being generally shorter than 200 m, while LF can never be caught in tunnels which contain no longitudinal wire. This is illustrated by fig. 5 and 6. Fig. 5 represents the field level of a radio broadcasting station in LF (620 kHz) in outside and inside the tunnels of the ring of Brussels. All the tunnels are similar, except in length to the road tunnel studied above. Their length varies between 100 m for the shortest tunnels and 300 m for the longest. Even in the shortest tunnels, field level is lower than noise in most part of the tunnel. The same measurement made at 100 MHz, is presented in fig. 6. The results obtained show that field level is higher than noise level everywhere in the short tunnels and in most parts of the others.

4. NATURAL MODES AND TRANSMISSION LINES

The principle of transmission lines, used as supports for guided propagation in tunnels consists in transmitting the electromagnetic energy by using the coaxial or bifilar mode, while the monofilar or natural modes ensure the coupling between the line and the transceivers in the gallery. Monofilar and natural modes being highly attenuated, exchanges of energy between these modes and the coaxial or bifilar mode are needed. These exchanges may be continuously realized when using a leaky cable (Martin, 1975), or at discrete places when using mode converters, radiating devices or antennas (Delogne, 1974; Deryck, 1975). In the latter case, the coupling loss depends also on the monofilar mode attenuation between two mode converters.

Monofilar mode has been studied extensively (Gabillard, 1970; Deryck, 1973, 1975). Its attenuation increases with frequency and is strongly dependent on the cable position, as it is shown on fig. 7. In practice, the transmission line must generally be hung in a very eccentric position which increases the monofilar mode attenuation and the coupling loss between this mode and the transceivers. As a result, when using monofilar mode, the cable must be hung carefully, and its distance from the wall must be as high as possible, and surely not lower than 5 cm.

The monofilar mode attenuation is an increasing function of frequency, while the natural mode attenuation can decrease with increasing frequency.

Comparing experimental results such as those presented in fig. 1 and 7, it is easy to determine, for a given position of the line in the tunnel, the frequency beyond which natural modes are less attenuated than the monofilar one. As an example, in the tunnel of Lanaye, when the distance between the line and the wall cannot be higher than 10 cm, natural propagation is more favourable than the monofilar one at frequencies higher than 80 MHz. Knowing the frequency limit where natural propagation is better than monofilar propagation is very important. Indeed, in some cases, it determines the optimal use of the transmission line.

Let us consider the line constituted by a non-leaky coaxial cable used together with radiating devices inserted at discrete places. If natural modes are not very attenuated, hanging such a line carefully becomes useless, except around the radiating devices. This line can be hung against the wall which greatly reduces the cost of its installation. Let us note that this is never possible with leaky cable without reducing its performances. Indeed, hanging a leaky cable close to a wall increases the attenuation of the coaxial or bifilar mode (Deryck, 1973; Delogne, 1976).

5. EXPERIMENTAL CONTROL

Various experiments can be carried out with a view to making conspicuous the advantages of using natural modes.

In the tunnel of Lanaye, we measured the field level along a non-leaky coaxial cable containing a short leaky section. A generator, coupled at the begining of the line excites the coaxial mode. The leaky section excites both monofilar and natural modes in the gallery (Delogne, 1976). The cable is hung against the wall. Frequency is 450 MHz. The field level beyond the leaky section is shown on fig. 8. We can observe that the attenuation value (15 dB/100 m) is similar to the attenuation of natural propagation as given by fig. 1.

Other experiments were made in a rectangular road tunnel. This tunnel is 5 m high and 9 m wide. Its walls are made of concrete. We measured the attenuation of natural propagation in this tunnel when it contains no conductor except its walls at 70 MHz, 100 MHz and 150 MHz. Then, a 300-metre non-leaky coaxial cable was hung at 10 cm from the roof and at 1 m from the side wall; a radiating device (Delogne, 1974) was placed in the middle of this cable, with a view to exciting both monofilar and natural modes in the tunnel, when the coaxial mode is excited by a generator coupled to this line. The field level was measured on each side of the radiating device at 70 MHz, 100 MHz and 150 MHz. As an example, the result obtained at 70 MHz is represented on fig. 9. The field attenuation at these frequencies was similar to that obtained in the case of natural propagation Analyzing the field level around the end of the line (noted EL on fig. 9) we observe that this end provokes no discontinuity in this level. This clearly indicates that propagation outside the cable is essentially due to natural modes.

The following experiments enable us to find the frequency beyond which natural modes are less attenuated than the monofilar mode. In a first experiment, the field level was measured in the axis of the gallery, at 70 m from the radiating device for frequencies between 15 MHz and 150 MHz, the whole line and the radiating device being both at 10 cm from the tunnel roof. In a second experiment, the distance between the line and the tunnel roof was increased up to 1 m near the receiving antenna (fig. 10 a), with a view to reducing the coupling loss between the monofilar mode and this antenna. The increase in the field level, compared with the previous measurement is plotted on fig. 11, curve a. The results obtained clearly show that the coupling loss is reduced (G > 0) only for frequencies lower than 60 MHz, showing that at these frequencies, propagation is due to the monofilar mode.

In a third experiment, the whole line was hung at 10 cm from the roof, but the distance between the radiating device and the roof was increased up to 1 m (fig. 10 b). This has no effect on the monofilar mode level, but it has some influence on the field radiated. The increase in the field level, compared with the first experiment is plotted on fig. 11, curve b. The results obtained show that the field level is increased for frequencies higher than 50 MHz, which implies that propagation outside the coaxial cable is then due to natural modes.

An important conclusion from these three experiments is that in this road tunnel, for frequencies higher than 50 MHz, hanging the cable against the wall or at 10 cm from it, has no influence on the propagation phenomena. Moreover, increasing the distance between the radiating device and the wall improves the performances of the transmission line.

A further experiment enables us to verify this conclusion. From the radiating device hanging from the roof, the cable was laid down in a gutter. The field level in the tunnel was measured at 150 MHz. The results are represented on fig. 12. These results are rather similar to those obtained when the cable is hung at 10 cm from the roof (fig. 10), which shows that in this case, there is no need to hang the cable carefully.

6. CONCLUSIONS

The experimental results presented above enable us to distinguish three different ranges of frequency, characterized by three different propagation mechanisms. Below cutoff frequency, waves propagate through the rock in the same manner as if there were no cavity. Attenuation increases with the square root of frequency and is a function of the conductivity of the underground. Close by cutoff, just below it, attenuation is only determined by frequency and by the shape and transverse dimensions of the tunnel.

At frequencies above cutoff, at a given frequency, attenuation depends on both conductivity and transverse dimensions of the tunnel, and is strongly correlated with the theoretical attenuation of the waveguide modes. Though, it is rather difficult to determine the conductivity of the tunnel walls with a good precision, the results presented, by showing the predominant influence of the transverse size of the gallery, can help to estimate the value of the attenuation of electromagnetic waves propagating in a tunnel.

When the electromagnetic waves are guided in the tunnel, with the help of a non-leaky transmission line, used together with radiating devices, the existence of natural modes makes the propagation phenomena independent of its position. As a result, when natural modes are less attenuated than the monofilar mode, the line can be hung against the wall, except the radiating device, without reducing its performances. Moreover, increasing the distance between the radiating device and the wall can improve the performances of this system.

REFERENCES

Anderson, J.B., Berntsen, S., and Dalsgaard, P., 1975, "Propagation in rectangular waveguides with arbitrary internal and external media", IEEE Trans. Microwaves Theory and Tech., vol. MTT-23, (7), pp. 555-560.

Comstock, W.D., 1971, "Electromagnetic wave propagation in an underground duct", Proc. IREE, Australia, vol. 32, pp. 159-162.

Cree, D.J., and Giles, L.J., 1975, "Practical performance of radiating cables", The Radio and Electronic Engineer, vol. 45, pp. 215-223.

Damoso, E.D., and De Padova, S., 1976, "Propagation and radiation of VHF - Radio signals in motorway tunnels", IEEE Trans. on Vehicular Technology, vol. VT-25, pp. 39-45.

Delogne, P., 1974, "Les systèmes INIEX de communication par radio", Ann. Mines de Belgique, vol. 10, pp. 951-962.

Delogne, P., and Safak, M., 1975, "Electromagnetic theory of the leaky coaxial cable", The Radio and Electronic Engineer, vol. 45, pp. 233-240.

Delogne, P., 1976, "Basic mechanisms of tunnel propagation", Radio Sci., vol. 11, pp. 295-303.

Deryck, L., 1973, "Etude de la propagation des ondes électromagnétiques guidées dans les galeries souterraines", Thèse de doctorat, Université de Liège, Belgium.

Deryck, L., 1975, "Control of mode conversions on bifilar line in tunnels", The Radio and Electronic Engineer, vol. 45, pp. 241-247.

Emslie, A.G., Lagace, R.L., and Strong, P.F., 1975, "Theory of the propagation of UHF radio waves in coal mine tunnels", IEEE Trans. Antennas Propagat., AP-23 (2), pp. 192-205.

Fontaine, J., Demoulin, B., Degauque, P., and Gabillard, R., 1973, "Feasibility of radio communications in mine galleries by means of a coaxial cable having a high coupling impedance", in Proceedings of Thru-the-Earth Electromagnetics Workshop, Colorado School of Mines, Golden, Colorado, National Technical Information Service, Springfield, VA 22151, acc. no. PB 231154, pp. 130-139, 15-17. August 1973.

Fritsch, V., 1963, "Propagation of radio frequency electromagnetic fields in geological conductors", J. Res. Nat. Bur. Std., vol. 67 D no. 2.

Gabillard, R., 1970, "Propagation des ondes électromagnétiques dans une galerie souterraine", Revue de l'Industrie Minérale, vol. 52, pp. 83-108.

Glaser, J.I., 1969, "Attenuation and guidance of modes in hollow dielectric waveguides", IEEE Trans. Microwaves Theory and Tech., vol. MTT-17, pp. 173-174.

Goddard, A.E., 1973, "Radio propagation measurements in coal mines at UHF and VHF", in Proceedings of Thru-the-Earth Electromagnetics Workshop, Colorado School of Mines Golden, Colorado, National Technical Information Service, Springfield, VA 22151 acc. no. PB 231154, pp. 54-61, 15-17. August 1973.

Mahmoud, S.F., and Wait, J.R., 1974, "Geometrical optical approach for electromagnetic wave propagation in rectangular mine tunnels", Radio Sci., vol. 9, pp. 1147-1158.

Marcuvitz, N., 1951, Waveguide Handbook. New York, Toronto, London : Mc Graw-Hill, pp. 55-72.

Martin, D.J.R., 1975, "A general study of the leaky feeder principle", The Radio and Electronic Engineer, vol. 45, pp. 205-214.

Terman, F.E., 1947, Radio Engineering. New York and London : Mc Graw-Hill, pp. 137-143.

Wait, J.R., 1975, "Theory of transmission of electromagnetic waves along multi-conductor lines in tunnels", The Radio and Electronic Engineer, vol. 45, pp. 229-232.

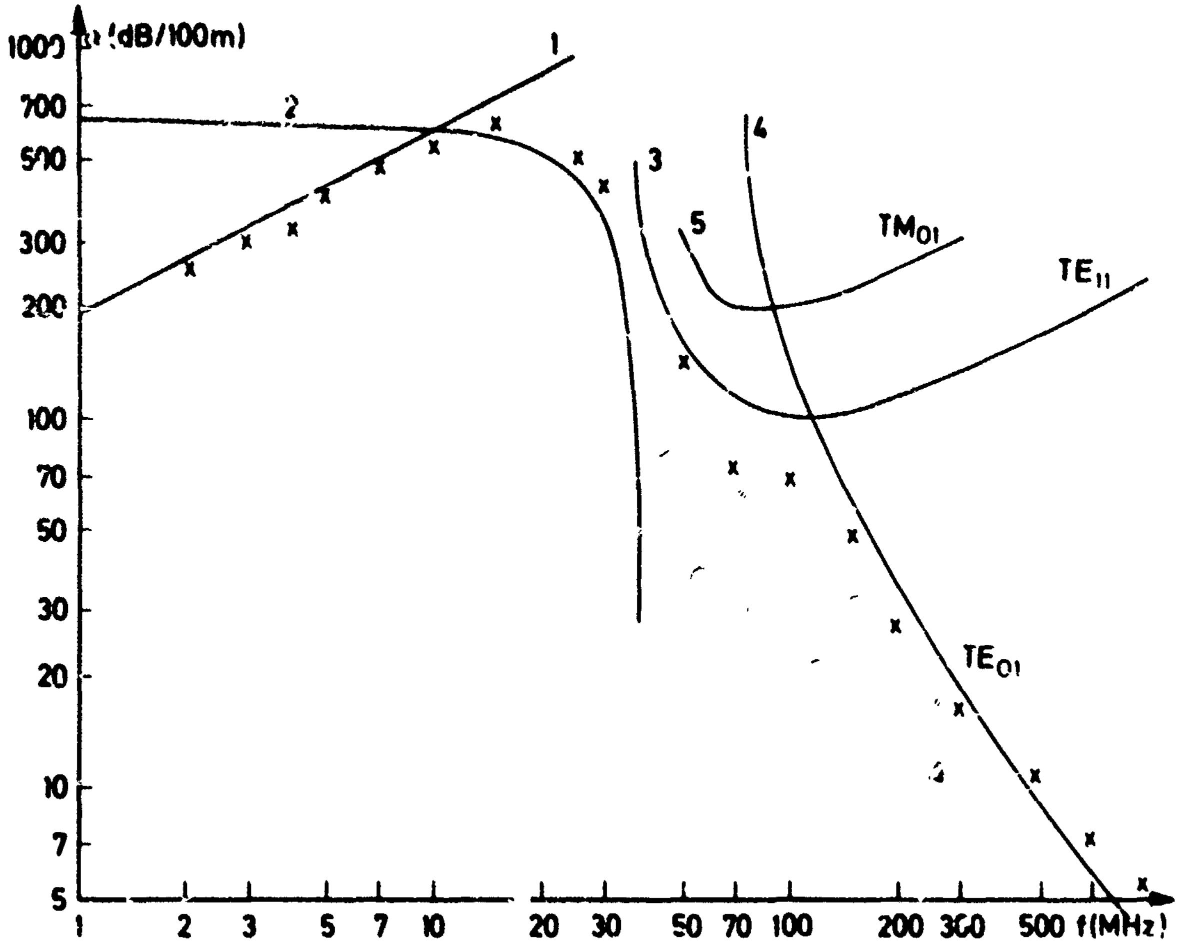


Fig 1 Experimental results and theoretical curves of the attenuation versus frequency in the tunnel of Lanaye

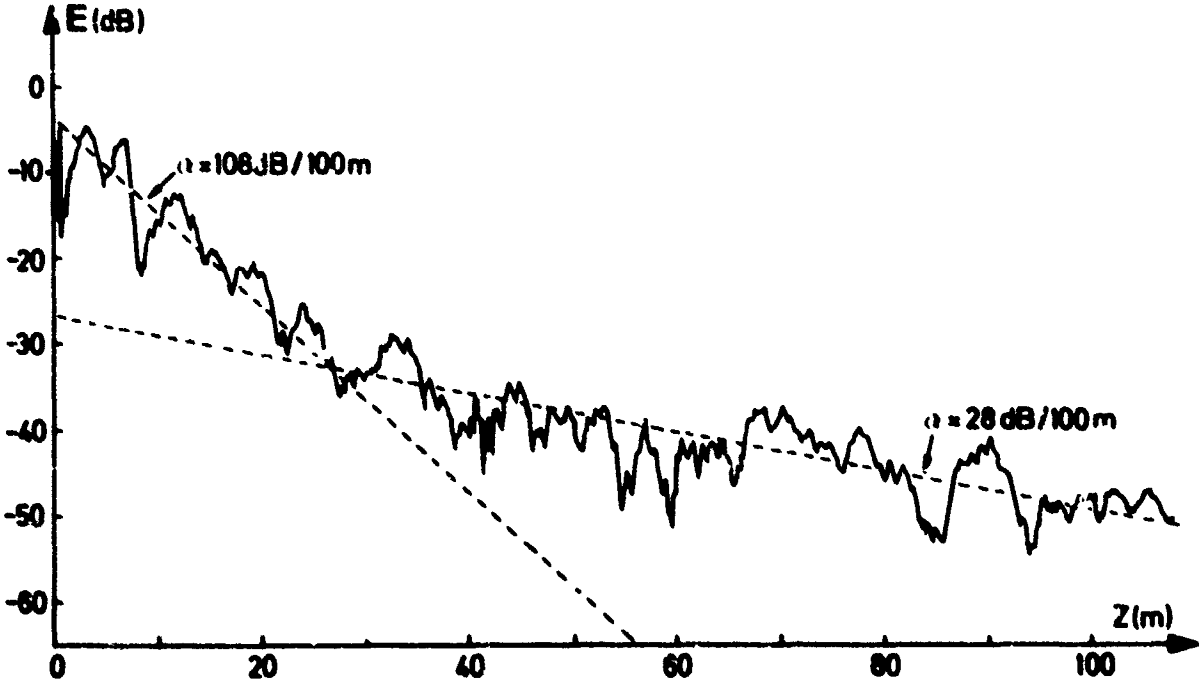


Fig.2 Field decrease in the tunnel of Lanaye when at least two modes can propagate. Frequency is 200 MHz

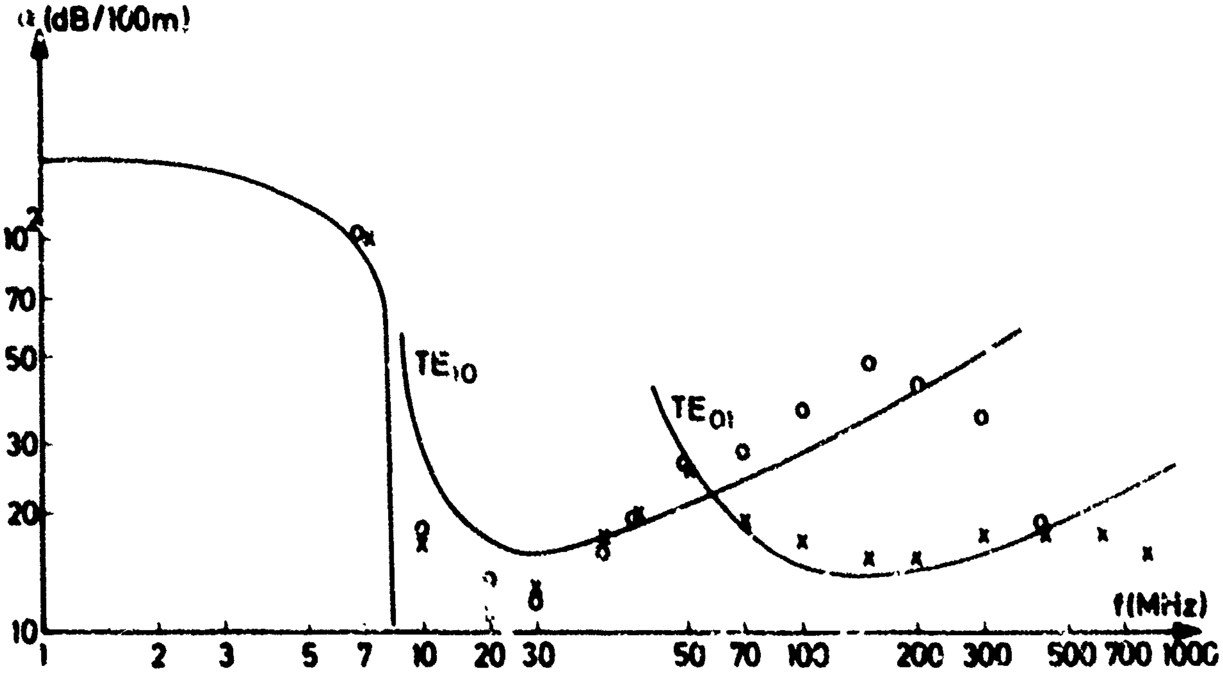


Fig 3 Attenuation versus frequency in a rectangular tunnel. Crosses represent vertical polarization, and circles horizontal polarization.

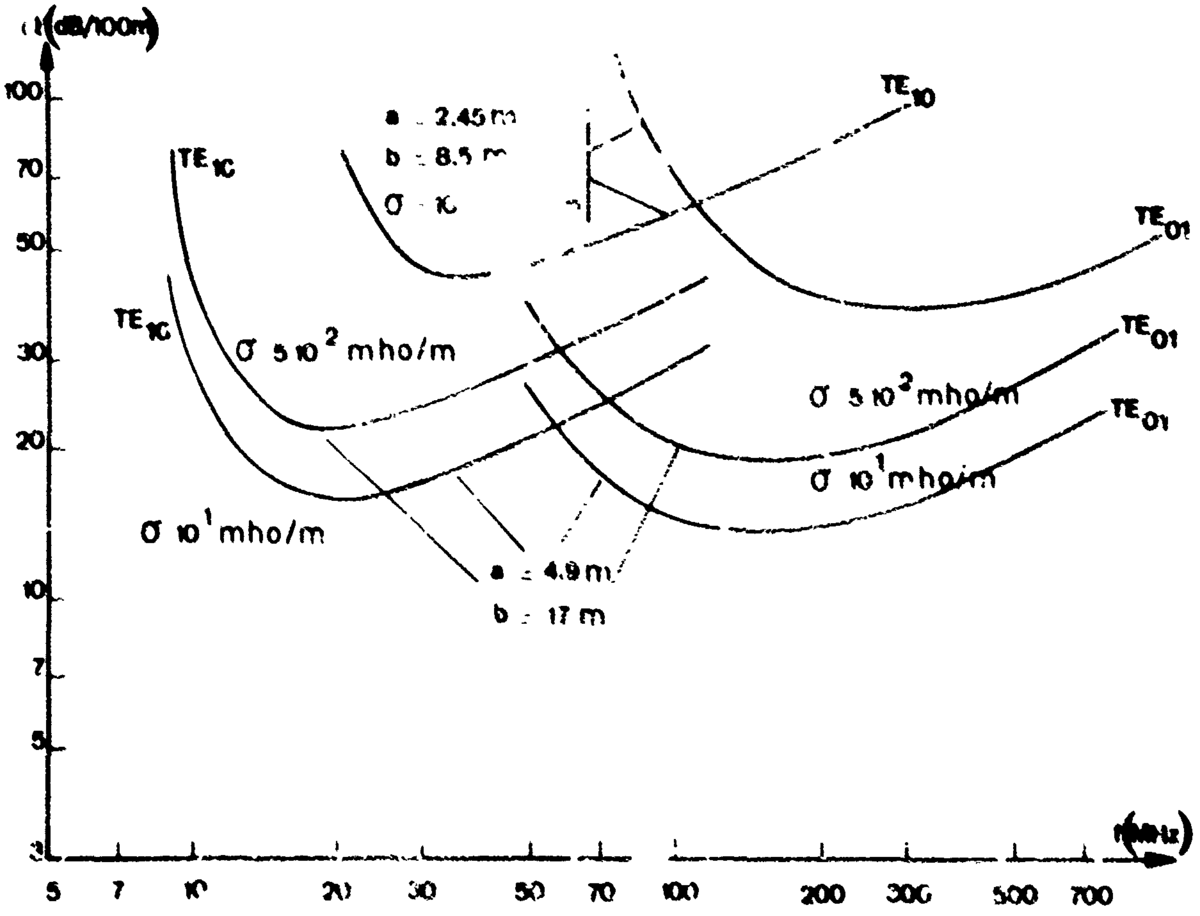


Fig.4 Influence of the transverse dimensions and wall conductivity on the theoretical attenuation of TE$_{10}$ and TE$_{01}$ modes in a rectangular tunnel

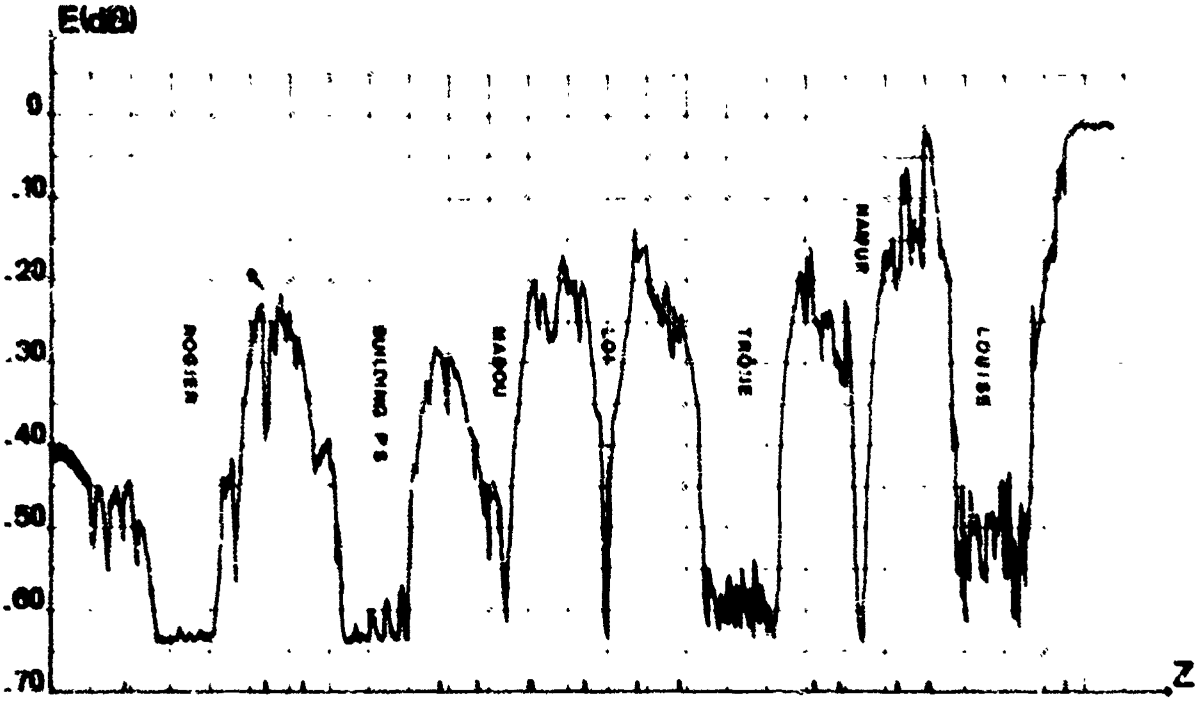


Fig.5 Field level of an AM radio broadcasting along the ring of Brussels. Tunnels are indicated by their names and by two vertical dashes. Frequency is 620 kHz

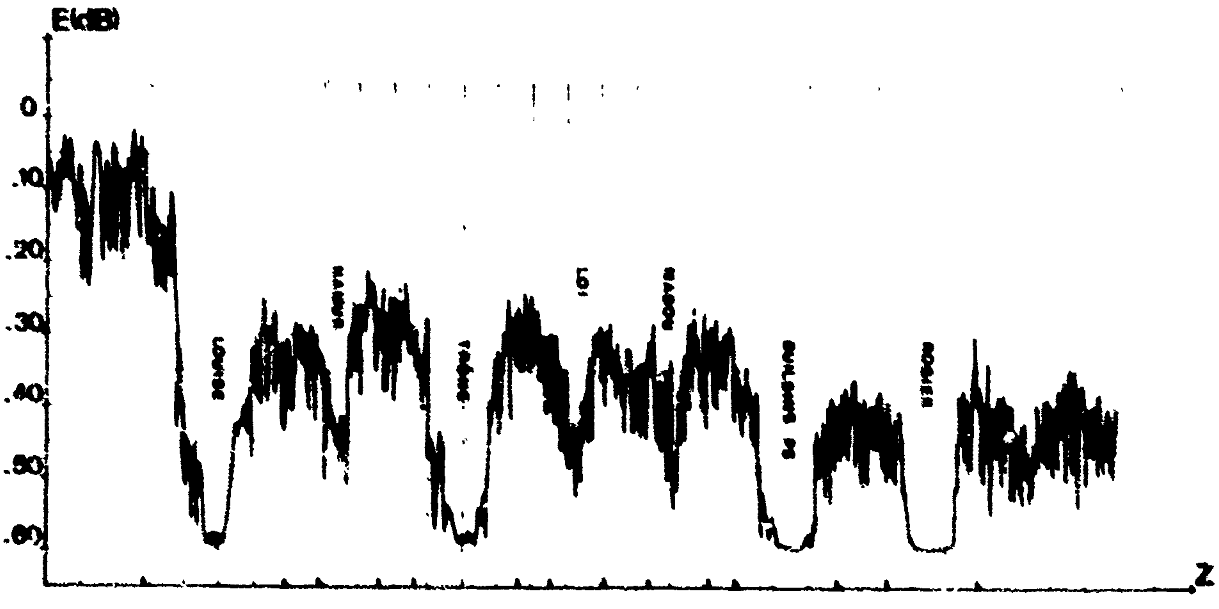


Fig.6 Field level of a FM radio broadcasting along the ring of Brussels. Tunnels are indicated by their names and by two vertical dashes. Frequency is 100 MHz.

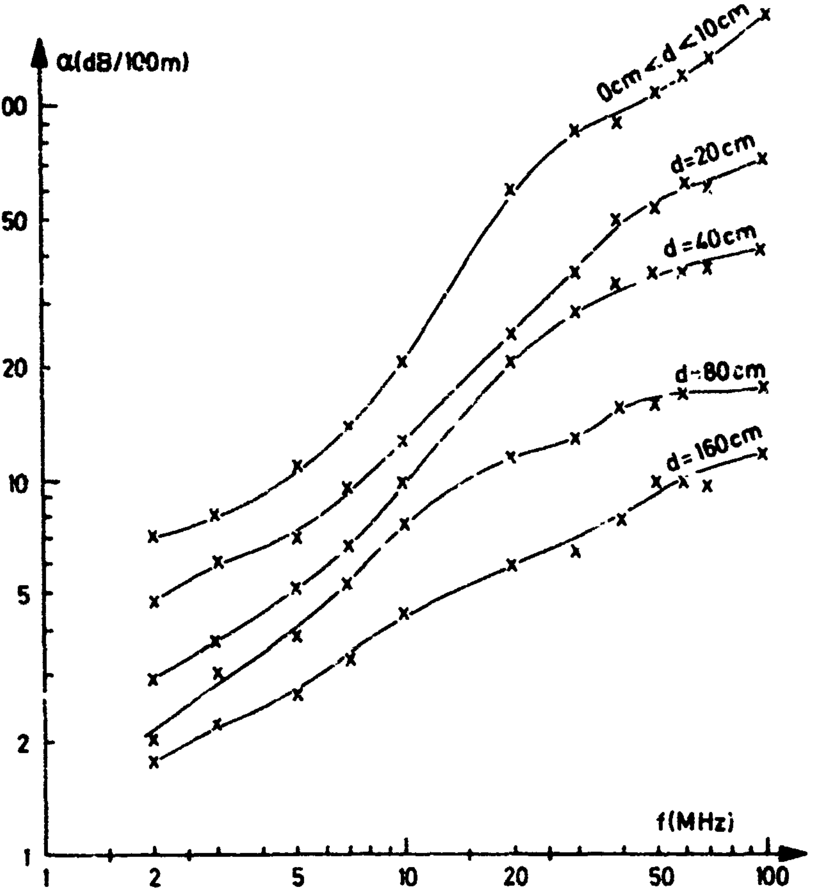


Fig.7 Field attenuation along a single-wire line in the tunnel of Lanaye

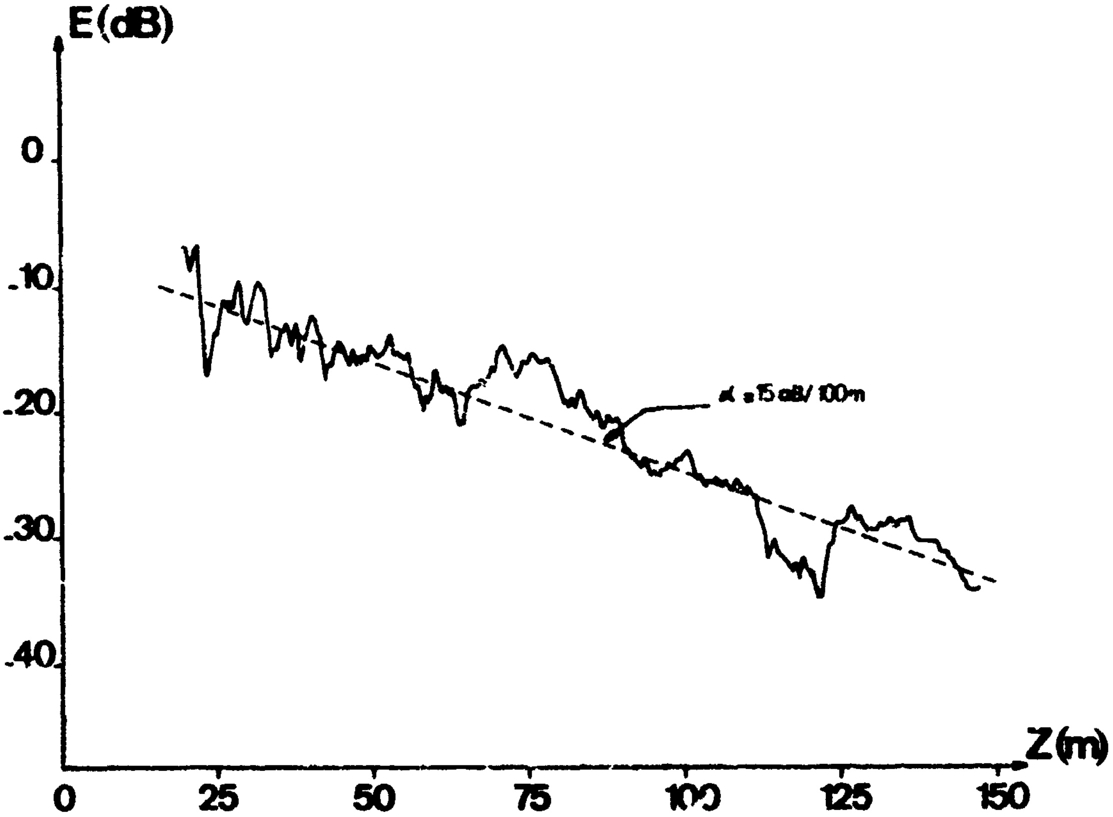


Fig.8 Field level in the tunnel of Lanaye when it contains a transmission line. Frequency is 450 MHz

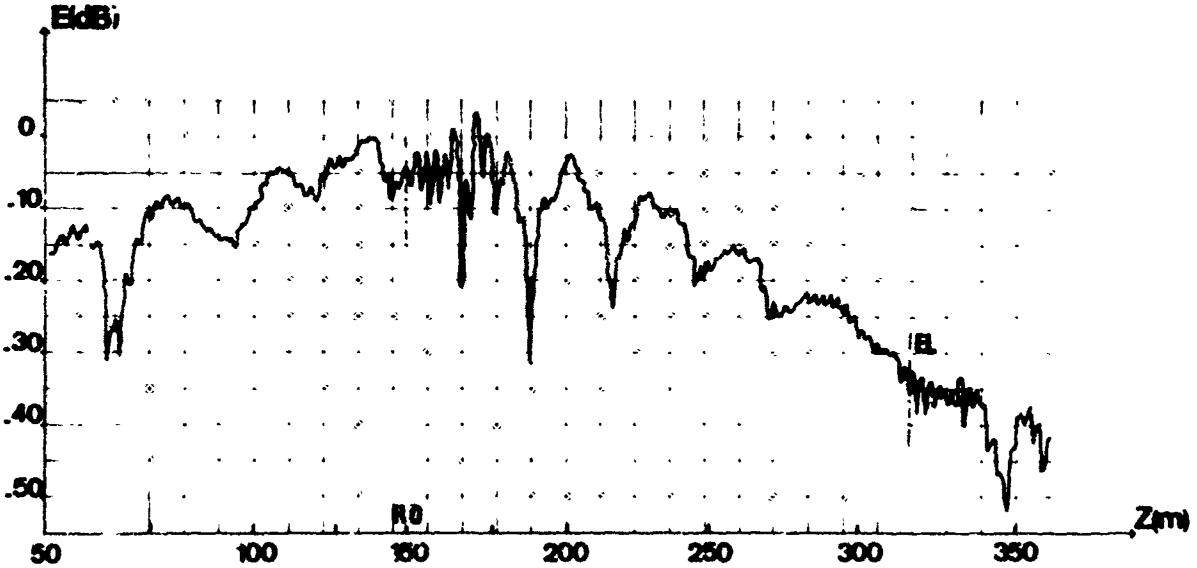


Fig.9 Field level in a road tunnel containing a transmission line with a radiating device (RD) Frequency is 70 MHz

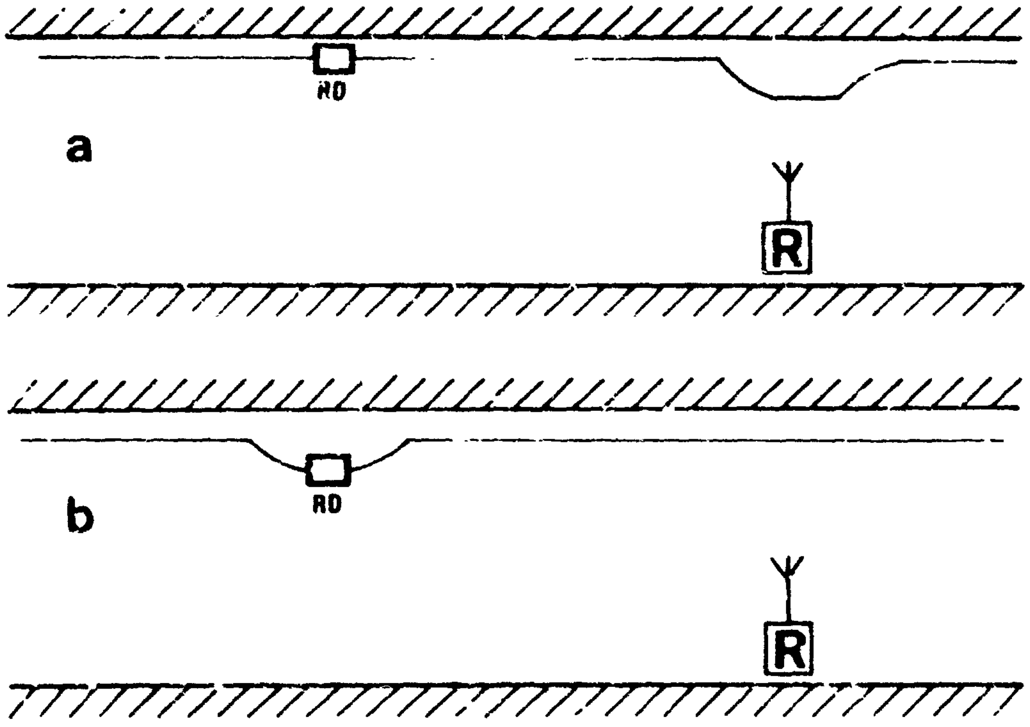


Fig.10 Line position in the tunnel

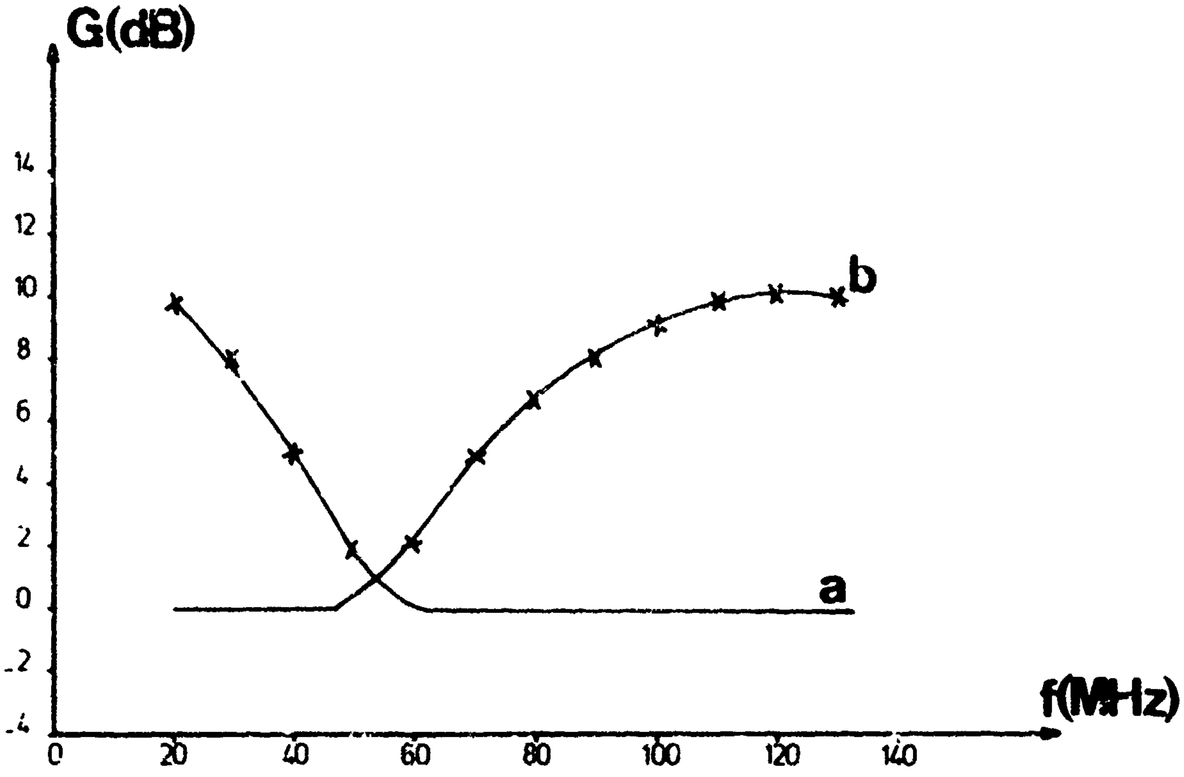


Fig.11 Influence of the line position on the field level versus frequency

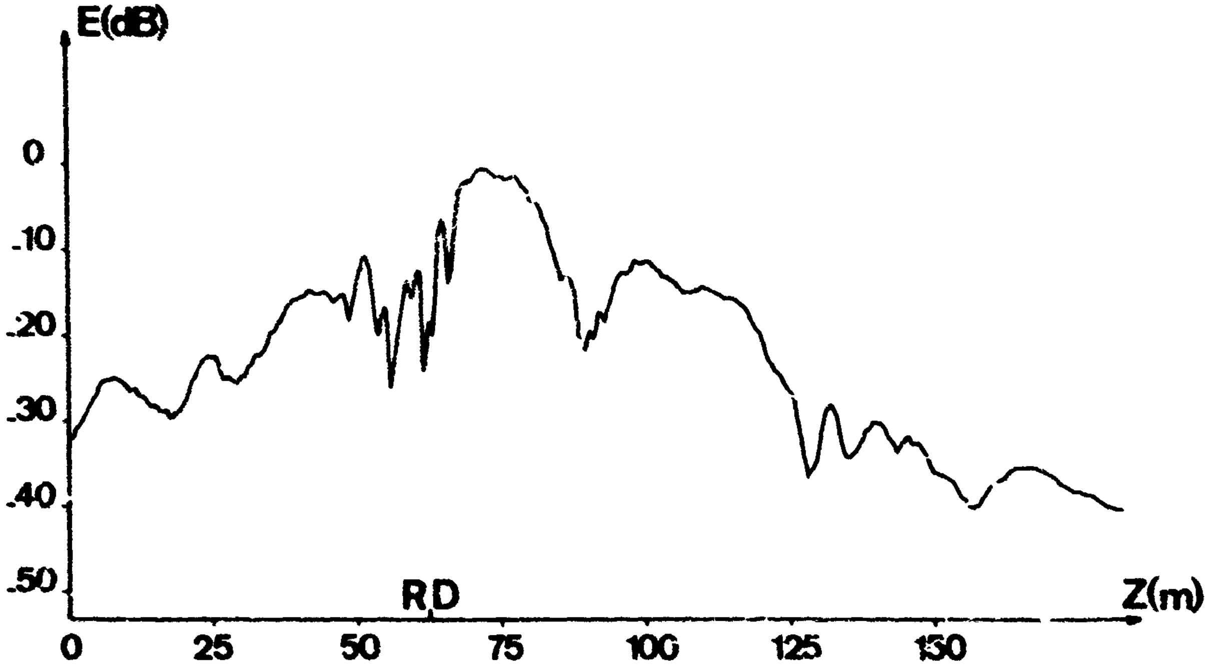


Fig.12 Field level in a road tunnel when the transmission line is laid in a gutter

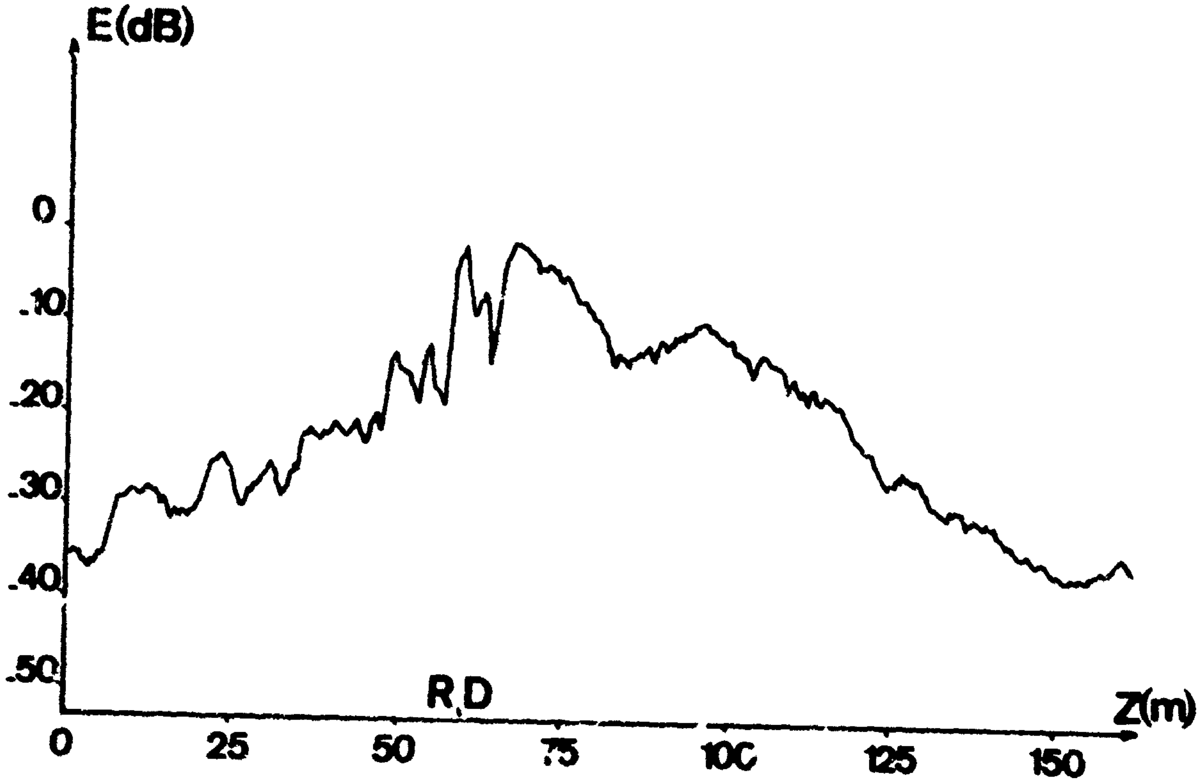


Fig.13 Field level in a road tunnel when the transmission line is hung at 10 cm from the roof

DISCUSSION

B.W. Reinisch, US

Are you saying, in conclusion, that transmission cables are of no real importance?

Author's Reply

I said that when natural modes are present, the line can be laid on the ground or in any position you wish, because it has no influence on the coupling loss between the transceivers and the radiating devices. However, the transmission line is needed to ensure the propagation of the coaxial mode between the radiating devices.

P.F. Checcacci, It

In 1949 the Italian State Railway made propagation experiments in a 18 Km double track tunnel using 3 GHz signals. The results were positive, although high attenuation arose when another train was in the tunnel

Author's Reply

I agree with these observations I made some experiments on the influence of vehicles on wave propagation in road tunnels, as I have described. The results from these experiments indicate that cars have practically no influence on the field level, but that big trucks generally provoke an important decrease of the field level I think it is strongly dependent on the ratio between the vehicle size and the tunnel size.

COMPARISON OF LOOP AND DIPOLE ANTENNAS IN LEAKY FEEDER COMMUNICATION SYSTEMS

David A. Hill and James R. Wait
U.S. Department of Commerce
Boulder, Colorado 80303

ABSTRACT

The mutual impedance between electric and/or magnetic dipoles in a circular tunnel containing a leaky coaxial cable is derived for arbitrary dipole positions. For the usual case of large dipole separation, the bifilar mode approximation is used to compute the mutual impedance and associated transmission loss. For typical cable parameters, a broad minimum in transmission loss is found in the range from 2 to 10 MHz. In this frequency range, the transmission loss is slightly less for magnetic dipoles (loops) than for electric dipoles. While the analytical model is a cable within a tunnel, the general approach and many of the conclusions would apply to similar above-ground systems. Thus the results are relevant to railway and roadway communications of the limited access type.

1. INTRODUCTION

The leaky feeder technique [Beal et al., 1973] has been developed for communication in mine tunnels [Martin, 1975]. The leaky transmission line is usually a coaxial cable [Fontain et al., 1973], and the energy is coupled into or out of the channel by antennas in the vicinity of the transmission line. Both electric dipole [Haining, 1974; Harms et al., 1974; Okada et al., 1975] and monopole [Delogne, 1974, Cree, 1975] antennas have been used. Small loops or coils [Breitenbach, 1974; Cribbens, 1974] have also been used to couple to the magnetic field.

In a previous paper [Hill and Wait, 1976], we calculated the mutual impedance and transmission loss between a pair of electric dipole antennas located in a circular tunnel which contained a leaky coaxial cable. The previous formulation utilized a reciprocity derivation [Mahmoud, 1974; Hill and Wait, 1974] which was useful only when the transmitting and receiving dipoles were widely separated. Here we employ a more direct derivation which allows arbitrary dipole locations and includes both electric and magnetic (small loop) dipole antennas. In addition to computing the transmission loss, we utilize the computed input resistance of the dipoles in the circular tunnel environment [Hill and Wait, 1978]. Thus the near field losses due to the lossy tunnel walls are accounted for.

When the dipoles are widely separated, the total transmission loss reduces to the sum of the dipole-to-cable coupling losses and the attenuation of the dominant mode. Numerical results are presented in order to compare the relative effectiveness of electric and magnetic dipole antennas.

2. DIPOLE EXCITATION

The cross section of the circular tunnel is shown in the cylindrical coordinate system (ρ,ϕ,z) in Figure 1. The air-filled tunnel of radius a_o has free space permittivity ε_o and permeability μ_o. The surrounding rock has conductivity σ_e, permittivity ε_e, and permeability μ_e.

The coaxial cable of radius c is centered at (ρ_c,ϕ_c), and it is assumed that c is small compared to both the free space wavelength and the distance of the cable from the tunnel wall. The specific leaky cable of interest here is modeled by: a center conductor of radius a and conductivity σ_w, insulation of outer radius b and permittivity ε, a thin shield of radius b that can be characterized by a transfer inductance per unit length L_T, a protective jacket of outer radius c and permittivity ε_c, and a thin lossy film of radius c that can be characterized by a conductivity-thickness product σd. Actually, the results given in this paper apply for any thin axial conductor which can be represented by a series impedance per unit length $Z(\lambda)$ where λ is the axial wavenumber. The appropriate expression for $Z(\lambda)$ for this coaxial cable model has been derived previously for use in the solution of the mode equation [Wait and Hill, 1975].

The x-directed dipole source at $(\rho_o,\phi_o,0)$ can be either electric or magnetic. The electric dipole source has a dipole moment $I\ell_{oe}$ where I is the current at the feed point and ℓ_{oe} is the effective length. The usual exp$(i\omega t)$ time dependence is suppressed. The magnetic dipole source has a magnetic dipole moment IA_{oe} where A_{oe} is the effective area.

The expressions for the fields of electric and magnetic dipoles of arbitrary orientation have been derived previously [Hill and Wait, 1978]. Here we consider only transverse orientations which generally are more effective in coupling to the leaky cable. In addition, x-directed orientation is sufficient to cover all transverse cases by proper selection of ϕ_o and ϕ_c.

The total fields at any point in the tunnel can be derived from electric and magnetic Hertz potentials U_t and V_t which are the z-components of the electric and magnetic Hertz vectors, respectively [Wait, 1959]. Furthermore, we can write

$$U_t = U + U_c \tag{1}$$

and

$$V_t = V + V_c \tag{2}$$

U and V are the Hertz potentials that would exist in an empty tunnel, while U_c and V_c are the contributions (outside the cable) due to the axial cable current. The expressions for U_c and V_c have been derived previously [Hill and Wait, 1978], and the derivation will not be repeated here. The specific expressions for electric and magnetic dipole excitation are given in Appendices A and B.

The components of the total field are obtained from U_t and V_t by the well known [Wait, 1959] relations:

$$E_\rho = \frac{\partial^2 U_t}{\partial\rho\partial z} - \frac{i\omega\mu_o}{\rho}\frac{\partial V_t}{\partial\phi} , \tag{3}$$

$$E_\phi = \frac{1}{\rho}\frac{\partial^2 U_t}{\partial\phi\partial z} + i\omega\mu_o \frac{\partial V_t}{\partial\rho} , \tag{4}$$

$$E_z = \left(\frac{\partial^2}{\partial z^2} - \gamma_o^2\right)U_t , \tag{5}$$

$$H_\rho = \frac{i\omega\epsilon_o}{\rho}\frac{\partial U_t}{\partial\phi} + \frac{\partial^2 V_t}{\partial\rho\partial z} , \tag{6}$$

$$H_\phi = -i\omega\epsilon_o \frac{\partial U_t}{\partial\rho} + \frac{1}{\rho}\frac{\partial^2 V_t}{\partial\phi\partial z} , \tag{7}$$

$$H_z = \left(\frac{\partial^2}{\partial z^2} - \gamma_o^2\right)V_t , \tag{8}$$

where $\gamma_o^2 = -\omega^2\mu_o\epsilon_o$.

3. MUTUAL IMPEDANCE

We consider now a small receiving dipole of arbitrary transverse orientation located at (ρ_r,ϕ_r,z_r) as indicated in Figure 1. For a receiving electric dipole of effective vector length $\vec{\ell}_{re}$, the received voltage v_r is

$$v_r = -\vec{\ell}_{re} \cdot \vec{E}_t \tag{9}$$

where

$$\vec{E}_t = \hat{\rho}E_\rho + \hat{\phi}E_\phi .$$

Thus the mutual impedance Z_m is

$$Z_m = v_r/I = -\vec{\ell}_{re} \cdot \vec{E}_t/I . \tag{10}$$

For a receiving magnetic dipole (loop of effective area A_{re}), the received voltage v_r is

$$v_r = i\omega\mu_o A_{re}\, \hat{n} \cdot \vec{H}_t ,$$

where (11)

$$\vec{H}_t = \hat{\rho}H_\rho + \hat{\phi}H_\phi ,$$

and $\hat{n}$ is a unit normal to the loop surface. Thus the mutual impedance Z_m for a small receiving loop is

$$Z_m = v_r/I = i\omega\mu_o A_{re}\, \hat{n} \cdot \vec{H}_t/I . \tag{12}$$

At this point, there is no restriction on the dipole locations (provided they are inside the tunnel and outside the cable). The formulation is also valid for mixed dipole types, such as electric source dipole and magnetic receiving dipole. When the dipoles are located close together (small $|z_r|$), the integral forms for the fields (as given in Appendices A and B) must be evaluated numerically in order to compute the mutual impedances. This is a rather tedious task, but it has been done in order to evaluate the change in the self impedance of electric and magnetic dipoles due to the tunnel walls and leaky cable [Hill and Wait, 1978].

4. BIFILAR MODE APPROXIMATION

For large values of $|z|$, the dominant contribution to the fields is from the mode of lowest attenuation. For most cases of interest, this is the "bifilar" mode which carries most of its energy between the center conductor and the braid but also has leakage fields outside the cable. The propagation constant λ_b of this mode satisfies the mode equation $D(\lambda_b) = 0$ where $D(\lambda)$ is given by (B-7). This mode equation has been solved numerically [Wait and Hill, 1975] and analytically for the quasi-static case [Seidel and Wait, 1978a]. The propagation constant λ_b is approximately that of the cable insulation, i.e. $\lambda_b \simeq \omega(\mu_o\epsilon)^{1/2}$.

The approximate expressions for the Hertz potentials for large $|z|$ are obtained by contour integration of (B-1) and (B-2). If only the pole contribution from the bifilar mode (at $\lambda = \lambda_b$) is included, the approximate expressions for U_t and V_t are found to be

$$U_t = -2\pi i \frac{\hat{E}_z(\lambda_b)\exp(-i\lambda_b|z|)}{[dD(\lambda)/d\lambda]\big|_{\lambda=\lambda_b}} \cdot \left[K_o(v_b\rho_o) - \sum_{m=-\infty}^{\infty} R_m(\lambda_b) \times \frac{K_m(v_b a_o)}{I_m(v_b a_o)} I_m(v_b\rho) I_m(v_b\rho_c)\exp[-im(\phi-\phi_c)]\right] \tag{13}$$

and

$$V_t = -2\pi i \frac{\hat{E}_z(\lambda_b)\exp(-i\lambda_b|z|)}{[dD(\lambda)/d\lambda]\big|_{\lambda=\lambda_b}} \sum_{m=-\infty}^{\infty} I_m(v_b\rho_c)\Delta_m(\lambda_b) \times I_m(v_b\rho)\exp[-im(\phi-\phi_c)] \quad , \tag{14}$$

where

$$v_b = (\lambda_b^2+\gamma_o^2)^{1/2} \quad .$$

Our previous solution of the mode equation, $D(\lambda) = 0$, utilized a version of Newton's method [Hamming, 1973] which yields the required derivative $dD(\lambda)/d\lambda$ with no additional effort. If other modes are needed, they can be handled in the same fashion. Then (13) and (14) would be a modal sum over λ_p where λ_p are the solutions of $D(\lambda_p) = 0$.

For the mine communication application, the bifilar mode approximation is sufficient. The field components at large $|z|$ are approximately given by substituting (13) and (14) into (3)-(8). The approximate expressions for the mutual impedance of widely spaced dipoles are, in turn, given by substituting the resultant field expressions into (10) or (12). When this is done, the result in (10) agrees with the previous approximation for the mutual impedance of electric dipoles as obtained by reciprocity [Hill and Wait, 1976].

5. TRANSMISSION LOSS

If the input resistance of the transmitting antenna is R_o, then the input power P_o is

$$P_o = |I|^2 R_o/2 \quad . \tag{15}$$

If the input resistance of the receiving antenna is R_r and the antenna is terminated in a matched load, then the received power P_r is

$$P_r = |I|^2 |Z_m|^2/(8R) \quad . \tag{16}$$

Thus the transmission loss L (in dB) is

$$L = 10 \log_{10}\left(\frac{P_o}{P_r}\right) = 10 \log_{10}\left(\frac{4R_o R_r}{|Z_m|^2}\right) \tag{17}$$

Since the transmitter power is limited in mine communications, it is the transmission loss which is of most interest to the systems engineer. The expression in (17) applies to both electric and magnetic dipoles as either transmitting or receiving antennas.

In order to compare the relative merits of dipole and loop antennas, a computer program was written for transmission loss between either a pair of electric dipoles or between a pair of magnetic dipoles. The required mutual impedances were computed from (10) and (12) using the bifilar mode approximation in (13) and (14). For the input resistances, the actual computed values in the tunnel environment [Hill and Wait, 1978] were used. Thus the near field loss resistance due to the tunnel walls was accounted for.

In all numerical results, the following parameters were used. For the tunnel, $a_o = 2$m, $\sigma_e = 10^{-3}$mho/m, $\epsilon_e/\epsilon_o = 10$, and $\mu_e/\mu_o = 1$. The cable was located at $\rho_c/a_o = 0.9$, and the parameters adopted were: $a = 1.5$ mm, $b = 10$ mm, $c = 10$ mm, $\sigma_w = 5.7 \times 10^7$mho/m, $\epsilon/\epsilon_o = 1.5$, and $\sigma_d = 0$. Thus the dielectric jacket and the lossy thin jacket have been excluded since they have been found to have a negligible effect.

Both electric dipoles were x-directed with $\phi_o=\phi_r=\phi_c=0^\circ$ for maximum cable coupling. The effective dipole lengths, ℓ_{oe} and $|\ell_{re}|$, were both taken to be 0.5m. This could correspond to a physical length of 1m with the usual triangular current distribution for electrically short dipoles.

Both magnetic dipoles were x-directed with $\phi_o=\phi_r=\phi_c=90°$ for maximum cable coupling. The loop areas, A_{oe} and A_{re}, were both taken to be 1m^2.

In Figure 2, the transmission loss L is shown for two dipole locations, $\rho_o/a_o = \rho_r/a_o = 0.5$ and $\rho_o = \rho_r = 0$. The transmission loss for the dipoles at the center of the tunnel ($\rho_o=\rho_r=0$) is larger simply because the dipoles are farther from the cable. The curves for various separations z_r show the effect of the attenuation of the bifilar mode. The attenuation even at 2 km is seen to be quite small at 1 MHz but much larger at 50 MHz. The curves for $z_r = 0$ are shown merely to illustrate the two-way coupling loss to and from the cable. This coupling loss is fairly small because we have taken a rather large value of 40 nH/m for the surface transfer inductance L_T of the cable braid [Fontaine et al., 1973]. Also, the low value of $\epsilon/\epsilon_o = 1.5$ yields a fairly high velocity for the bifilar mode which serves to decrease the coupling loss. In addition, the larger coupling loss quoted by many cable manufacturers applies to antennas located 20 feet (≈ 6m) from the cable. The antenna-cable separations in Figure 2 are only 1.8m and 0.8m. Thus the parameters and the configuration represented in Figure 2 represent a very favorable situation. Still the results are in reasonable agreement with the free space coupling loss calculations of Rawat and Beal [1974] using somewhat different assumptions.

In Figure 3, we show the corresponding transmission loss for a pair of identical magnetic dipoles (loops). The results are similar to those of the electric dipole, and the transmission loss is slightly lower for the low frequencies.

As seen by (17), the transmission loss for matched loads depends only on the mutual impedance and the input resistances. Since matched loading is difficult for electrically small antennas, the mutual impedances and input resistances are of interest in their own right. In Figure 4, the input resistances and the magnitude of the mutual impedances are shown for one dipole location. Below about 20 MHz the loss resistance dominates the radiation resistance for both electric and magnetic dipoles [Hill and Wait, 1978]. The result is that the electric dipole resistance is nearly independent of frequency and the magnetic dipole resistance is proportional to the square of the frequency. Another point is that both the input resistance and the mutual impedance of short electric dipoles are proportional to the length squared. For small loops (magnetic dipoles), both the input resistance and mutual impedance are proportional to the loop area squared. Thus, for the adopted model, the transmission loss is independent of dipole length or loop area as indicated by (17). Of course, this simple result neglects any difficulties in antenna matching or any wire resistance of the antennas.

In Figures 5 and 6, the transmission losses for electric and magnetic dipoles are shown for smaller values of L_T. A value of L_T = 10 nH/m is probably more realistic for most cable shields, and a value of L_T = 2 nH/m represents a very low leakage cable braid. A decrease in L_T causes a large increase in coupling loss as seen in Figures 5 and 6 and also a decrease in attenuation of the bifilar mode. Since the coupling loss dominates the attenuation loss for most cases of interest, a large value of L_T (very leaky cable) is most favorable. For very large dipole separations, z_t, this conclusion would no longer hold.

5. CONCLUDING REMARKS

The mutual impedance between electric and/or magnetic dipoles has been derived for arbitrary dipole positions and orientations. For the usual case of large dipole separation, the single bifilar mode approximation is adequate. Numerical results for transmission loss indicate that a cable with a large transfer inductance and a low dielectric constant is most favorable. The large leakage fields of such a cable results in a slightly higher attenuation rate, but this is more than offset by a large reduction in coupling loss.

For typical parameters, a broad minimum in transmission loss is observed somewhere between 2 and 10 MHz. This is in agreement with Fontaine et al. [1973] who find 7 MHz to be an optimum frequency. In this frequency range, the transmission loss is somewhat lower for magnetic dipoles than for electric dipoles. This difference results primarily from the lower loss resistance (due to the tunnel walls) of the magnetic dipole. At higher frequencies where the loss resistance is less important, this advantage disappears.

The analysis given here assumes a uniform tunnel and cable with no mode conversion. A remaining extension would be to include the effects of mode conversion between the monofilar bifilar modes. Such mode conversion can be either intentional due to cable design [Deryck, 1975, Seidel and Wait, 1978b] or inadvertant due to nonuniformities in the tunnel walls [Seidel and Wait, 1978a]. However, it is doubtful if this more complicated analysis would modify our general conclusions on the relative merits of loop and dipole antennas in mine communications.

ACKNOWLEDGMENTS

The problem was actually suggested by Dr. H.K. Sacks and we appreciate his sustained interest in this work. Also, we would like to thank Mrs. Lana Hope for help in preparing the manuscript and Mr. Robert McDonald for drafting the figures.

REFERENCES

BEAL, J.C., JOSIAK, J., MAHMOUD, S.F., 1973, "Continuous-access guided communications (CAGC) for ground-transportation systems," Proc. IEEE, 61(5), 562-568

BREITENBACH, O., 1974, "Leaky coaxial cables for radio communication to traffic services," in Proc. Int. Colloquium Leaky-Feeder Communications, Surrey, England, 197-225.

CREE, D.J., 1975, "Practical performance of radiating cables," Radio Electron. Eng., 45(5), 215-223.

CRIBBENS, A.H., 1974, "Some properties of slotted coaxial waveguides," in Proc. Int. Colloquium Leaky-Feeder Communications, Surrey, England, 250-262.

DELOGNE, P., 1974, "The INIEX mine communications system," in Proc. Int. Conf. Radio: Roads, Tunnels and Mines, 129-136.

DERYCK, L., 1975, "Control of mode conversions on bifilar lines in tunnels," Radio Electron. Eng., 45(5), 241-247.

FONTAINE, J., DE MOULIN, B., DEGAUQUE, P., and GABILLARD, R., 1973, "Feasibility of radio communication in mine galleries by means of a coaxial cable having a high coupling impedance," in Proceedings of Thru-the-Earth Electromagnetics Workshop, National Technical Information Service, Springfield, VA 22151, Acct. No. PB 231 154, pp. 130-139.

HAINING, K.[illegible], 1974, "Experimental modelling of leaky feeders," in Proc. Int. Colloquium Leaky Feeder Communications, Surrey, England, 236-249.

HAMMING, R.W., 1973, Numerical Methods for Scientists and Engineers, McGraw-Hill, New York.

HARMS, P.L., JOHANNESSEN, R., and FERNANDES, A.S.C., 1974, "Some engineering considerations affecting the use of radiating cables in communications between vehicles and traffic control centres," in Proc. Int. Colloquium Leaky-Feeder Communications, Surrey, England.

HILL, D.A., and WAIT, J.R., 1974, "Excitation of monofilar and bifilar modes on a transmission line in a circular tunnel," J. Appl. Phys., 45(8), 3402-3406.

HILL, D.A., and WAIT, J.R., 1976, "Calculated transmission loss for a leaky feeder communication system in a circular tunnel," Radio Sci., 4(11), 315-321.

HILL, D.A. and WAIT, J.R., 1978, "The impedance of dipoles in a circular tunnel with an axial conductor," IEEE Trans. Geoscience Electronics, GE-16(2), 118-126.

MAHMOUD, S.F., 1974, "Characteristics of EM guided waves for communication in coal mines," IEEE Trans. Commun., COM-22(10), 1547-1554.

MARTIN, D.J.R., 1975, "A general study of the leaky feeder principle," Radio Electron. Eng., 45(5), 205-214.

OKADA, O., KISHIMOTO, T., ADAGAWA, K., NAKAHARA, Y., MIKOSHIBA, K., HORIGUCHI, F., and OKAMOTO, K., 1975, "Leaky coaxial cable for communication in high speed railway transportation," Radio Electron. Eng., 45(5), 224-228.

RAWAT, V., and BEAL, J.C., 1974, "Leaky cables treated as open waveguides," Proceedings of the Colloquium on Leaky-Feeder Radio Communications held from 8-10 April 1974, 108-124, University of Surrey, Guildford.

SEIDEL, D.B. and WAIT, J.R., 1978a, "Transmission modes in a braided coaxial cable and coupling to a tunnel environment," IEEE Trans. Microwave Theory Tech., MTT-26(7), 494-499.

SEIDEL, D.B. and WAIT, J.R., 1978b, "Role of controlled mode conversion in leaky feeder communication systems," IEEE Trans. Antennas Propagat., AP-26(5), 690-694.

SEIDEL, D.B. and WAIT, J.R., 1978c, "Mode conversion by tunnel non-uniformities in leaky feeder communication systems," Preliminary Report on Contract H0155008 to U.S. Bureau of Mines, 1 September 1978.

WAIT, J.R., 1959, Electromagnetic Radiation from Cylindrical Structures, Pergamon, New York.

WAIT, J.R. and HILL, D.A., 1975, "Propagation along a braided coaxial cable in a circular tunnel, IEEE Trans. Microwave Theory Tech., MTT-23(5), 401-405.

APPENDIX A - EMPTY TUNNEL FIELDS

The empty tunnel fields are obtained from the Hertz potentials U and V which are given by [Hill and Wait, 1978]

$$U = \Gamma[A_m(\lambda)K_m(v\rho) + P_m(\lambda)I_m(v\rho)] \quad , \tag{A-1}$$

$$V = \Gamma[B_m(\lambda)K_m(v\rho) + Q_m(\lambda)I_m(v\rho)] \quad , \tag{A-2}$$

where

$$\Gamma[\quad] = \int_{-\infty}^{\infty} \sum_{-\infty}^{\infty} [\quad]\exp[-im(\phi-\phi_o)]\exp(-i\lambda z)d\lambda \quad ,$$
$$v = (\lambda^2+\gamma_o^2)^{1/2} \quad , \quad \gamma_o = i\omega(\mu_o\epsilon_o)^{1/2} \quad , \tag{A-3}$$

and I_m and K_m are modified Bessel functions. The integration contour is taken along the real axis in the complex λ plane, and the summation is taken over all integer values of m from $-\infty$ to $+\infty$.

The terms involving K_m represent the primary field (for $\rho_o<\rho<a_o$). For the electric dipole source, A_m and B_m are given by

$$A_m(\lambda) = \frac{-I\ell_{oe}}{4\pi^2\omega\epsilon_o}\left(\frac{\lambda}{2v}\right)[\exp(i\phi_o)I_{m+1}(v\rho_o) + \exp(-i\phi_o)I_{m-1}(v\rho_o)] \quad . \tag{A-4}$$

$$B_m(\lambda) = \frac{I\ell_{oe}}{4\pi^2}\left(\frac{i}{2v}\right)[\exp(-i\phi_o)I_{m-1}(v\rho_o) - \exp(i\phi_o)I_{m+1}(v\rho_o)] \quad . \tag{A-5}$$

For the magnetic dipole source, A_m and B_m are given by

$$A_m(\lambda) = \frac{i\omega\mu_o IA_{oe}}{4\pi^2}\left(\frac{i}{2v}\right)[\exp(-i\phi_o)I_{m-1}(v\rho_o) - \exp(i\phi_o)I_{m+1}(v\rho_o)] \quad . \tag{A-6}$$

$$B_m(\lambda) = \frac{IdA}{4\pi^2}\left(\frac{-i\lambda}{2v}\right)[\exp(i\phi_o)I_{m+1}(v\rho_o) + \exp(-i\phi_o)I_{m-1}(v\rho_o)] \quad . \tag{A-7}$$

The terms involving I_m in (A-1) and (A-2) represent the secondary fields due to the tunnel walls. P_m and Q_m are related to A_m and B_m by the following

$$P_m(\lambda) = -\left\{A_m(\lambda)\left[\left(\frac{m\lambda}{v^2a_o} - \alpha_m\right)^2 + \left(\frac{\gamma_o K_m'}{vK_m} + \eta_o Y_m\right)\left(\frac{\gamma_o I_m'}{vI_m} + \frac{Z_m}{\eta_o}\right)\right]I_mK_m + B_m(\lambda)\left(\frac{m\lambda}{v^2a} - \alpha_m\right)\frac{i\omega\mu_o}{v^2a}\, D_m^{-1}\right\} , \tag{A-8}$$

$$Q_m(\lambda) = -\left\{B_m(\lambda)\left[\left(\frac{m\lambda}{v^2a_o} - \alpha_m\right)^2 + \left(\frac{\gamma_o I_m'}{vI_m} + \eta_o Y_m\right)\left(\frac{\gamma_o K_m'}{vK_m} + \frac{Z_m}{\eta_o}\right)\right]I_mK_m - A_m(\lambda)\left(\frac{m\lambda}{v^2a_o} - \alpha_m\right)\frac{i\omega\epsilon_o}{v^2a}\right\} D_m^{-1} , \tag{A-9}$$

where

$$D_m = \left[\left(\frac{m\lambda}{va_o} - \alpha_m\right)^2 + \left(\frac{\gamma_o I_m'}{vI_m} + \eta_o Y_m\right)\left(\frac{\gamma_o I_m'}{vI_m} + \frac{Z_m}{\eta_o}\right)\right]I_m^2 \quad , \quad \alpha_m = m\lambda/(u^2a_o) \quad ,$$

$$Y_m = [i\gamma_e^2/(u\omega\mu_e)]K_m'(ua_o)/K_m(ua_o) \quad , \gamma_e^2 = i\omega\mu_e(\sigma_e + i\omega\epsilon_e) \quad ,$$

$$Z_m = -(i\omega\mu_e/u)K_m'(ua_o)/K_m(ua_o),$$

and

$$u = (\lambda^2 + \gamma_e^2)^{1/2} .$$

In (A-8) and (A-9), the following notational simplicity has been used: $I_m = I_m(va_o)$, $I_m' = I_m'(va_o)$, $K_m = K_m(va_o)$, and $K_m' = K_m'(va_o)$.

APPENDIX B - CABLE FIELDS

The fields due to the axial cable current are obtained from the Hertz potentials U_c and V_c which are given by [Hill and Wait, 1978].

$$U_c = \int_{-\infty}^{\infty} F_c(\lambda)\Big[K_o(v\rho_d) - \sum_{m=-\infty}^{\infty} R_m(\lambda)\,\frac{K_m(va_o)}{I_m(va_o)}\, I_m(v\rho) I_m(v\rho_c)\exp[-im(\phi-\phi_c)]\Big]\exp(-i\lambda z)\,d\lambda \ , \tag{B-1}$$

$$V_c = \int_{-\infty}^{\infty} F_c(\lambda) \sum_{m=-\infty}^{\infty} I_m(v\rho_c)\Lambda_m(\lambda) I_m(v\rho)\exp[-im(\phi-\phi_c)]\exp(-i\lambda z)\,d\lambda \ , \tag{B-2}$$

where

$$\rho_d = [\rho^2 + \rho_c^2 - 2\rho\rho_c \cos(\phi-\phi_c)]^{1/2} \ .$$

To satisfy the boundary conditions at the tunnel walls, $R_m(\lambda)$ and $\Lambda_m(\lambda)$ are given by

$$-R_m(\lambda)\,\frac{K_m(va_o)}{I_m(va_o)} = \frac{P_m(\lambda)}{\Delta_m(\lambda)}\Bigg|_{B_m=0} \tag{B-3}$$

and

$$\Lambda_m(\lambda) = \frac{Q_m(\lambda)}{\Delta_m(\lambda)}\Bigg|_{B_m=0} \ , \tag{B-4}$$

where $P_m(\lambda)$ and $Q_m(\lambda)$ are given by (A-8) and (A-9)

The factor $F_c(\lambda)$ is given by

$$F_c(\lambda) = \hat{E}_z(\lambda)/D(\lambda), \tag{B-5}$$

where

$$\hat{E}_z(\lambda) = \hat{E}_z^p(\lambda) - v^2 \sum_{m=-\infty}^{\infty} P_m(\lambda) I_m(v\rho_c)\exp[-im(\phi_c-\phi_o)] \tag{B-6}$$

$$D(\lambda) = 2\pi i\omega\epsilon_c Z(\lambda) + v^2\Big[K_o(vc) - \sum_{m=-\infty}^{\infty} R_m(\lambda)\,\frac{K_m(va_o)}{I_m(va_o)}\, I_m(v\rho_c) I_m[v(\rho_c+c)]\Big] \ . \tag{B-7}$$

The modal equation to determine the discrete propagation modes in the tunnel [Wait and Hill, 1975] is simply $D(\lambda) = 0$.

The quantity $\hat{E}_z^p(\lambda)$ is actu[illegible] the Fourier transform of the axial electric field due to the direct or primary influence of the source di[illegible], evaluated at the cable. For an x-directed electric dipole, $\hat{E}_z^p(\lambda)$ is given by

$$\hat{E}_z^p(\lambda) = \frac{I\ell_{oe}}{4\pi^2\omega\epsilon_o}\,(\lambda v)\,\frac{\rho_c\cos\phi_c - \rho_o\cos\phi_o}{\hat{\rho}_c}\,K_1(v\hat{\rho}_c) \ . \tag{B-8}$$

Finally, for an x-directed magnetic dipole, $\hat{E}_z^p(\lambda)$ is given by

$$\hat{E}_z^p(\lambda) = -\frac{i\omega\mu_o I A_{oe}}{4\pi^2}\, v\,\frac{\rho_c\sin\phi_c - \rho_o\sin\phi_o}{\hat{\rho}_c}\,K_1(v\hat{\rho}_c) \ . \tag{B-9}$$

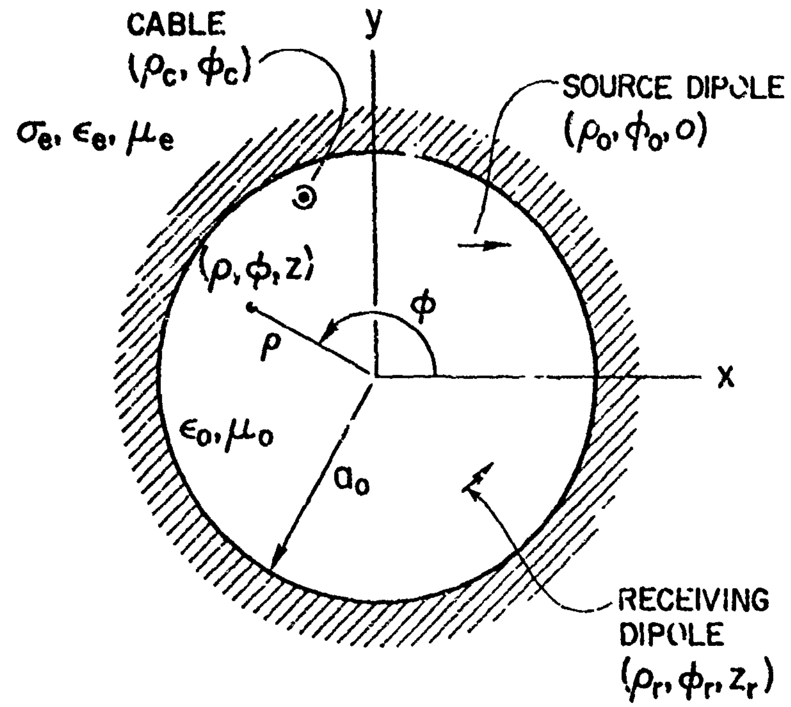

CABLE
(ρ_c, ϕ_c)
$\sigma_e, \epsilon_e, \mu_e$
y
SOURCE DIPOLE
$(\rho_0, \phi_0, 0)$
(ρ, ϕ, z)
ρ
ϕ
X
ϵ_0, μ_0
a_0
RECEIVING
DIPOLE
(ρ_r, ϕ_r, z_r)

Fig 1 Transmitting and receiving dipoles (electric or magnetic) in a circular tunnel which contains a leaky coaxial cable

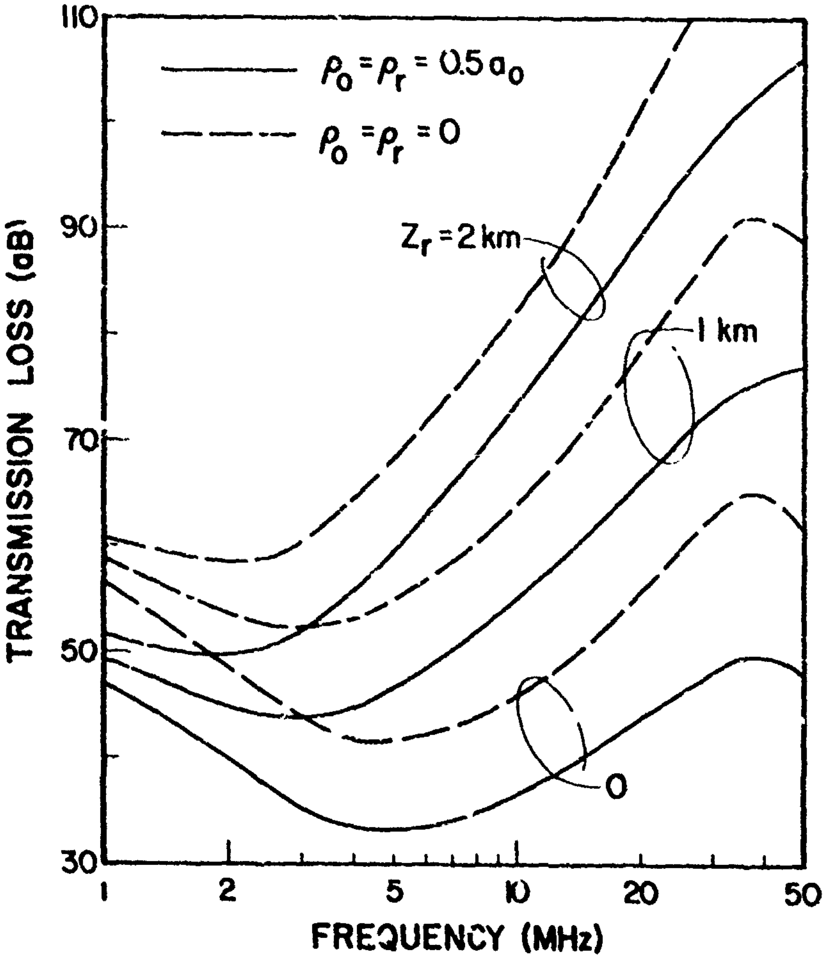


Fig.2 Transmission loss between a pair of identical electric dipoles for various dipole locations and axial separations The curves for $z_r = 0$ represent the two-way coupling loss in and out of the cable A high value of 40 nH/m is used for the surface transfer inductance of the cable braid

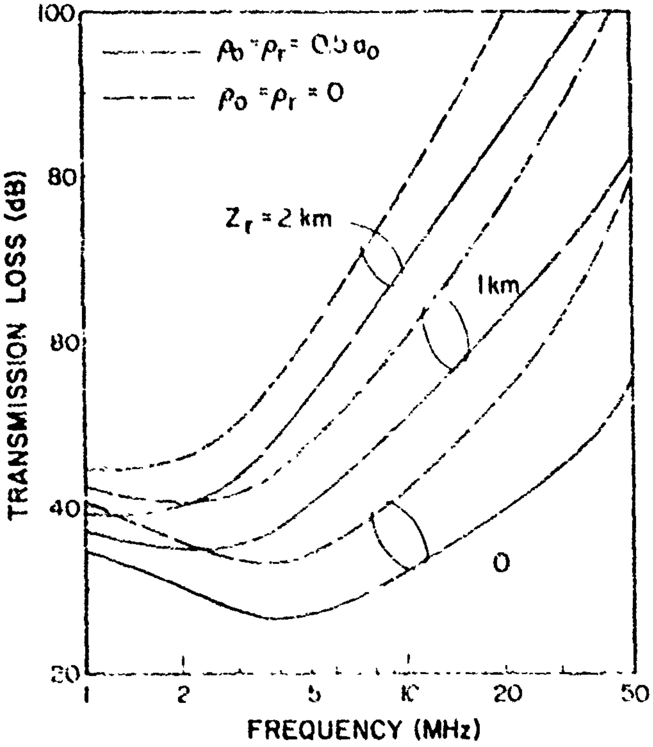


Fig.3 Transmission loss between a pair of identical magnetic dipoles (loops) for various dipole locations and axial separations. A high value of 40 nH/m is used for the surface transfer inductance of the cable braid.

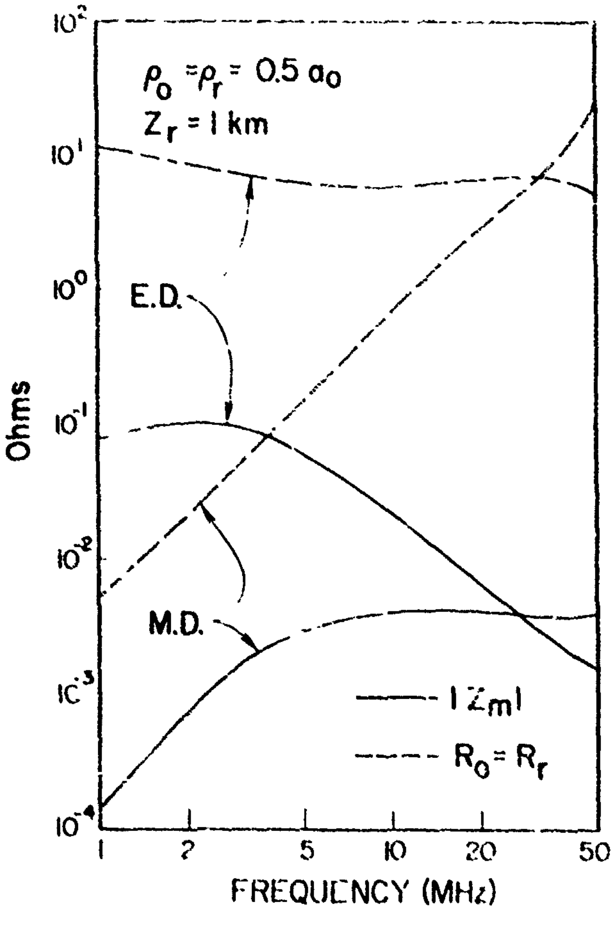


Fig.4 Mutual impedance, Z_m, and input resistance, $R_o = R_r$, for electric dipoles (E.D.) and magnetic dipoles (M.D.). Azam, $L_T = 40$ nH/m

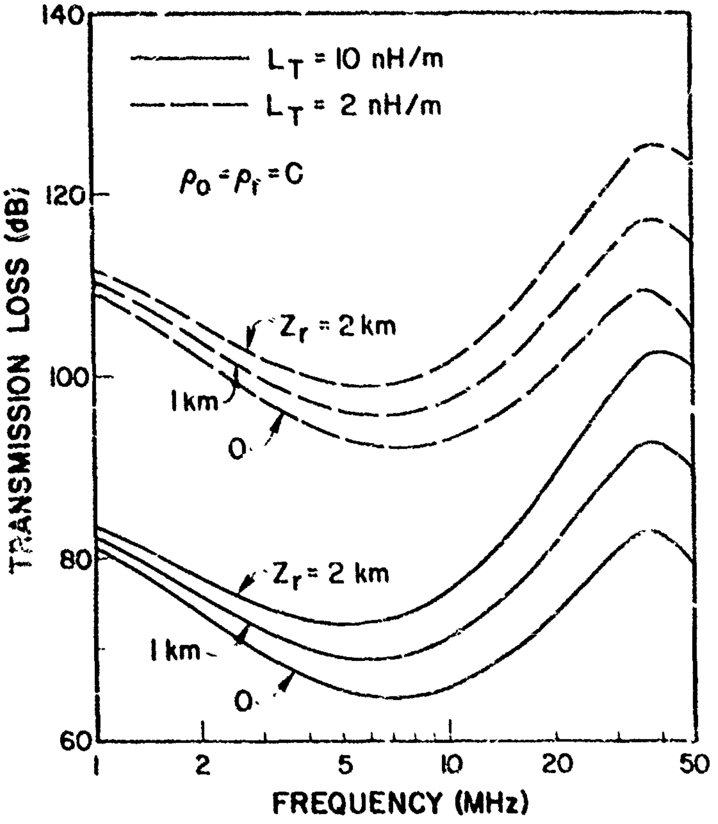


Fig 5 Transmission loss between a pair of identical electric dipoles for cables of lower leakage

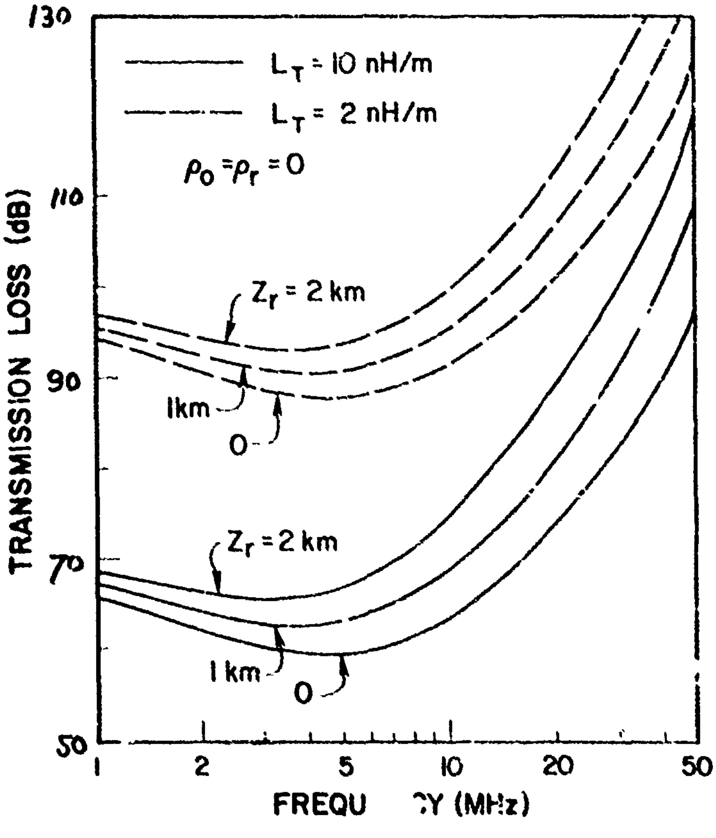


Fig.6 Transmission loss between a pair of identical magnetic dipoles for cables of lower leakage

MODE CONVERSION BY TUNNEL NON-UNIFORMITIES

IN LEAKY FEEDER COMMUNICATION SYSTEMS

David B. Seidel and James R. Wait
Cooperative Institute for Research in Environmental Sciences
University of Colorado
Boulder, Colorado 80309

ABSTRACT

Using an idealized theoretical model, we deal with the inadvertent mode conversion between the bifilar and monofilar modes in a tunnel that contains a braided coaxial cable. The tunnel is allowed to have various kinds of lateral non-uniformities such as changes of wall conductivity and permittivity. We conclude that such effects are very important in understanding leaky feeder systems.

1. INTRODUCTION

The leaky feeder concept has been utilized effectively in intra-mine communications [Delogne 1976, Mahmoud and Wait, 1976]. Previously, it was found that such a system using a braided coaxial cable will propagate two low frequency transmission line modes [Hill and Wait, 1975]. One of these eigen modes (i.e., the monofilar mode) is readily excited from an arbitrary point within the tunnel but suffers high attenuation because the return current flows mainly in the tunnel wall. The other eigen mode (i.e., the bifilar mode) has low attenuation because the return current flows mostly in the sheath; but it is poorly excited for an external source in the tunnel. Thus, for an efficient limited access communication system, it is desirable to convert energy from one mode to the other. Such mode conversion can be obtained through intentional insertion of axial non-uniformities in the coaxial line or within the tunnel environment. Alternately, such conversion may occur inadvertently due to existing nonuniformities along the tunnel's length. In an earlier paper [Seidel and Wait, Sept. 1978), transmission line techniques were used to develop a theory of controlled mode conversion for a braided coaxial cable in a semi-circular tunnel. Specifically, we had considered a cable that contains a short section of higher leakage corresponding to the "leaky stub" converter suggested by Delogne [Delogne, Sept. 1978]. Such results are relevant to mine communication systems, using the continuous leaky feeder concept [Martin, 1971], which rely mainly upon such inadvertent mode conversion for their operation. However, this point of view has not been universally adopted by the leaky-feeder community.

2. FORMULATION

The model assumed is described in terms of a cylindrical coordinate system (ρ,ϕ,z) and is shown in Fig. 1. The tunnel wall is located at $\rho = a_o$ for $0 < \phi < \pi$, and the assumed perfectly conducting tunnel floor is located at $\phi = 0$ and $\phi = \pi$ for $0 < \rho < \infty$. The region defined by $\rho > a_o$ and $0 < \phi < \pi$ is a homogeneous medium with conductivity σ_e and permittivity ε_e. The coaxial cable with outside radius c is centered at $\rho = \rho_o$ and $\phi = \phi_o$. The region defined by $\rho \le a_o$, $0 < \phi < \pi$ and $\rho' > c$, where ρ' is the radial component of a cylindrical coordinate system (ρ',ϕ',z) centered at (ρ_o,ϕ_o), is described by the free space permittivity and permeability ε_o and μ_o, respectively. The geometry of the coaxial cable is shown in Fig. 2. The inner conductor, of radius a, has a high but finite conductivity σ_w. The surrounding insulation of radius b is a lossless dielectric with permittivity ε. The braided sheath located at $\rho' = b$ is characterized by a surface transfer impedance Z_T. The outer dielectric coating has radius c and permittivity ε_c. We also allow for the possibility that a thin lossy film is located at $\rho' = c$ which is characterized by a transfer impedance $Z_L = (2\pi c(\sigma d))^{-1}$ where d is the conductivity-thickness product of the lossy layer. We assume that the fields of each mode of this structure vary as $\exp(-\Gamma z+i\omega t)$ where ω is the angular frequency and Γ is the complex propagation constant for the particular mode.

In a previous paper [Seidel and Wait, July 1978], we obtained a quasistatic mode equation for the low frequency quasi-TEM modes for this structure. There it was found that, in general, there are three distinct modes of this type. The corresponding propagation constants Γ_j were found to be the roots of a cubic equation with complex coefficients. In the special case where the lossy film layer is absent, there are only two modes, whose propagation constants are obtained from a quadratic equation. In each case, two modes can be identified as the well-known monofilar and bifilar modes. The third mode, when it exists, is referred to as the jacket mode [Seidel and Wait, July 1978].

Because the dominant low frequency modes of this system behave like TEM transmission-line modes, we can use transmission-line concepts to consider the effects of axial discontinuities along the extent of the tunnel [Seidel and Wait, Sept. 1978]. This is accomplished by defining characteristic impedances which relate the currents and voltages on the various conductors present. These voltages and currents are then expressed in terms of incident, reflected, and transmitted waves on each side of any discontinuity, and associated coefficients of reflection and transmission between the various modes are obtained by enforcing Kirchoff's voltage and current laws at the discontinuity. It goes without saying that this quasi-static approach is valid only in an approximate sense since higher-order evenescent modes are neglected.

A solution for the reflection and transmission matrices $\overline{\overline{R}}$ and $\overline{\overline{T}}$ was obtained earlier [Seidel and Wait, Sept. 1978] for the case of a tunnel containing a section of length ℓ over which some physical characteristic of the tunnel was different from that in the remainder of the tunnel. Although the method was applied there only to one type of intentional converter, it is also suited to the consideration of many types of inadvertent mode conversions, such as uniform discontinuities of wall permittivity of conductivity. However, before pursuing these extensions, it is worthwhile to somewhat extend the previous quasi-static analysis [Seidel and Wait, Sept., 1978].

It is well known [Collin, 1960] by the reciprocity theorem, that for reciprocal, isotropic media, the reflection and transmission matrices $\overline{\overline{R}}$ and $\overline{\overline{T}}$ defined in [Seidel and Wait, 1978] will be symmetric if the modes are normalized such that the power flow in the various modes are equal. However, in [Seidel and Wait, Sept., 1978], the modes were normalized by the rather arbitrary convention that the current on the coaxial we will denote the non-symmetric reflection and transmission matrices associated with this normalization by $\bar{\bar{R}}'$ and $\bar{\bar{T}}'$.

Now we define a new mode normalization by the convention that $I_s^{(j)} = a_j$ where $I_s^{(j)}$ is the coaxial braid current in the jth mode and $\{a_j\}$ by demanding that the matrices $\bar{\bar{R}}$ and $\bar{\bar{T}}$ associated with this normalization be symmetric. Since [Seidel and Wait, Sept. 1978] provides a solution for $\bar{\bar{R}}'$ and $\bar{\bar{T}}'$, we simply need to relate $\bar{\bar{R}}$, $\bar{\bar{T}}$, and $\{a_j\}$ to these quantities.

It is easy to show that $\bar{\bar{R}}$ and $\bar{\bar{T}}$ are related to $\bar{\bar{R}}'$ and $\bar{\bar{T}}'$ by

$$R_{ij} = (a_j/a_i)R'_{ij} \tag{1}$$

and

$$T_{ij} = (a_j/a_i)T'_{ij}$$

Since we demand $R_{ij} = R_{ji}$ and $T_{ij} = T_{ji}$, (1) implies that

$$(a_j/a_i) = (R'_{ji}/R'_{ij})^{1/2} = (T'_{ji}/T'_{ij})^{1/2} \tag{2}$$

and thus

$$\begin{aligned} R_{ij} &= (R'_{ij}R'_{ji})^{1/2} \\ T_{ij} &= (T'_{ij}T'_{ji})^{1/2} \end{aligned} \tag{3}$$

From (2) we see that only the relative amplitudes of the eigencurrents need be restricted and hence we arbitrarily set $a_1 = 1$. Thus, for a structure propagating N modes, (2) gives us N(N-1) equations for the N - 1 unknowns $\{a_j\}$, j = 2,...,N. Obviously, this system of equations is overdetermined for N > 1. If, in fact, the system has a solution, it is an indication that the modelling process has preserved reciprocity.

3 SOME NUMERICAL RESULTS

In what follows, we will adopt the convention that modes 1, 2, and 3 are the monofilar, bifilar, and jacket modes, respectively. For regions of the tunnel where $z < 0$ (Region 1) or $z > \ell$ (Region 3), the parameters describing the physical properties of the tunnel are the same, and unless specifically indicated otherwise, will have the following typical values: $a_o = 2$m, $\rho_o = 1.8$m, $\phi_o = 45°$, $\epsilon_e = 10\epsilon_o$, $\sigma_e = 10^{-2}$mhos/m, a = 1.5mm, b = 1.0cm, c = 1.15cm, $\epsilon = 2.5\epsilon_o$, $\epsilon_L = 3.0\epsilon_o$, $\sigma_w = 5.7 \times 10$ mhos/m, $\psi = 45°$, $\sigma d = 0$, and $f = \omega/2\pi =$ 36MHz. The coaxial braid is fairly leaky, with a surface transfer inductance $L_T = 40$ nH/m.

In the region described by $0 < z < \ell$ (Region 2), we will let one of the parameters assume a different value than in the remainder of the tunnel; all other parameters are equal to those given above. Data will be given for the monofilar, bifilar, and conversion reflection coefficients (R_{11}, R_{22}, R_{12}) and similarly for the monofilar, bifilar, and conversion transmission coefficients (T_{11}, T_{22}, T_{12}). The situation is depicted in Fig. 4. Data are given in dB, that is

$$X_{ij}(\text{in db}) = 20 \log_{10}|X_{ij}|$$

where X_{ij} denotes any one of the various reflection or transmission coefficients.

First we consider the effect of a change in the tunnel wall's conductivity σ_e in region 2. Figure 4 plots X_{ij} as a function of σ_e for various values of ℓ. It is seen that the reflection coefficients and the conversion transmission coefficient vary in the same manner, exhibiting maximum conversion and reflection when the conductivity is higher in region 2. When σ_e is lower in region 2, these coefficients level off, and are relatively insensitive to further downward changes in σ_e. Of course, at $\sigma_e = 10^{-2}$mhos/m, the regions are the same and we expect R_{11}, R_{22}, R_{12}, and T to vanish. It is also interesting to note that T_{ij} (in db) ≈ $\alpha_j\ell$ where $\alpha_j = 8.686$ Real(Γ_j) is the attenuation rate of the jth mode in region 2 (in dB/m). This means that, essentially, all energy in each mode is transmitted with only small reflection and conversion, and transmission loss is due, primarily, to the modal attenuation rate in region 2.

Figure 5 shows a similar relation of these coefficients to a change in tunnel wall permittivity (ϵ_e). However, not unexpectedly, the maximum conversion and reflection occurs when ϵ_e is smaller in region 2 than in regions 1 and 3. Correspondingly, conversion and reflection are relatively insensitive to upward changes in region 2 permittivity.

Another interesting source of inadvertent mode conversion is due to variations in the lossy film on the cable's outer surface. To consider this, we now let $\sigma d = 10^{-2}$mhos in both regions 1 and 3. Figure 6 shows the effect of varying σd in region 2 upon $\bar{\bar{R}}$ and $\bar{\bar{T}}$. We find that when σd is lower in region 2, there is little mode conversion or reflection, and the coefficients R_{11}, R_{22}, R_{12}, and T_{12} are relatively insensitive to downward changes in σd in region 2. Conversely, when σd in region 2 is increased above that in regions 1 and 3, we see an immediate increase in reflection and conversion. Again, as in previous examples, T_{11} and T_{22} are essentially dependent upon the attenuation rates of the respective modes in region 2.

Although variations in σd along an actual mine tunnel are an unavoidable source of mode conversion, the behavior of T_{12} and R_{12} suggest the interesting idea of intentionally coating periodically spaced segments of cable with a conductive material to be used as mode converters. Indeed, we see that if $\sigma d = 10^{-2}$mhos for the unperturbed system and 5 to 10m lengths of cable are coated to increase this to 1 mho, we obtain a conversion transmission coefficient of about 30 dB. From [Seidel and Wait, Sept. 1978], we found that for a typical leaky-stub (i.e., leaky section) converter (L_T = 40 nH/m), T_{12} was about 20 to 25 dB. Intentional converters of this type would be easy and inexpensive, and might be a quick and simple fix for areas of poor reception in existing continuous access leaky feeder systems.

In Fig. 7, we consider the effect that different cable insulation dielectrics have upon mode conversion for a given discontinuity. Specifically, consider $\sigma_e = 10^{-4}$ mhos/m in region 2 ($\sigma_e = 10^{-2}$ mhos/m in regions 1 and 3) and consider $\bar{\bar{R}}$ and $\bar{\bar{T}}$ for various permittivites of the cable insulation (ε). Note that ε is the same throughout all three regions. For this particular wall conductivity discontinuity, we see that mode conversion is maximized with $\varepsilon \approx 1.25\varepsilon_0$, indicating a fairly high velocity cable. This maximum seems to occur when the phase velocities of the two modes (monofilar and bifilar) are equal.

Thus far, our technique has been applied only to changes in material parameters of the system. But it can also be applied to changes in tunnel geometry. Obviously, because of the resulting abrupt geometrical discontinuities, our confidence in the method is somewhat less: such results are not given here (but they are available from the authors on request).

At this point, it is interesting to consider the consistency of the solutions for the normalization constants $\{a_j\}$ which symmetrized $\bar{\bar{R}}$ and $\bar{\bar{T}}$. Actually, in all cases, $\bar{\bar{R}}$ and $\bar{\bar{T}}$ were obtained by (3). However, in each case, all N(N-1) solutions to (2) were also examined. For the results presented in Figs. 4-7 (that were characterized by changes in material parameters), all solutions for (2) were always in agreement in excess of four decimal places. It is worth mentioning here that this consistency was not found when the geometrical characteristics such as tunnel radius and cable location were modified over the section of length ℓ. However, we do not dwell on this matter here since it does not affect the results presented.

4. CONCLUDING REMARKS

We have demonstrated by a rudimentary theory that inadvertent mode conversion can be a significant aspect of leaky feeder communication systems. In particular, if the phase velocities of the monofilar and bifilar mode are of the same order, relatively small lateral changes in the tunnel characteristics will produce significant energy conversion between the modes. This is probably the dominant mechanism in leaky feeder systems that employ a continuous loosely braided coaxial cable with a foam dielectric insulator. Unfortunately, it is difficult to predict the performance of such systems in a real world environment. On the other hand, controlled mode conversion by means of line converters can use cables with a higher permittivity of the dielectric. Then the inadvertent conversion is of much less consequence.

REFERENCES

COLLIN, R.E., 1960, Field Theory of Guided Waves, New York: McGraw-Hill, p. 80.

DELOGNE, P. April 1976, "Basic mechanisms of tunnel propagation", Radio Sci., Vol. 11, pp. 295-303.

MAHMOUD, S.F., and WAIT, J.R., January 1976, "Calculated channel characteristics of a braided coaxial cable in a mine tunnel", IEEE Trans. Communications, Vol. COM-24, pp. 82-87.

MARTIN, D.J.R., May 1975, "A general study of the leaky-feeder principle", The Radio and Electronic Engineer, Vol. 45, No. 5, pp. 205-214.

SEIDEL, D.B., and WAIT, J.R., July 1978, "Transmission modes in a braided coaxial cable and coupling to a tunnel environment", IEEE Trans. Microwave Theory Tech., Vol. MTT-26, pp. 494-499.

SEIDEL, D.B., and WAIT, J.R., Sept. 1978, "Role of controlled mode conversion in leaky feeder mine-communication systems", IEEE Trans. Antennas Propagat., Vol. AP-26, pp. 690-694.

WAIT, J.R., and HILL, D.A., May 1975, "Propagation along a braided coaxial cable in a circular tunnel", IEEE Microwave Theory Tech., Vol. MTT-23, pp. 401-40[illegible].

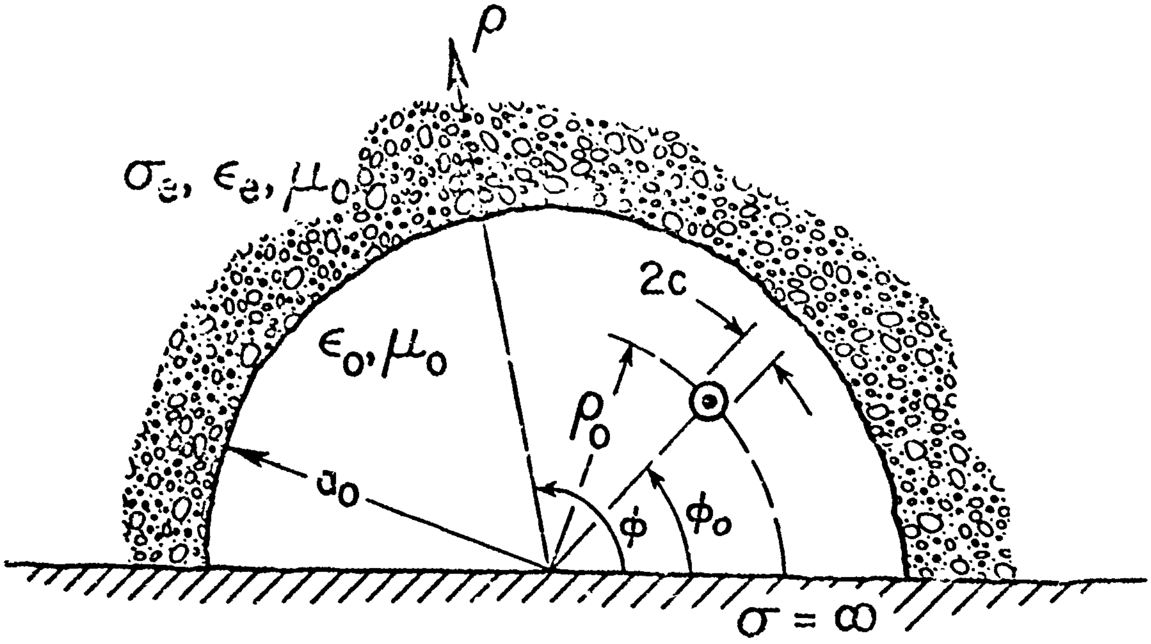

ρ
$\sigma_e, \epsilon_e, \mu_0$
ϵ_0, μ_0
2c
ρ_0
a_0
ϕ
ϕ_0
$\sigma = \infty$

Fig.1 Cross section of semi-circular tunnel containing coaxial cable

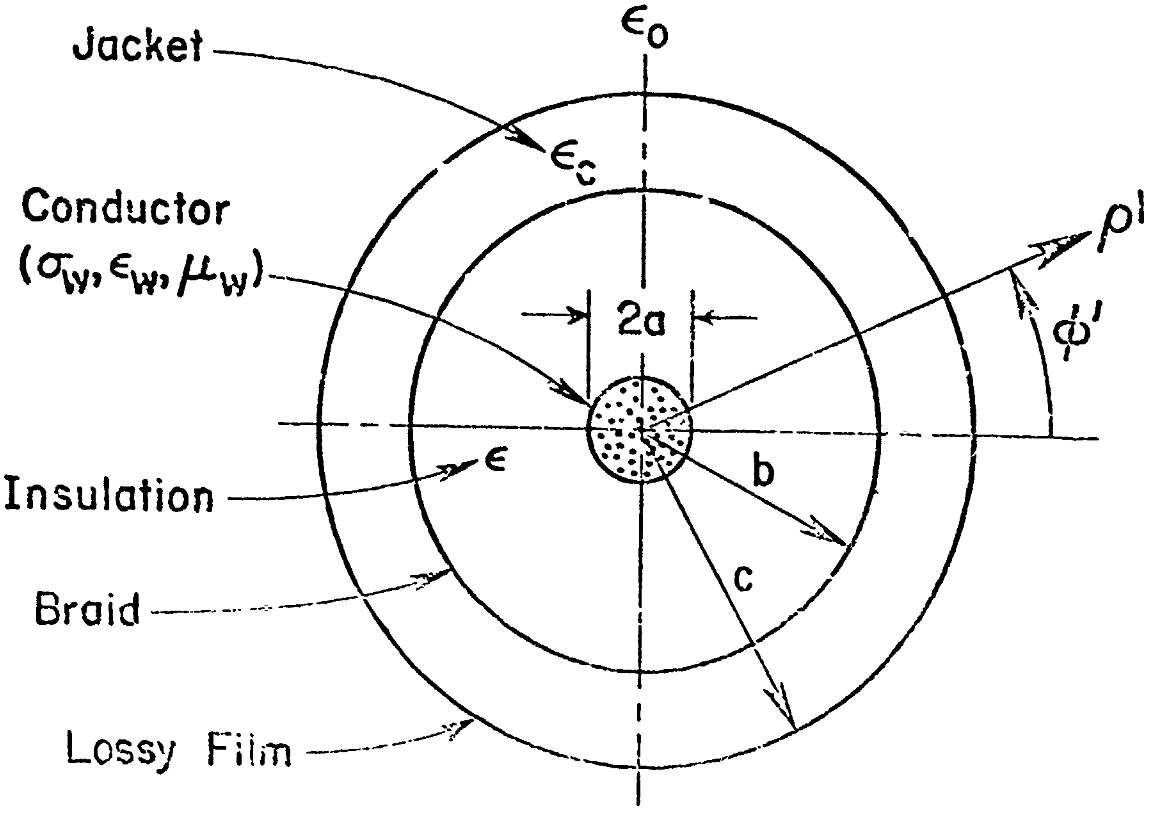

Jacket
ϵ_0
ϵ_c
Conductor
$(\sigma_w, \epsilon_w, \mu_w)$
2a
ρ'
ϕ'
Insulation
ϵ
b
Braid
c
Lossy Film

Fig.2 The braided coaxial cable

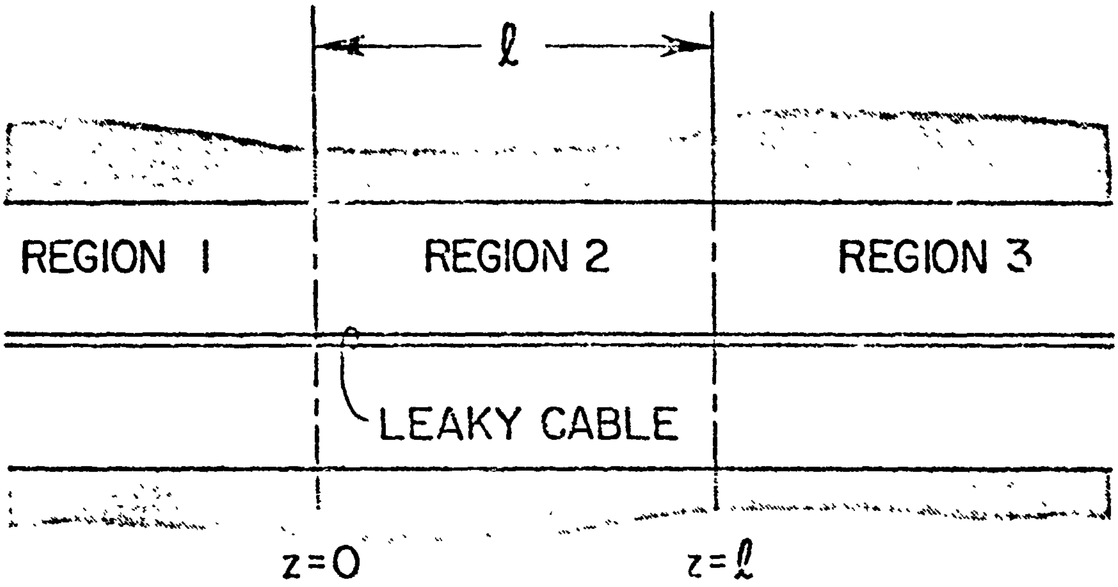

ℓ
REGION 1
REGION 2
REGION 3
LEAKY CABLE
z=0
z=ℓ

Fig.3 Side view of the tunnel showing the three regions

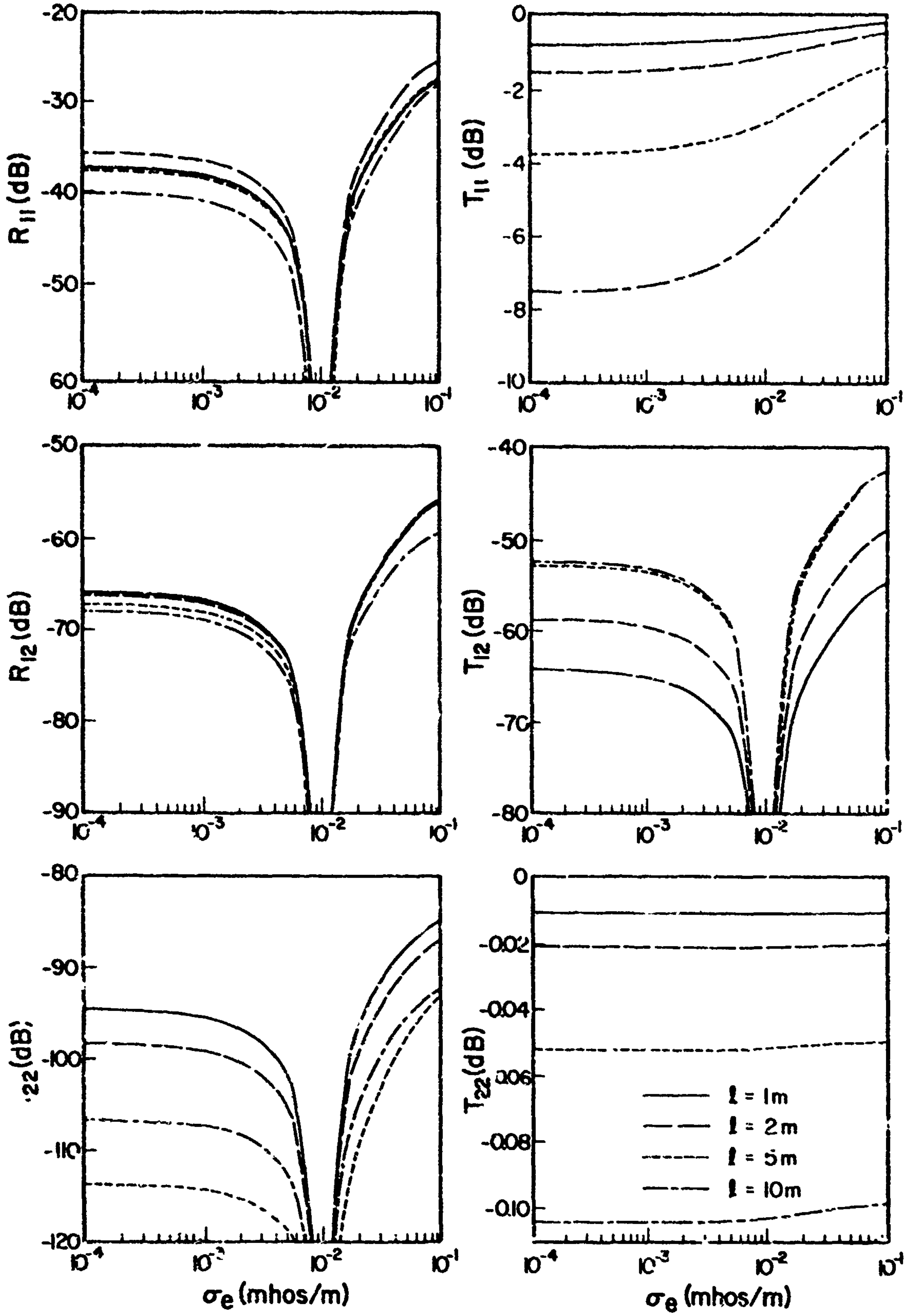


Fig.4 Reflection and transmission coefficients as a function of the rock conductivity σ_e in region 2

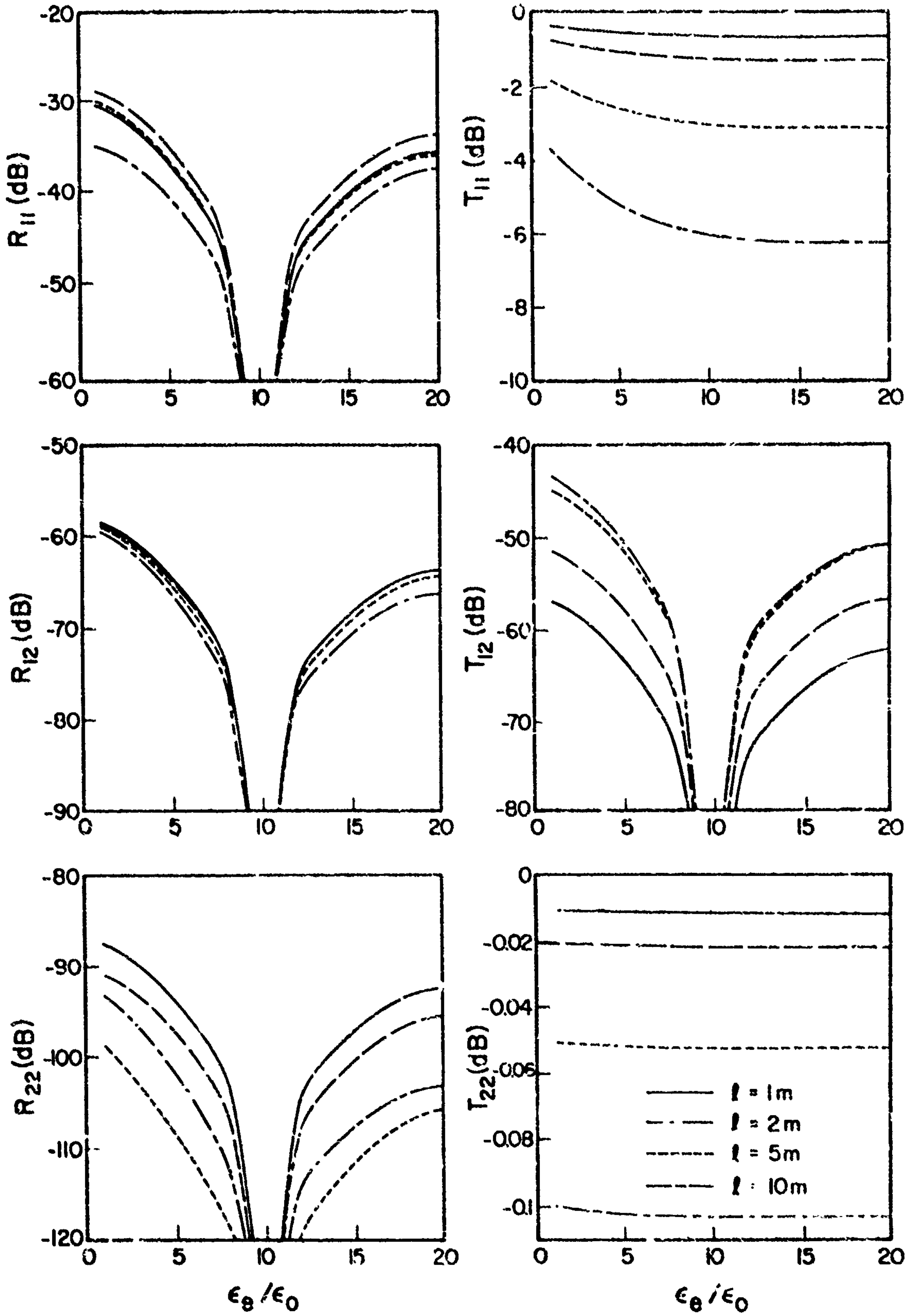


Fig.5 Reflection and transmission coefficients as a function of the wall permittivity ϵ_e in region 2

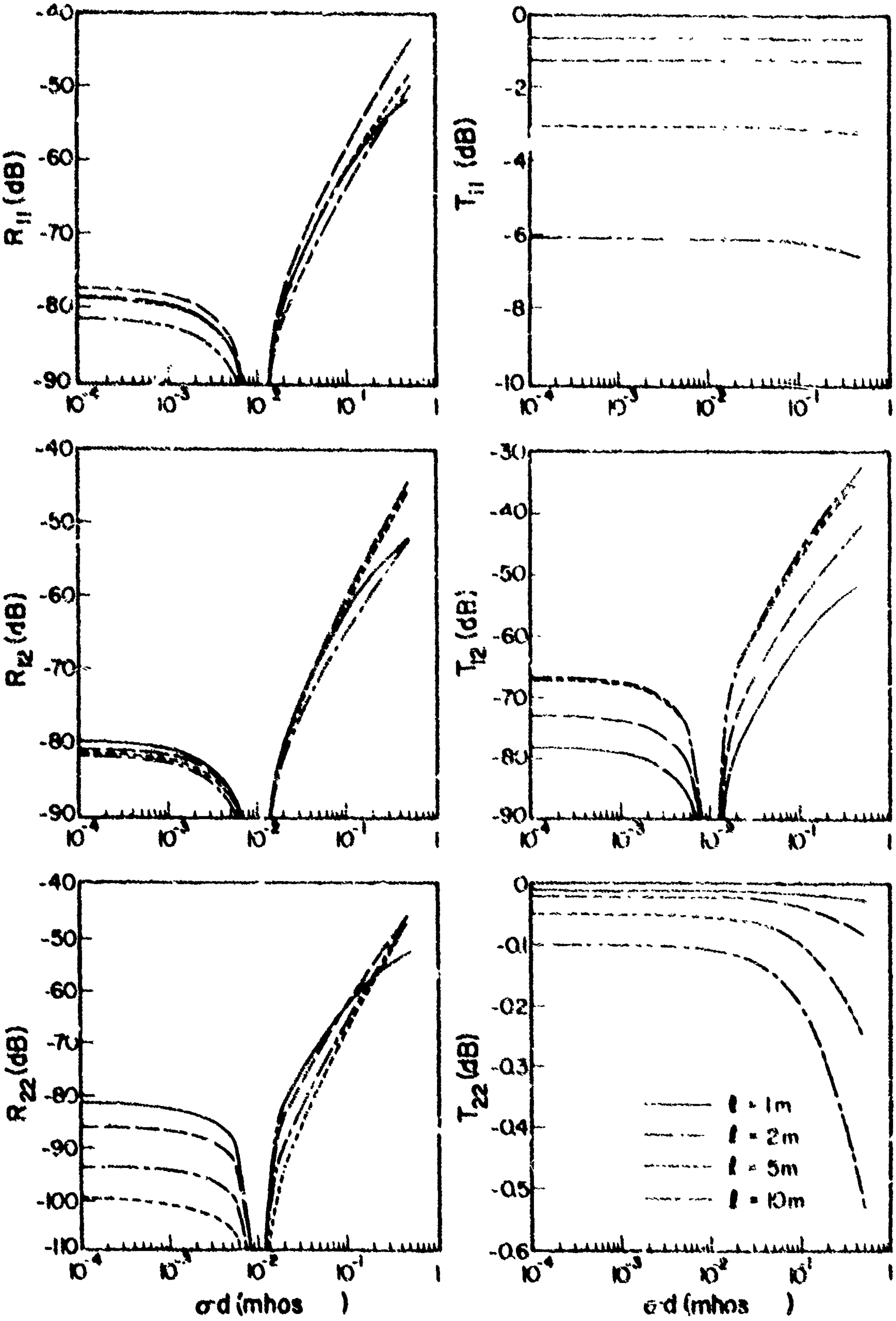


Fig.6 Reflection and transmission coefficients as a function of [illegible] cket conductance σd in region 2

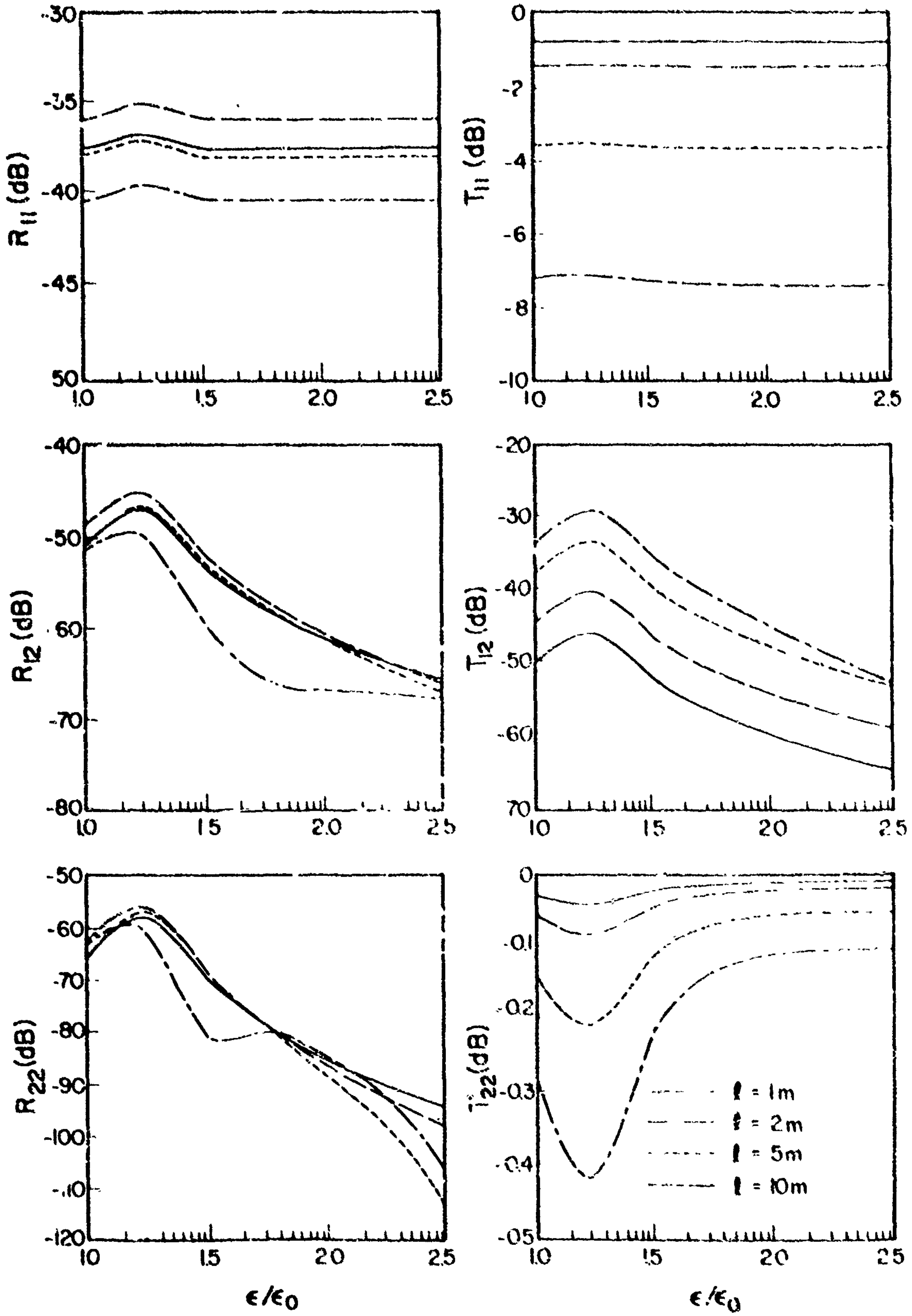


Fig 7 Reflection and transmission coefficients as a function of the insulator permittivity ε of the cable in region [illegible]

SUMMARY OF SESSION VII

EQUIPMENT CONSIDERATIONS

This final session focused on equipment and antennas in connection with HF radio-wave propagation research and application

Tanner (US) described a computerized HF direction finding system with a novel antenna array capable of determining the azimuth and elevation of a radio signal of unknown origin Thus an approximate location can be estimated from a single site (normally this can be achieved only by means of cross bearing). Other advantages of this system are the high speed of measurements being performed and the minimal requirements for trained technical personnel

Bibl (US) explained the various measurements which can be performed simultaneously with the new digital ionosonde, the Digisonde 128 PS. The measurements include amplitude, phase, angle of arrival, polarization, frequency shift and signal travel time of a radio wave. For future applications a topside digital ionosonde with low output data rate is being developed

Belrose (Ca) described an HF communication set comprising a modern single-sideband transceiver with voice compression and quickly erectable antennas optimized for low heights. The system is intended to be used from remote sites in northern Canada, where neither landline nor VHF radio is available

Sweeting's (UK) paper was read by **D.J. Gale**. It is concerned with a computer-based method for optimizing the radial system of relatively short vertical whip antennas. Some users (mobile radio) are bound to use whip antennas, where the efficiency mainly depends on the radial system Radials of different lengths reduce resonance effects which otherwise appear in highly variable impedances.

Stark (Ge) introduced a method for optimizing the radiation pattern of HF antennas. For different ranges to be bridged by HF radio, the angle of incidence is different and also the frequency used usually is different! A great number of radio propagation predictions of the Deutsche Bundespost is used to find the angle distribution as a function of the distance. These plots were then used to design logarithmic-periodic antennas where the vertical radiation patterns change with frequency so that for each distance the radiation is optimum

A MOBILE HF IMPULSE SOURCE LOCATOR

by

M.L.HERON
Physics Department,
James Cook University,
Townsville, 4811
Australia.

SUMMARY.

The design strategy is described for an instrument which will record the direction of arrival of radio waves from noise pulses generated during discharges of static electricity in the atmosphere. The system can be adapted to LF which would be more useful for thunderstorm location and tracking, but the advantage of the HF instrument is that it can be mounted on a mobile platform, for example an aircraft, and taken near to and into storms to study storm system infrastructure. The problem of extracting the directionality from otherwise wide band noise is solved electronically by using coherency in a narrow frequency band. A simple data logging system is described. The present system, configured for an airborne platform has azimuthal resolution of about 0.2° and elevation resolution of about 5°.

1. INTRODUCTION.

For many years it has been a standard technique to locate and tract meteorological disturbances by measuring time delays of L.F. noise pulses or 'spherics on long base line triangulation networks. These observations have, in the past, generally been costly to maintain and operate and recently, to the detriment of operational forecasting, there has been a trend towards discontinuing the procedure. As well as this forecasting aspect, the relationship between static electricity in cloud and vortex systems and the other meteorological parameters is not well understood. There is a need, therefore, for two types of noise pulse locating devices. The first is a long range device which will replace the long baseline 'spherics system to allow storm location and tracking, but on a reduced cost basis. The second requirement is a device for short range high resolution location of active cells within storm systems.

This paper describes a noise pulse locator designed to meet the needs of research into the structure of tropical cyclone (hurricane) systems but which may be easily adapted to storm location and tracking applications. Tropical cyclones do not easily submit to detailed observation because of their infrequent, but devastating visits to a given location and unpredictable paths. Coastal rain radars are adequate for path tracking and severity estimates and coupled with satellite-borne infrared cloud observations form an indispensable real time information base for coastal communities. Limits to resolution of these and other ground based techniques at remote sites render them of little use for observations of the detailed internal structure of cyclones. Researchers wanting observations with high spatial resolution in tropical cyclones have been faced with the problem of maintaining arrays of closely spaced coastal or island sites and simply waiting for a cyclone to pass. The comparatively recent development of aircraft instrumentation for meteorological research is removing these observational frustrations by taking the observation site to the storm rather than waiting for the storm to come to the site. This capability is opening the way for detailed study of active cells within tropical cyclone and other storm systems.

This paper is directed towards:

(i) extracting from the extent literature a set of properties of lightning noise at HF upon which the instrumentation is based,

(ii) describing the principle and instrumentation of the noise pulse locator,

(iii) estimating resolution, errors and limitations of the equipment as deployed on an Orion aircraft,

(iv) noting the areas of application and adaption of the instrument.

2. ELECTROSTATIC DISCHARGE PROCESSES.

Most observations of lightning phenomena over many decades have been made of cloud to ground flashes. The accumulated literature gives a fairly consistent description of the observed phenomena during a cloud to ground flash but the characteristics of cloud to cloud, intracloud, and cloud to air discharges do not emerge so clearly. Most recorded observations are on thunderstorms but the few observations on other systems like dust devils and tornadoes provide useful contrasts on which to base empirical theories for the purpose of instrumenting the tropical cyclone experiment. Theoretical bases for the electrical processes are not well established.

2.1 Cloud to Ground

The classical description of a cloud-to-ground flash has a stepped leader lasting for about 20 ms which radiates at HF and creates an ionized channel between the negatively charged cloud base and the positively charged ground. (This ignores the small p region of positive charge at the cloud base.) (Uman 1969). A high current return stroke lasting a few milliseconds does not radiate at HF but is the most luminous feature. At 30-50 m/sec intervals a fast dart leader (~2 m sec) followed by another return stroke usually occurs for a total duration of up to 0.5 sec. During the interstroke period smaller scale discharge processes called K-streamer and J-streamer processes make charge available to the stroke channel by extending conduction paths from the top of the channel further into the negatively charged region of the cloud. These processes radiate at HF.

The radiation at HF begins at the commencement of the first stepped leader as a general enhancement superposed on short spikes of the order of microseconds probably originating from each step. (figure 1) A few milliseconds of no signal at 11 MHz accompanies the return stroke, followed by noise from the J and K processes which has a similar spikey appearance to that from the stepped leader. (Horner 1964)

2.2 Cloud to Cloud

The noise at HF from cloud to cloud discharges is similar to that from the stepped leader and interstroke period of a cloud to ground flash. The presence or absence of a quiet time provides identification of cloud to ground and cloud to cloud flashes (Kreielsheimer 1965). For thunderstorms, Brook and Kitagawa (1960) report a fairly consistent ratio of three cloud to cloud flashes for every one cloud to ground flash and discuss charge continuity under these conditions. This ratio varies from 2.5 to 1 in temperate zones to 10 to 1 in the tropics (Pierce 1955, 1956, and Prentice 1960). The discussion by Brook and Kitagawa (1960) probably does not apply to large cyclone systems but may apply to sub-cells.

2.3 Noise Pulse Source Location

High frequency radiation does not originate from the full channel length of the high current return stroke but from smaller scale discharges. The stepped leader of a cloud to ground flash emits HF and VHF noise as a series of pulses typically 1-2 µsec in duration and separated by about 50 µsec. Initial discharge in the cloud preceding the stepped leader and junction processes (J and K) in the cloud between strokes give off HF radiation. Hewitt (1957), using 50 cm radar observations found that the inter-stroke streamer activity was normally most intense and prolonged in the lower parts of a thundercloud (i.e., 4-7 km altitude) but occasionally develop to around 10 km. Krehbiel et al (1976) used a ground array of stations to measure fast changes in electric field to locate charge centres in a thundercloud and found a horizontal displacement between ground strokes.

The experiment as planned will not have the time resolution to locate the radiation sources of individual 1-2 µsec pulses from each filament step but will average over a longer period of about 200 µsec. At the standard data storage rate of 10 per second this will not follow the progression of radiation centres throughout a flash. The instrumentation will allow switch selection of higher data rates (100 per second) for later inclusion of these fast observations.

The basic data rate will locate the discharge centres for comparison with precipitation radar patterns. Krehbiel et al (1976) observed a close association between the location of charge centres and shafts of precipitation located by radar for one thunderstorm in New Mexico. The cause and effect relationship between precipitation and electrification is not understood but the two are often observed to be linked (Vonnegut et al 1959). In addition to the link with precipitation, work by Reynolds and Brook (1956) has demonstrated a direct relationship between vertical convection and electric fields in isolated thunderstorms and Brook and Kitagawa (1960) suggest that for isolated thunderstorms the mean duration of lightning discharges (the total flash) is a measure of the extent and intensity of vertical convection. The association with convection is in general agreement with the results of Taylor (1975) who found greater radio noise activity in windstorms and funnel clouds than in thunderstorms.

It must be stressed that the radiation centre at HF will not necessarily be the luminosity centre or the point of maximum current or of maximum atmospheric heating. Rather it is the centre of initial breakdown. Malan (1963) suggests a filament breakdown mechanism between large drops ($\geq$ 2mm diameter) low down in a thundercloud when E $\geq$ 100 kv/m. Higher up, the droplets and ice particles polarise, and breakdown will occur when E $\geq$ 300 kv/m (Chalmers 1967) and the breakdown is most likely to begin where the droplet spacing is smallest. It is these breakdown regions which will be located by the radio direction finder. Ambiguities and errors due to extended sources are discussed later under system evaluation.

2.4 Discharges in Tropical Cyclone Systems.

No reports of the nature of electrical discharges observations can be found which have sufficient resolution to add any understanding to the internal structure of a tropical cyclone or hurricane. Ward (1977) used the 2[illegible] MHz radar facility at Townsville, North Queensland to examine the time sequence of radio noise from cyclones in that region and reported a cycle of noise burst activity going from a minimum in burst rate through a peak and back to a null in about [illegible]0 minutes. This cycle of activity needs to be confirmed with more observations and compared with other results. This period is about the same as the time for a thunderstorm to develop and may point to substorm cells within the cyclone structure. However it appears likely that the lightning in cyclonic conditions will not have the ratio of cloud to cloud to cloud-to-ground flashes typical of thunderstorms. The results of Ward (1976, 1977) and the burst rate data taken in this present experiment will need to be compared with results for other storm systems. Taylor (1975 and 1973) found that thunderstorms produce the lowest burst rate, with hail, wind and funnel clouds producing respectively increasing burst rates. Taylor's results for tornado conditions show high burst rates with average values in excess of 20 bursts per minute. Brook and Kitagawa (1960) point out that the interval between flashes depends upon the number of cells in a thunderstorm system with each cell discharging about every 30 seconds. This explanation may be relevant to tropical cyclones if they have sub-cells.

3. INSTRUMENTATION

3.1 Principle of Operation

The low power pulsed emissions from streamer processes during the reorganisation of electrostatic charge inside a cloud appear as wide band noise at the antenna of a radio receiver. Apart from its sporadic occurrence and pulsed structure in the time domain, the distinguishing feature separating it from normal front-end noise is that it originates from a point source discharge. This means that the radio waves at various frequencies arriving at the antenna have ordered phase relationships. A narrow band receiver responds to wide band noise by following the phase of the wave whose frequency is that of the centre of the receiver band. Conceptually, the radio emission needs to be described as a packet of energy inside which, in the time domain, there is a fine structure of pulses. The Fourier Transform of the packet is then definable and for a narrow frequency interval we can discuss the amplitude and the phase.

In order to observe the phase relationships of the point source of noise pulses, an interferometer is required which will operate on the radio waves during one of the packets. A three antenna interferometer is the bare minimum for three-dimensional direction finding. The success of the system rests upon the ability to make relative phase measurements between waves of the same frequency at the three different antennas, during the short-lived packet of energy.

3.2 Antennae

The antenna locations are chosen on the aircraft for operational convenience under the restriction that they be not linearly arranged (figure 2). On an alternative ground based platform a fourth antenna in an elevated position would be of considerable advantage for error reduction on elevation as well as azimuth measurement. The NOAA WP-3D aircraft equipped for meteorological sensing ([illegible] 1976) is the basis of figure 2 and suitable strengthened locations for vertical whip antennas one metre long are shown.

3.3 Receivers

The three channel HF receiver shown in block form in figure 3 maintains phase relativity across the channels. The frequency [illegible] MHz has been chosen because the corresponding wavelength of [illegible] m is appropriate for the aircraft borne antenna layout. Also it is a sufficiently high frequency to allow discrimination between cloud-to-cloud and cloud-to-ground discharges. A higher frequency would have less radiant energy and a lower frequency system (LF) would respond more to cloud-to-ground strokes.

The 455 kHz intermediate stages have precision matched ceramic filters to ensure that the pass bands seen by the phase lock loops are identical. For a narrow band system the phase lock loops may be considered to be following the centre frequency of a noise band. The difference between centre frequencies must have a reciprocal which is very much greater than the time required for the phase lock loops to acquire locking conditions.

The phase comparators operate on fixed amplitude square waves at 455 kHz and are 0° to 360° detectors. The outputs have single pole time constants of 200 µs and the sample-and-hold circuits operate nominally 200 µs after the amplitude threshold is crossed.

The amplitude threshold level is externally adjusted to overcome unknown background noise levels problems. The threshold level is set using a small monitor oscilloscope. The amplitude modulation of one of the channels is detected and the output ([illegible]) is recorded directly in analogue form on one channel of a magnetic tape. This tape is for performance monitoring and occurrence statistics. When the threshold circuit is triggered, a marker pulse is set on a second channel of the tape and the sample-and-hold circuits are triggered. Lock-out circuits prevent further threshold triggering for [illegible] ms under normal operation and [illegible] ms for the fast data rate, after which the threshold circuit will either immediately trigger again or go into a wait state in quiet conditions.

An eight bit binary counter is incremented each time the sample-and-holds have settled down. This is an event counter indicating the lightning activity level but more importantly it serves as a status register for asynchronously passing data to the data acquisition system.

3.4 Data Acquisition

Three analogue to digital (A/D) channels are required for the three data lines and a fourth input being an eight bit digital register serves as a status register for the data transfer. After each increment of the status word the three analogue channels are ready to be digitised and stored. Since the data is intermittent the time of each triplet of data has to be stored. If the normal maximum data rate of 10^1 per second, the status word has to be interrogated at least 10, and preferably [illegible] times per second, while at the fast rate (max 10^2 per sec) the interrogations should be greater than 100 per second and preferably 200 per second to avoid losing points in the asynchronous data transfer.

The essential software of data logging is outlined in figure 4 to be implemented on a [illegible] microprocessor recording data onto magnetic tape. This system uses four A/D channels and tests the lowest bit of the status word for a change of state when data is ready to be read out of the sample-and-hold circuits.

4 ANALYSIS

The trigonometry required to derive azimuth and elevation from the relative phases of the signals at the three antennas is outlined in this section. Errors are estimated and some limitations of this experiment are discussed.

4.1 Theory

Figure 5 shows the coordinate system based on the location of the three antennas. θ is the azimuth and ϕ the elevation. The symbol $\phi 21$ is used to represent the observable phase of antenna 2 with respect to the phase of antenna 1. In this notation

$$\phi 21 = (4\pi A/\lambda)\cos\phi\sin\theta$$
$$\phi 32 = (2\pi D/\lambda)\cos\phi\cos\theta - (2\pi A/\lambda)\cos\phi\sin\theta$$
$$\phi 31 = (2\pi D/\lambda)\cos\phi\cos\theta + (2\pi A/\lambda)\cos\phi\sin\theta$$

There is redundancy in the set of equations but some solutions for θ and ϕ are ill-conditioned under certain conditions. The solution for θ using

$$\tan\theta = \frac{D\phi 21}{A(\phi 32 + \phi 31)}$$

is satisfactory for all θ and gives θ to $\pm$ 0.[illegible] degrees for errors of $\pm 1^\circ$ in the phase measurements. The quadrant information for θ is obtained by adding 180° if $(\phi 32 + \phi 31)$ is negative which gives $-90^\circ < \theta < 270^\circ$.

The best solution for ϕ depends on the value of θ and the following scheme is used

$$\cos\phi = \lambda\phi 21/4\pi A\sin\theta$$
$$\text{for } |\sin\theta| \geq 0.707$$
$$\cos\phi = \lambda\phi 32/2\pi(D\cos\theta - A\sin\theta)$$
$$\text{for } -0.707 < \sin\theta < 0$$

and

$$\cos\phi = \lambda\phi 31/2\pi(D\cos\theta + A\sin\theta)$$
$$\text{for } 0 < \sin\theta < 0.707$$

For observational phase errors of $\pm 1^\circ$, ϕ will be evaluated to better than 5°. The observations do not allow resolution of the sign of ϕ.

4.2 Limitations

One serious limitation is the poor experimental resolution of elevation angle particularly when it is less than 10°. This arises through the geometry of the antennas and could be avoided by running a duplicate experiment on the same antennas but at a higher frequency (~300 MHz).

The problems of interference between point sources in different locations and of extended sources, like a long streamer, are more fundamental. If several points radiate during the 300 µs required for each observation then the composite radiation is received and processed. It is shown here that for two sources of equal radiant power, the instrument will record the mean phase and hence the mean direction in space. If however, as is more likely, one point is radiating more power that others then the system will favour the strongest source.

Consider two waves with amplitudes A and B at the same frequency but with a relative phase shift ψ. The antenna will sense a voltage V given by

$$V = A\sin\omega t + B\sin(\omega t + \psi)$$

and this signal will drive the phase lock loop. This can be written

$$V = \{(A+B\cos\psi)^2 + (B\sin\psi)^2\}^{\frac{1}{2}} \sin(\omega t + \alpha)$$

when α is the phase of the sine wave driving the phase circuits and is given by

$$\cos\alpha = (A+B\cos\psi)/\{(A+B\cos\psi)^2 + (B\sin\psi)^2\}^{\frac{1}{2}} .$$

For A = B, $\alpha = \psi/2$ and the phase taken by the loop is the mean of the two radiating waves. When A > B, α becomes less than $\psi/2$ as shown in figure 6. The net result is that the measured direction is weighted towards the stronger source.

This inability to detect or resolve simultaneous sources may cause interpretational difficulties. These problems if they prove to be serious, my be avoided by taking the observations over much shorter time intervals of ~1 µs. The short duration peaks of radiation from individual streamer activity are less likely to coincide with the same fast burst originating elsewhere.

5. CONCLUSION

A review of lightning physics and cloud electrification has been used as the basis of the design of a mobile HF noise pulse locating instrument. The three antenna version configured for the WP 3D(Orion) aircraft has some performance limitations which are intrinsic in a three element interferometer operating in a three dimensional situation. For other applications a four antenna mounted out of the plane of the other three would be sought. In this work the general direction has been towards the airborne application because of the poorly documented state of electrical discharge phenomena in tropical cyclone systems. However an immediate alternative application is in the location and tracking of thunderstorms from fixed stations using lower frequencies (eg 10kHz) and cloud-to-ground energy releases. While this may be seen as a useful updating of the long baseline 'spherics system (LF) it should be noted that two fixed stations are necessary in each region. Undoubtedly the major application of the noise pulse locator described here is to mobile platforms close to, and inside, storm systems.

REFERENCES

BROOK, M. and N. KITAGAWA, 1964, "Some aspects of lightning activity and related meteorological conditions" J. Geophys. Res. 69, [illegible].

CHALMERS, J.A., 1947, "The capture of ions by ice particles" Quart. J. Roy. Meteorol. Soc. 73, [illegible].

HEWITT, F.J., 1957, "Radarechoes from inter-stroke processes in lightning" Proc. Phys. Soc. 70, 961-979.

HORNER, F., 1964, "Radio noise from thunderstorms" in 'Advances in Radio Research' Vol. 2 Ed. J.A. Saxton, Academic Press, London, pp 121-204.

KREHBIEL, P., M. BROOK, R. McCRORY, and [illegible], 1976, "Lightning charge center locations relative to precipitation in a thunderstorm" International Conference on Cloud Physics, Colorado, American Meteorological Society.

KENTISCHEIMER, K., 1965, Contribution to "Problems of Atmospheric and Space Electricity" Ed. S.C. Coroniti, Elsevier, Amsterdam, p. 88.

MALAN, D.J., 1965, Contribution to "Problems of Atmospheric and Space Electricity" Ed. S.C. Coroniti, Elsevier, Amsterdam, p. [illegible].

PHILLIPS, BYRON B., 1976, "The NOAA aircraft observation system" International Conference on Cloud Physics, Boulder, Colorado, American Meteorological Society.

PIERCE, E.T., 1955, "Electrostatic field-changes due to lightning discharges" Quart. J. Roy. Met. Soc. 81, 211-228.

PIERCE, E.T., 1956, "The influence of individual variations in the field changes due to lightning discharges upon the design and performance of lightning flash counters" Arch. Met. Geo. 9, [illegible].

PRENTICE, S.A., 1960, "Thunderstorms in the Brisbane area" J. Inst. Elect. Engrs (Australia) 32, [illegible].

REYNOLDS, S.E. and M. BROOK, 1956, "Correlation of the initial electric field and the radar echo in thunderstorms" J. Meteorol. 13, 376-380.

TAYLOR, W.L. 1973, "Electromagnetic radiation from severe storms in Oklahoma during April 29-30, 1970" J. Geophys. Res., 78, 8761-8777.

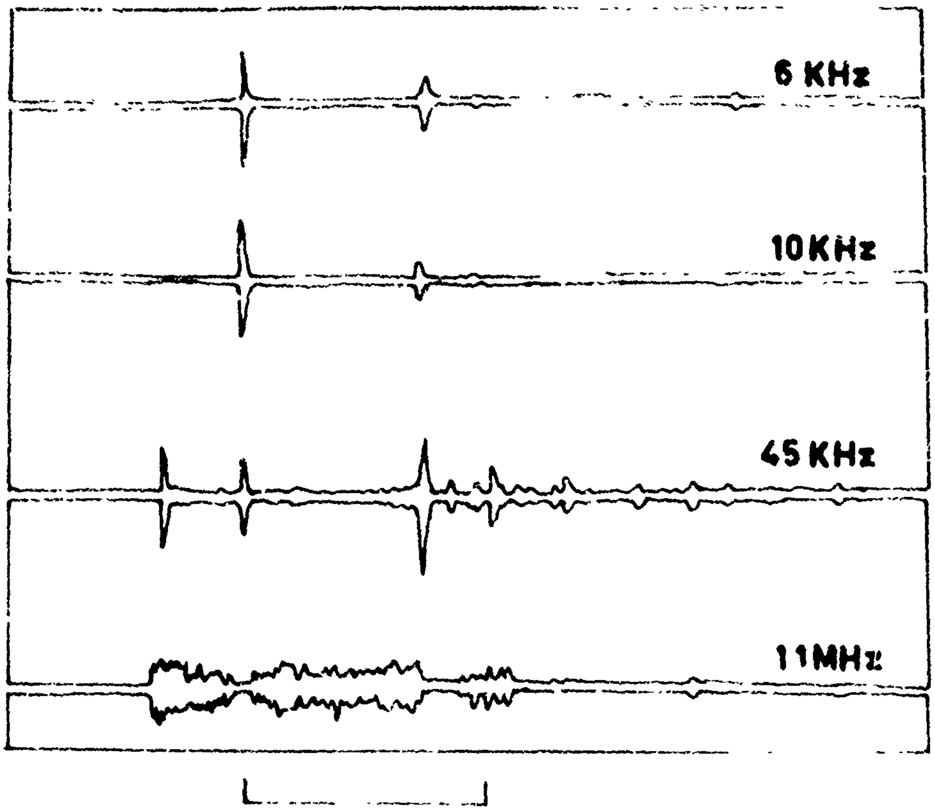
TAYLOR, W.L., 1975, "Detecting tornadic storms by the burst rate nature of electromagnetic signals they produce" 9th Conference on Severe Local Storms, Norman, Oklahoma 1975, American Meteorological Society.

UMAN, Martin A., 1969, "Lightning", McGraw Hill, New York.

VONNEGUT, B., C.B. MOORE and A. BOTKA, 1959, "Preliminary results of an experiment to determine initial precedence of organized electrification and precipitation in thunderstorms" J. Geophys. Res. 64, [illegible].

WARD, J.F., 1976, "The electrical characteristics and possible Modification of Tropical Cyclones" [illegible] International Symposium on Equatorial Aeronomy, Townsville, Australia, 1976.

WARD, J.F., 1977, "Electrical and Magnetic phenomena of Tropical Cyclones" Natural Philosophy Report No. [illegible] James Cook University.



(after HORNER, 1964)

Figure 1. Radiation from a ground flash. The return strokes show up clearly on 6 kHz and the characteristic ground stroke blanking is seen on 11 MHz. The inter-stroke radiation on 11 MHz here is similar to radiation from cloud-to-cloud flashes.

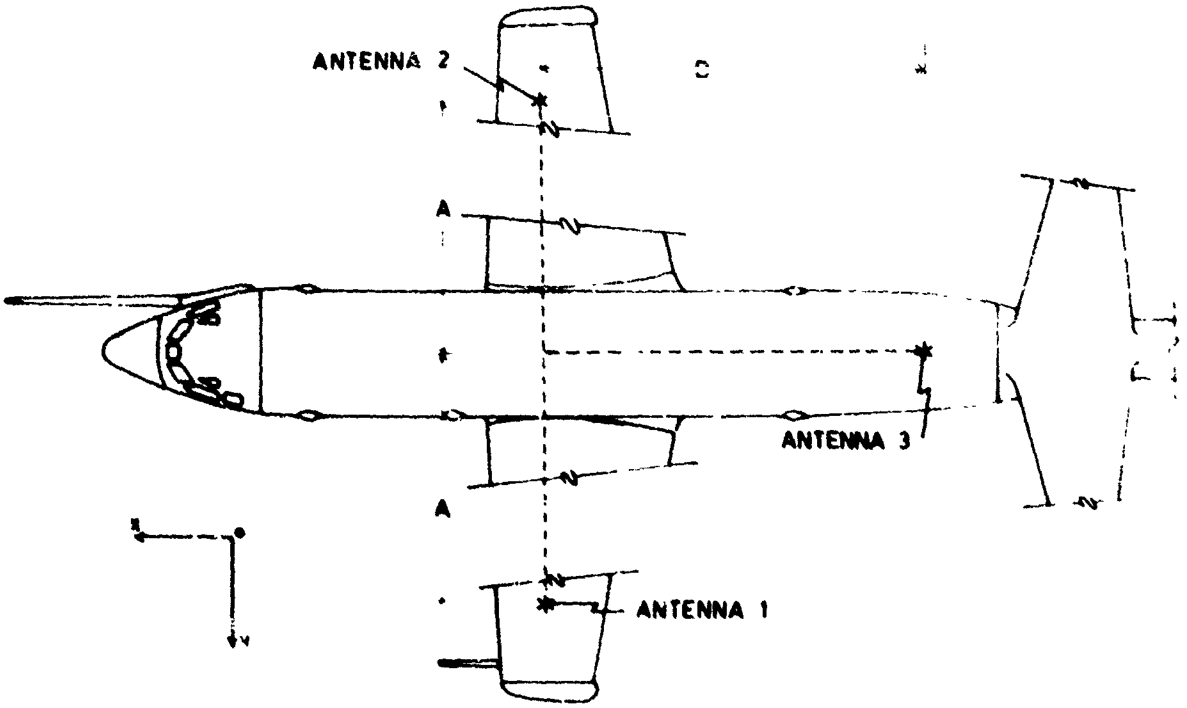


Figure 2. Antenna locations on the underside of the Lockheed WP-3D aircraft. 1 and 2 are on the wing sensor pylons and 3 is on the rear fuselage. The antennas are downward pointing one metre stubs.

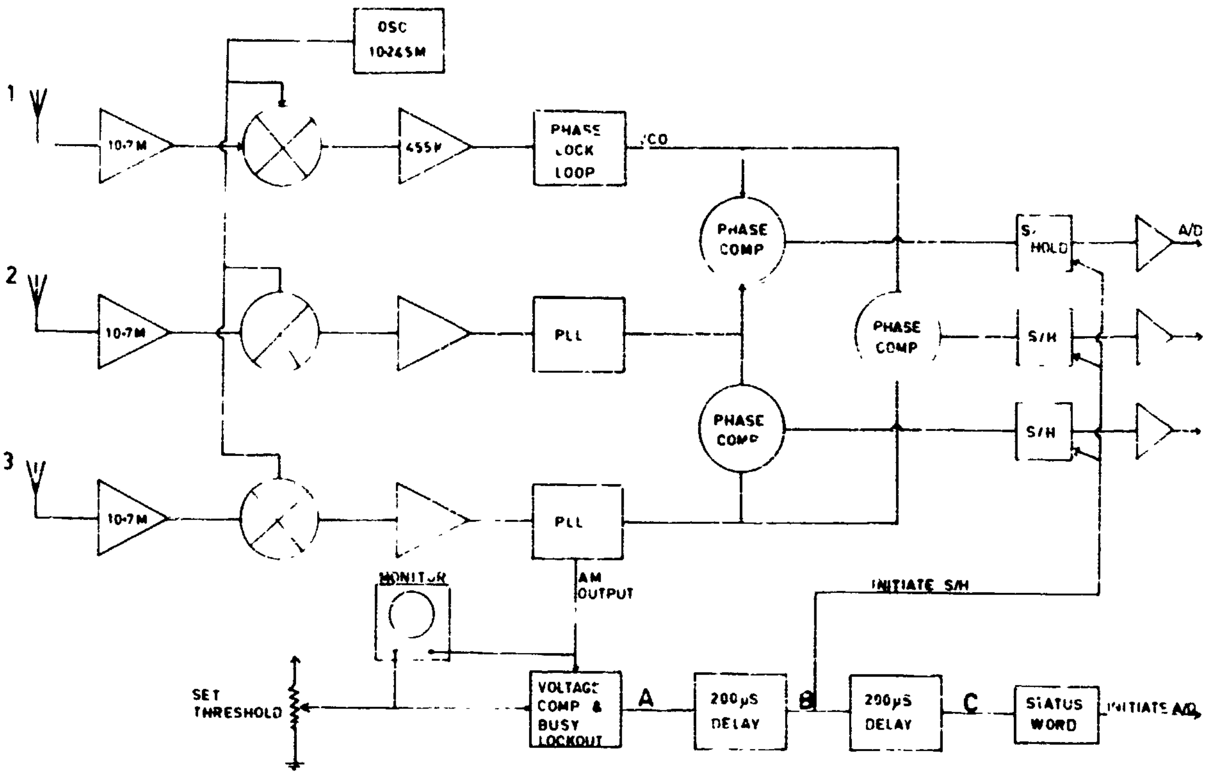

OSC
10.245M
1
2
3
10.7M
455K
PHASE LOCK LOOP
VCO
PLL
PHASE COMP
S/HOLD
S/H
A/D
MONITOR
AM OUTPUT
INITIATE S/H
SET THRESHOLD
VOLTAGE COMP & BUSY LOCKOUT
A
200μS DELAY
B
C
STATUS WORD
INITIATE A/D

Figure 3a. The block diagram for the hardware. The timing of the threshold detector and subsequent triggers is shown in figure 3b.

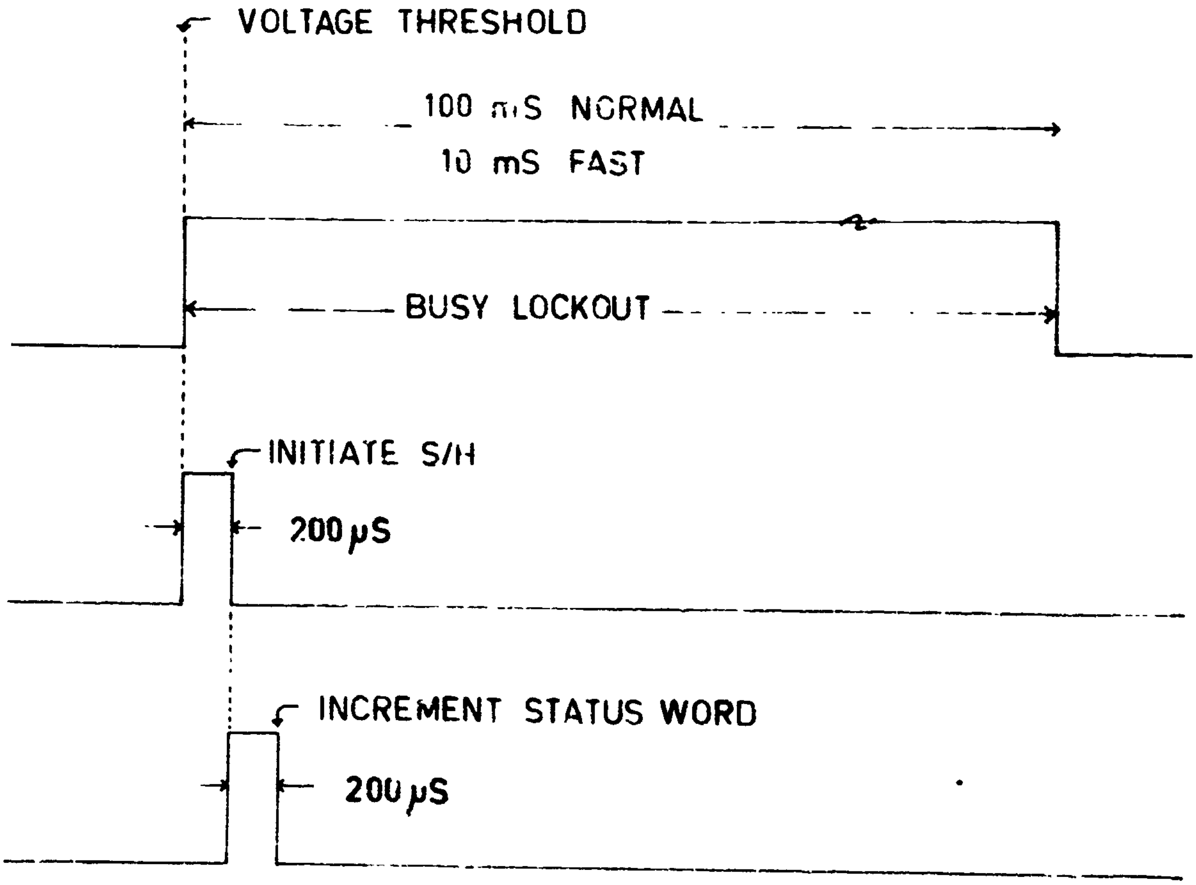

VOLTAGE THRESHOLD
100 mS NORMAL
10 mS FAST
BUSY LOCKOUT
INITIATE S/H
200μS
INCREMENT STATUS WORD
200μS

Figure 3b. The timing chain after the radiation amplitude rises above the threshold level.

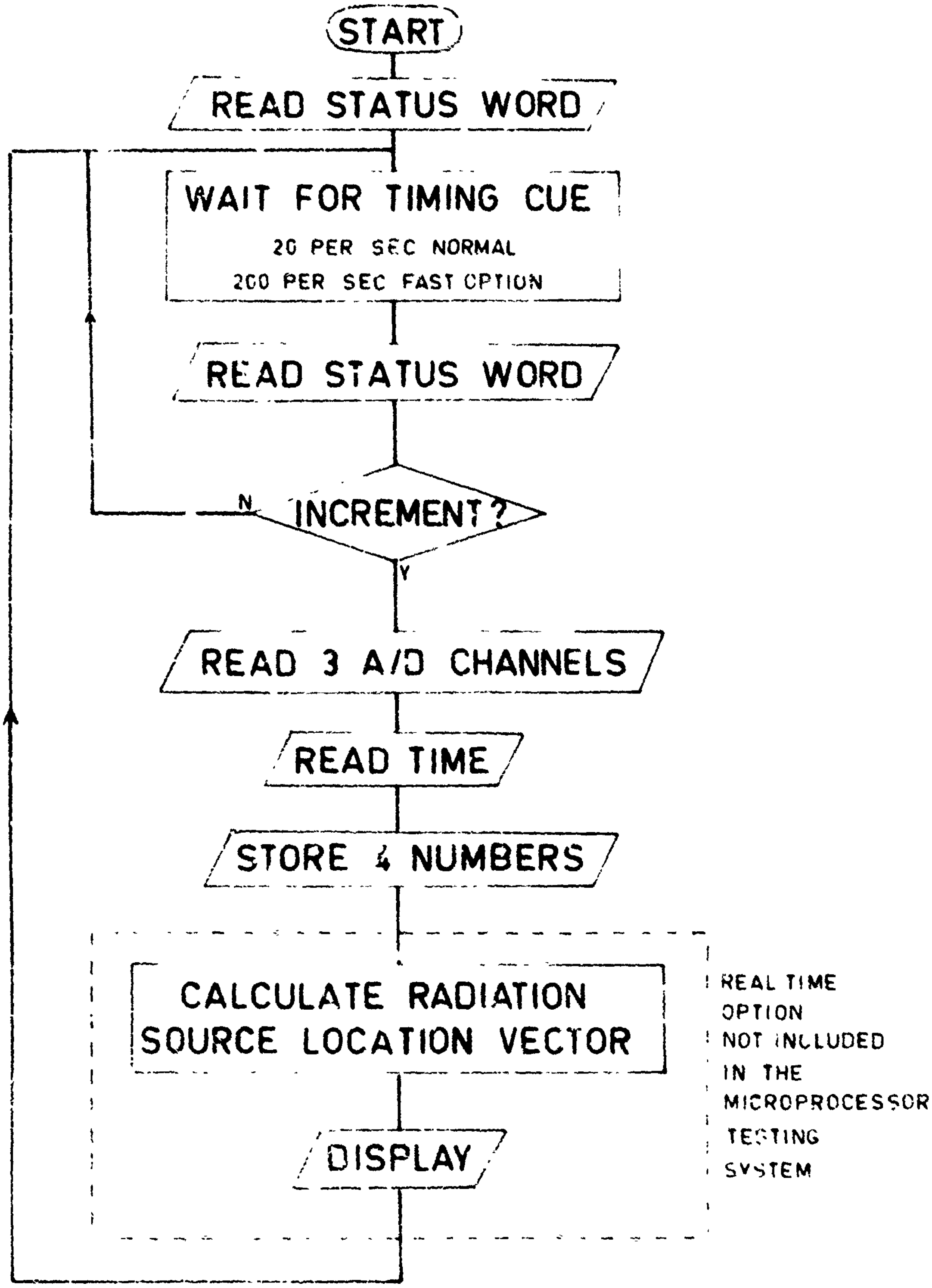

START
READ STATUS WORD
WAIT FOR TIMING CUE
20 PER SEC NORMAL
200 PER SEC FAST OPTION
READ STATUS WORD
INCREMENT?
N
Y
READ 3 A/D CHANNELS
READ TIME
STORE 4 NUMBERS
CALCULATE RADIATION SOURCE LOCATION VECTOR
DISPLAY
REAL TIME OPTION NOT INCLUDED IN THE MICROPROCESSOR TESTING SYSTEM

Figure [illegible]. A skeletal flow chart of the data acquisition software as implemented on a microprocessor.

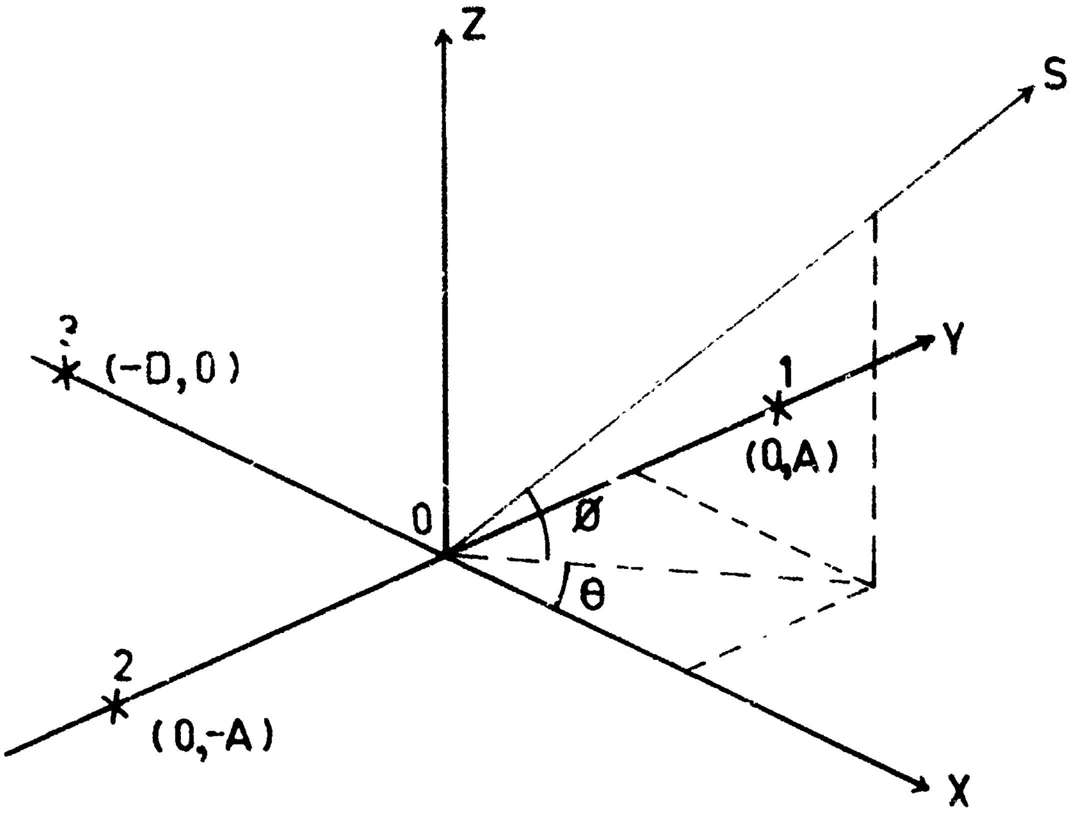


Figure 5. The coordinate system and line OS to the radiation source.

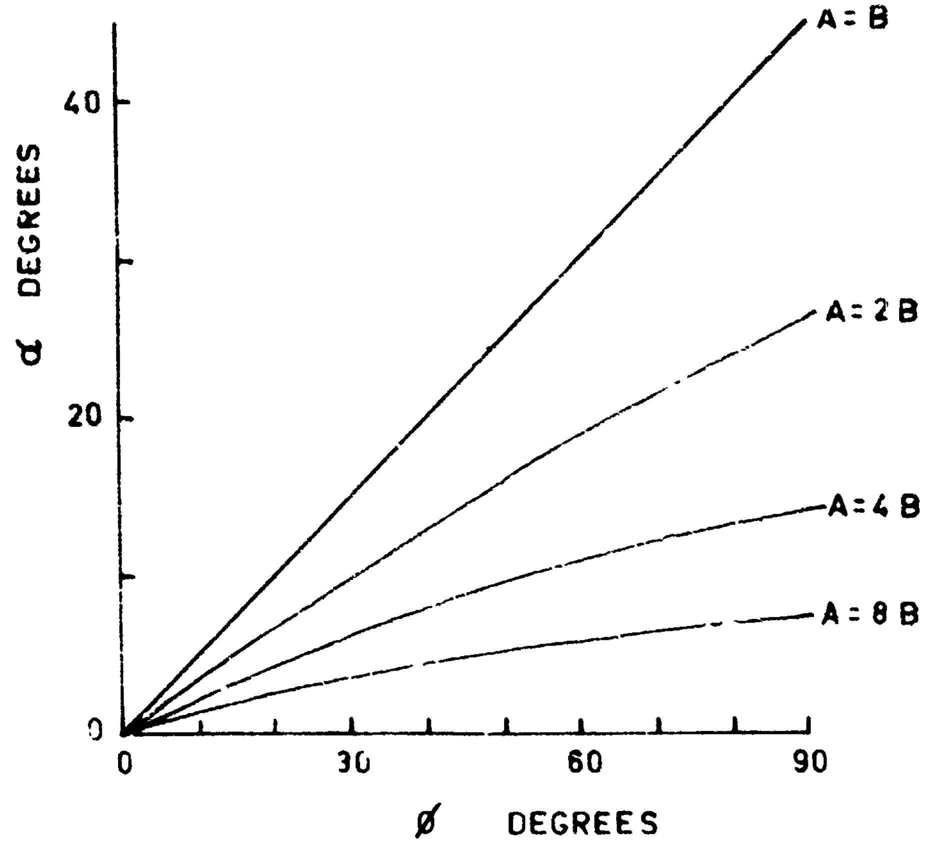


Figure 6. The phase meter is weighted towards the stronger of two radiation sources. A and B are amplitudes and the phase angles α and ϕ are defined in the text.

A NEW COMPUTER-CONTROLLED HIGH FREQUENCY DIRECTION-FINDING AND TRANSMITTER LOCATING SYSTEM

R. L. Tanner, Ph.D
Technology for Communications International (TCI)
1625 Stierlin Road
Mountain View, California 94043 USA

SUMMARY

The direction-finding and transmitter locating system to be described here represents a new approach to this old problem that offers a number of benefits when compared with more traditional systems. These include:

- High speed and the ability to operate against very short duration signals with high accuracy
- Much improved ability against short-range, high-angle sky-wave signals
- Single-station locator capability
- Much less critical siting requirements
- Easily adapted to automatic netting of a multi-station system
- Minimal requirements for trained technical personnel.

Operation of the TCI system depends heavily upon the ability to compute with high accuracy the precise response of the novel, inward-looking antenna array systems used. These computations provide both phase and amplitude response, both of which are necessary to the technique employed, and utilize accurate computational techniques based on numerical solution of the integral equations defining both the DF array and all relevant boundary conditions. The DF process is essentially the inverse of the array response computation. Under computer control the response of the array to the incoming wave is accurately measured and is matched to the re-computed response characteristics stored in the computer memory. Both azimuth and elevation angle of the arriving wave are given with an accuracy that is limited essentially only by variations in the ionosphere.

The DF and transmitter locating system is easily interfaced with computerized spectrum monitoring and signal classification and recognition equipment to provide an almost fully automatic signal intercept and response system.

1. INTRODUCTION

The TCI computer-controlled direction-finding system provides a standard of performance with respect to speed, required signal-to-noise ratio, and accuracy that has heretofore been unobtainable in even the largest and most complex HF direction-finding systems. This paper explains the underlying system design features that make this standard of performance possible and documents actual results to date.

Since the end of World War II almost all large aperture HF direction-finding systems have been of the Wullenweber type. Such systems have usually taken the form of several bands of vertical monopoles backed by reflecting screens. The DF process in these systems involves a rotating goniometer which produces an array beam that rotates in a circular scanning mode. The signal received by the scanning beam has traditionally been displayed on a cathode ray tube, where each successive scan produces a new image, generally corresponding to the shape of the scanning beam, but modified by modulations of the signal, fading and other factors.

In most such systems the operator visually integrates out the effects of modulation and fading, and seeks a line of symmetry in the displayed pattern that indicates the azimuthal angle of arrival. Recent improvements in some systems have replaced (or augmented) the visual displays by systems that digitize the goniometer outputs and utilize mini-computers to compute the probable bearings from the digitized goniometer data.

Traditional HF Wullenweber systems would be subject to polarization errors when operating against sky waves if horizontally-polarized antennas were used. To avoid such errors virtually all of the traditional systems have employed only vertically-polarized antenna elements which suppress response to horizontally-polarized components of the incoming waves. Such elements have nulls at the zenith and these systems therefore yield poor results when operating against high-angle sky-wave signals emanating from short-range emitters. Recently, due to rapid advances in computer technology, it has been possible to apply powerful, large-scale computers to the design of novel antenna arrays and much smaller, but still quite powerful, mini-computers to the processing and manipulating of signals received by these arrays. As a consequence, new DF systems have entered service which perform the DF estimate much more rapidly, provide automatic and very rapid netting of the several stations in a multiple station system, greatly reduce the adverse effects of signal modulation and fading, deal accurately with short-range high-angle signals, and provide a most useful single station locator capability. Moreover, the cost of these new systems is a small fraction of the cost of the better of the older systems.

2. DIRECTION-FINDING AND TRANSMITTER LOCATING TECHNIQUE

2.1 Array Response

The key to the direction-finding technique employed is a very accurate knowledge of the response characteristics of the array. Both phase and amplitude response at all points in space, at all frequencies, and for both vertically and horizontally-polarized incoming waves must be known with great precision and stored in the memory of the site computer.

The array response, including the effects of scattering and absorption by the antenna elements opposite the excited sector of the array, are obtained by computations in which the integral equations defining the entire array and all relevant boundary conditions are solved numerically. (TANNER, R. L. and ANDREASEN, M. G., 1967) (TANNER, R. L. and WHARTON, W., 1978) The array response characteristics must of course, be stored in the computer at a finite number of discrete frequencies and angles. Sophisticated and accurate interpolation techniques are employed to obtain the response at frequencies and angles different from the exact values for which the data are stored.

2.2 Data Collection System

The system by which the array response data are collected that permit the direction of arrival of the incoming wave to be computed comprises three major components: The computer-controlled switched beamformers, the dual-channel receiver, and the site computer.

2.2.1 Computer-Controlled Beamformers

The beamformers used for the DF function are essentially as described below, but incorporate PIN diode switching matrices that permit the beams to be switched by computer command to any of the beam positions available in the array. Either sum or difference beams can be selected. Switching from one beam position or mode to any other position is accomplished in approximately 200 µsec.

The system includes two identical but completely independent beamformers. One of these feeds the "reference channel" of the receiver; the other the "sampling channel."

2.2.2 The Dual-Channel DF Receiver

The dual-channel DF receiver is actually two separate high-quality receivers having a common, computer-controlled local oscillator. Each channel is essentially a high accuracy, high-dynamic range, computer tunable radio-frequency voltmeter that measures precisely the voltage applied to its input terminals. The output of each channel is a pair of numbers that give both the amplitude and the phase of the input voltage. These two numbers appear in digital form on a pair of 12-bit A/D converters that are read by the computer.

2.2.3 Site Computer and DF Algorithm

In gathering data for a DF computation, the "reference channel" is connected by computer command to the antenna beam pointed in the general direction of maximum signal and is not switched during the data gathering process. Voltages measured on this channel include all amplitude and phase variations -- such as those due to modulation -- associated with the incoming signal wave. The other channel of the receiver, called the "sampling channel," is switched by computer command among the different beams of the array. Voltages measured on this channel include all amplitude and phase variations associated with the signal itself and, in addition, include amplitude and phase variations that arise due to the differences in response to the incoming signal wave of the different beams of the array.

By comparing the voltages measured on the "sampling channel" to those measured on the "reference channel," the phase and amplitude variations of the incoming signal itself (such as those caused by modulation) are eliminated, leaving a set of complex numbers (phase and amplitude) caused by differences in phase and amplitude response to the incoming wave of the different beams of the array. Using these numbers the computer calculates the direction of the arriving wave, including both azimuth direction and elevation angle. It does so by comparing the measured responses of the different beams with their ideal pre-computed responses stored in the computer memory. Data for one DF "cut" or "snap' is collected in as little as 200 ms.

A block diagram of one station of a typical system is shown in Fig. 1. Figure 2 is a photograph of the electronic equipment complement at a typical station.

3. ANTENNA ARRAY AND BEAMFORMING

Perhaps the single most important component determining system performance is the antenna array which will now be discussed.

3.1 Array Configuration

The arrays employed are novel, utilizing quasi-log-periodic elements arrayed in an inward-looking annulus. Vertically-polarized, horizontally-polarized and dual-polarized systems are available. A dual-polarized array, the TCI Model 410, is illustrated in Fig. 3. Figure 4 is a photograph of a horizontally-polarized array, the TCI Model 402. The dual-polarized array consists of independent vertically and horizontally-polarized arrays, co-located and sharing a common structural support system.

The number of elements in the array is typically 20 or 24, although arrays with as many as 36 elements and as few as 18 elements have been built. The array shown in Fig. 3 has an overall diameter of 300 meters which is the largest that has been built, although larger arrays are possible. The array of Fig. 4 is smaller, 216 meter diameter. Low frequency performance tends to suffer if the array is made smaller, although the 216m arrays of the type shown in Fig. 4 perform well to a lower frequency of 2 MHz.

3.2 Beamforming

The quasi log periodic antennas are arrayed with their vertices on an inner circle of finite radius and not at the center. This fact is of considerable importance. As will be described, it permits the formation of low frequency beams as narrow as permitted by the maximum diameter of the available site and at the same time allows the beamwidth at the highest operating frequency to be as narrow as is consistent with the number of beams available in the array.

As is well known log-periodic antennas tend to radiate in the direction of their vertices -- that is, toward the small end of the antenna or toward the center of the array. Directive beams are formed, just as in a more conventional outward-looking Wullenweber, by combining the outputs of a number of elements through phase delay lines which serve to collimate the radiation. In outward-looking circular arrays the phasing lines with the greatest delay are in series with the elements at the center of the illuminated sector. In the inward-looking arrays described here the phasing lines with greatest delay are in series with the elements at the edges of the illuminated sector. Typically the elements occupying an arc of 120 degrees are combined to form a beam. By combining the elements in each half of the sector separately, and then by combining these subsectors in a hybrid combiner with both sum and difference outputs, both sum and difference beams are obtained.

Figures 5 and 6 show representative patterns obtained with the arrays described. For monitoring or communications applications beams equal in number to the number of array elements can readily be made available. For the direction-finding and transmitter locating function, separate computer-controlled switchable beams are employed. These can be made available simultaneously with multiple fixed beams. A true omniazimuthal beam can also be furnished if required.

3.3 Advantages of Inward-Looking Array Configuration

For one accustomed to thinking of circular arrays in the context of the conventional outward-looking configuration, the type of arrays just described seem strange -- even bizarre -- and raise many questions. Upon further consideration, however, it becomes apparent that the advantages are many and significant while the disadvantages are more imaginary than real. Some of the advantages will now be discussed. The seeming disadvantages will also be considered more carefully.

3.3.1 Large Bandwidth and Simplified RF Distribution System

In the conventional Wullenweber systems the operating frequency band is covered as a series of sub-bands, each typically of 3:1 bandwidth. At least two and frequently three sub-bands are required to cover the bandwidth of interest. For the inward-looking arrays described, an operating range of nearly 20:1 (1.6 to 30 MHz) has been covered in a single band. This is possible because of the very broadband capabilities of the quasi-log-periodic antenna elements used in the array.

The fact that a wide operating bandwidth can be covered by the array in a single band greatly simplifies the RF distribution system in addition to making it much less costly. For example, a typical TCI system requires only 24 multicouplers and beamformers to provide 24 monitor beams and provide for the DF capability. By contrast, the FLR-9, perhaps the best of the conventional systems, has 48 low-band elements, 96 mid-band elements, and 48 high-band elements; a total of 192 elements. Typically 16 elements are combined in the beamformers to form the monitor beams with additional outputs to accommodate the DF function. To do this each element must be provided with multicouplers having many more outputs than are required in the inward-looking arrays. In addition, the monitor beamformers are both much more numerous and more complex. Each sub-band beam combines the outputs from many more elements, there are more beams required for adequate angular coverage in each sub-band, and each sub-band requires its own set of beamformers. Finally to obtain monitor beams that cover the entire operating frequency band, the separate beams from the three different bands must be combined in multiplexing units.

3.3.2 Efficient Use of Available Aperture -- Smoother and Smaller Variation of Beamwidth with Frequency

The inward-looking quasi-log-periodic arrays make uniquely efficient use of the available aperture. This results, in turn, in both smaller and smoother variation of array beamwidth over the operating frequency range. It is a well known characteristic of log-periodic antennas that their active regions change physical position with changes in frequency. At low frequencies the active, or radiating, region is at the large end of the antennas and progresses forward toward the vertex as the frequency increases. In the circular arrays described here this means that the effective aperture at low frequencies corresponds with the maximum diameter of the array. As frequency increases the active region moves toward the vertex or toward the center of the array. Consequently the effective physical size of the aperture decreases, but because the vertices of the antennas are on a circle of significant radius and not at the center of the array, the electrical size of the aperture increases, and the beamwidth narrows. With the design parameters available, the arrays can be designed to have a low-frequency beamwidth as narrow as the maximum diameter of the available site permits, and a high frequency beamwidth just wide enough to cover the required angular sector; for example, 15 degrees with a 24-element array. Typically, over a 15:1 frequency band the variation in beamwidth is approximately 3:1.

3.3.3 Symmetry of Electrical Environment and Relative Immunity to Site Irregularities

A further advantage of the inward-looking array is the symmetry of the electrical environment inherent in this configuration and the comparative immunity to site irregularities and obstacles in the area surrounding the array. In a conventional outward-looking array the area outside the array itself must be smooth, uniform, and free of scattering obstacles for a substantial distance. Otherwise the beam shape will change as the beam rotates, resulting in significant DF errors. For the inward-looking array, on the other hand, the separation between the active sector of the array and site irregularities or scattering obstacles is effectively increased by the diameter of the array. Irregularities or obstacles outside the array therefore have much less effect on beam shape and DF performance. None of the sites where such systems are presently installed and performing well would be suitable for installation of an outward-looking array.

3.3.4 Ground Screen Considerations

The ground screens used with present outward-looking arrays are much too short to affect radiation patterns and serve only to stabilize element impedance. In such arrays beams at different angular positions overlook different areas of ground, and to make the ground screen sufficiently long to benefit the radiation patterns would be both very expensive and increase by approximately nine-fold the land area required. For the inward-looking array, beams at different angles overlook the same common area. By covering the area occupied by the array with a ground screen, all beams are provided with a ground screen of length essentially equal to array diameter. The improvement in gain and low angle radiation which results more than compensates for the power absorbed or scattered by the unexcited elements opposite the excited sector.

3.3.5 Ease of Testing

Another advantage of the inward-looking configuration of considerable importance is the ease with which the array can be tested. By injecting a signal into a small omniazimuthal antenna at the geometric center of the array, the array elements, beamformers, multicouplers, and other components of the system can be tested easily, rapidly, and accurately, and any malfunctions diagnosed. The testing operation can be accomplished largely by the site computer.

4. TYPICAL RESULTS

Figures 7a and 7b are reproductions of actual hard copy printouts of DF results from one site of a four-site system. Each line of the printout represents the results of one "cut" for which the data were gathered in less than 300 ms. Computation of the bearing from the measured results requires approximately two seconds. Signal-to-noise ratio was approximately 15 to 20 dB. From left to right the printout columns give the date and the local time (to nearest second) that the data were gathered, the computed azimuth bearing, the computed elevation angle, and finally a quantity called the "confidence factor." There are several features evident in the data shown that deserve comment. First, the confidence factor: This number is very closely akin to a correlation coefficient between actual measured voltages and a hypothetical set of idealized voltages derived from the stored array response characteristics. The DF computation consists, in effect, of finding the best possible match between the actual (normalized) measured voltages at the different beam output ports and a corresponding set of voltages derived from the stored array response data as the hypothesized direction of arrival of the wave is shifted about. The hypothesized direction yielding the best correlation is taken as the correct direction and printed out.

It will be noted that all but one of the confidence factors in the examples of Fig. 7 exceed 0.9 and there are two instances where the confidence factor is 0.994. The theoretical maximum for this factor is 1.00 and it is appropriate to consider the many different circumstances that can result in a lower value. Errors in the pre-computed stored response data or in the interpolation between the discrete values of frequency and angle that are stored in the computer are two of the most interesting possibilities. The values of 0.974 achieved against actual signals implies very high accuracy of the stored data and consequently of the computational techniques by which it was derived. Factors which could result in values lower than 1.00 include noise, measurement error, interfering signals, contamination of the principal wave by other propagating modes and others.

Another point of interest is that the different "cuts," which were taken several seconds apart in time, do not scatter randomly but vary in a systematic manner with time. For example, the first group of results in Fig. 7a shows a small negative azimuth error at first. The bearing error then passes through zero, goes slightly positive, then drifts negative, reaching a maximum value of more than 4 degrees before it again turns in the positive direction and reaches a maximum positive value of 1.7 degrees before again turning negative. Such behavior is exactly what one might expect from the passage of a so-called TID (travelling ionospheric disturbance) (JONES, T. B. and REYNOLDS, J. S. B., 1971). Space limitations preclude the inclusion of many more similar examples.

5. INTEGRATED SYSTEM

Typically several sites are netted into a system. The individual sites automatically report their DF results to a central command center where a computer uses the site results to calculate the location of the transmitter.

Figure 8 reproduces typical hard copy reports generated in response to a system DF command. Figure 8a shows the report at one site of the sytem, while Fig. 8b shows the emitter location report at the command center for the same event. At the operator's option the site report could be altered to eliminate the results for individual cuts and to print out only the average summary. It will be noted in the command center report that latitude and longitude degrees have been deleted at the system owner's request to avoid revealing location. It should also be mentioned that the calculated radius of uncertainty (1 km) shown in Fig. 8b is somewhat optimistic in this example. The actual transmitter location was approximately 10 km from the calculated location.

6 SINGLE STATION LOCATION

The ability of the system to measure the elevation angle of arrival as well as its azimuth gives it the useful ability to locate the source of a signal from a single DF station Normally, emitter location is accomplished using azimuth bearing data and triangulation. However, on some occasions only one station will obtain a high confidence bearing, or two stations might obtain bearings but the angle between the lines of bearing from the two stations is too small to allow accurate triangulation. In such circumstances the single station location capability is the only means by which an estimate of emitter location can be obtained.

Single station location requires in addition to elevation and azimuth angles, a knowledge of the effective height of the ionosphere at the frequency of the target signal. Ideally this information should be available on a real-time basis together with information concerning ionospheric tilts, etc. Experience with the systems described has shown, however that useful results can be obtained using only the ionospheric propagation prediction data obtained from the Institute of Telecommunications Sciences (ITS). Table 1 shows a number of examples of the results obtained this way. Use of a real-time ionospheric sounder properly incorporated into the system would almost certainly improve upon the results obtained thus far.

7. COMPUTERIZED SPECTRUM SURVEILLANCE AND SIGNAL RECOGNITION

Although a complete description is beyond the scope of this paper, it is appropriate to mention that the computerized DF system just described is ideally adapted to operation with companion automated spectrum surveillance and signal classification or recognition equipment. TCI has furnished spectrum monitor equipment which, depending upon the channel width defined, is able to scan the HF spectrum at a rate exceeding 20 MHz sec, catalog and remember occupied channels, and report newly appeared signals to a computerized classifier/recognizer. The classifier/recognizer can be programmed to recognize signals of particular interest if their modulation characteristics can be defined with adequate precision, and to refer them automatically to the emitter locator system which then determines the location of the transmitter from which the signal originates.

8. CONCLUSION

The direction-finding and transmitter locating system just described offers many advantages of speed, accuracy, economy, and immunity to site irregularities over older, more traditional systems. These benefits derive to some extent from the novel approach used. Of central importance, however, is the extensive use of modern high speed computers. These permit the DF array to be designed and its response characteristics to be known with a precision heretofore not possible. They also permit the rapid and accurate collection of almost all signal response data potentially available from the array and its rapid processing to obtain the direction of arrival of the signal of interest.

TANNER, R. L. and ANDREASEN, M. G., September 1967, "Numerical Solution of Electromagnetic Problems", IEEE Spectrum, Vol. 4, No. 9, pp. 53-61.

TANNER, R. L. and WHARTON, W., 1978, "The Computer-Aided Design of LF, MF, and HF Antennas", presented at IEEE Conference.

JONES T. B. and REYNOLDS, J. S. B., January/February 1975, "Ionospheric Perturbations and Their Effect on the Accuracy of HF Direction-Finders", The Radio and Electronic Engineer, Vol. 45, No. 1/2.

Table 1. Summary of Typical Test Data of Transmitter Location from a Single Site

Local Time	Frequency (MHz)	Known Dist to Sig Source (km)	Elev Angle of Arrival (deg)	Refl Layer	Azimuth Error (Great Circle degrees)	Error in Locating Emitter by Single Site (km)
11 50	3 056	231	38 4	E	0 8	14
11.52	3 056	231	39 0	E	0 2	9
12 07	5 191	231	67 5	F	0 6	18
12 39	3 680	264	34 8	E	0 2	20
12 46	3 292	306	28 7	E	0 7	35
12.49	5 295	306	60 3	F	0 9	24
12 51	7 299	306	59 0	F	1 3	40
13 00	7 299	306	58 5	F	0 9	44
15 57	4 969	36 1	78 5	E	1 8	10
16 02	3 681	36 1	79 6	E	1 4	5

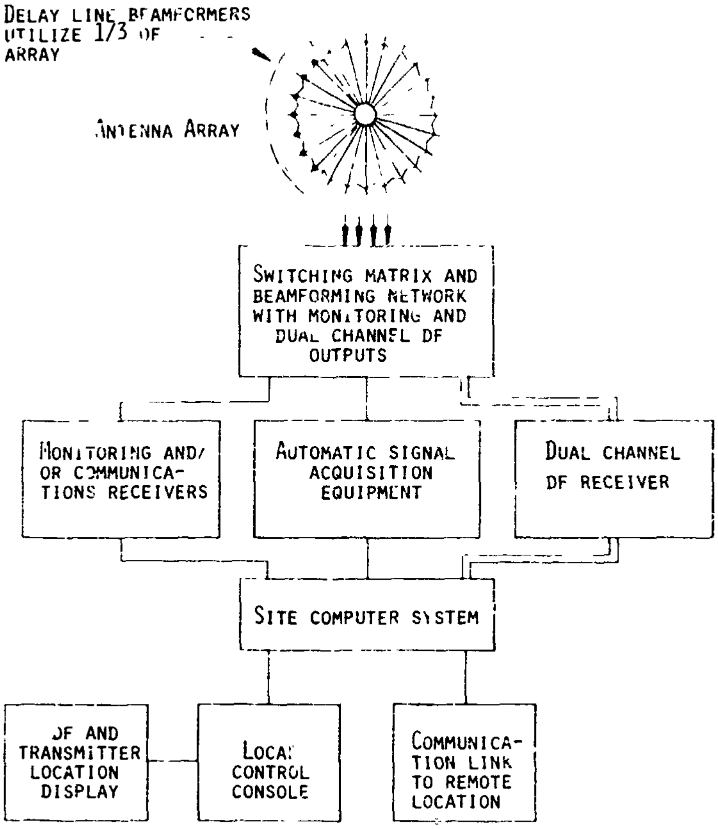

DELAY LINE BEAMFORMERS UTILIZE 1/3 OF ARRAY
ANTENNA ARRAY
SWITCHING MATRIX AND BEAMFORMING NETWORK WITH MONITORING AND DUAL CHANNEL DF OUTPUTS
MONITORING AND/ OR COMMUNICATIONS RECEIVERS
AUTOMATIC SIGNAL ACQUISITION EQUIPMENT
DUAL CHANNEL DF RECEIVER
SITE COMPUTER SYSTEM
DF AND TRANSMITTER LOCATION DISPLAY
LOCAL CONTROL CONSOLE
COMMUNICATION LINK TO REMOTE LOCATION

Fig. 1 System Block Diagram

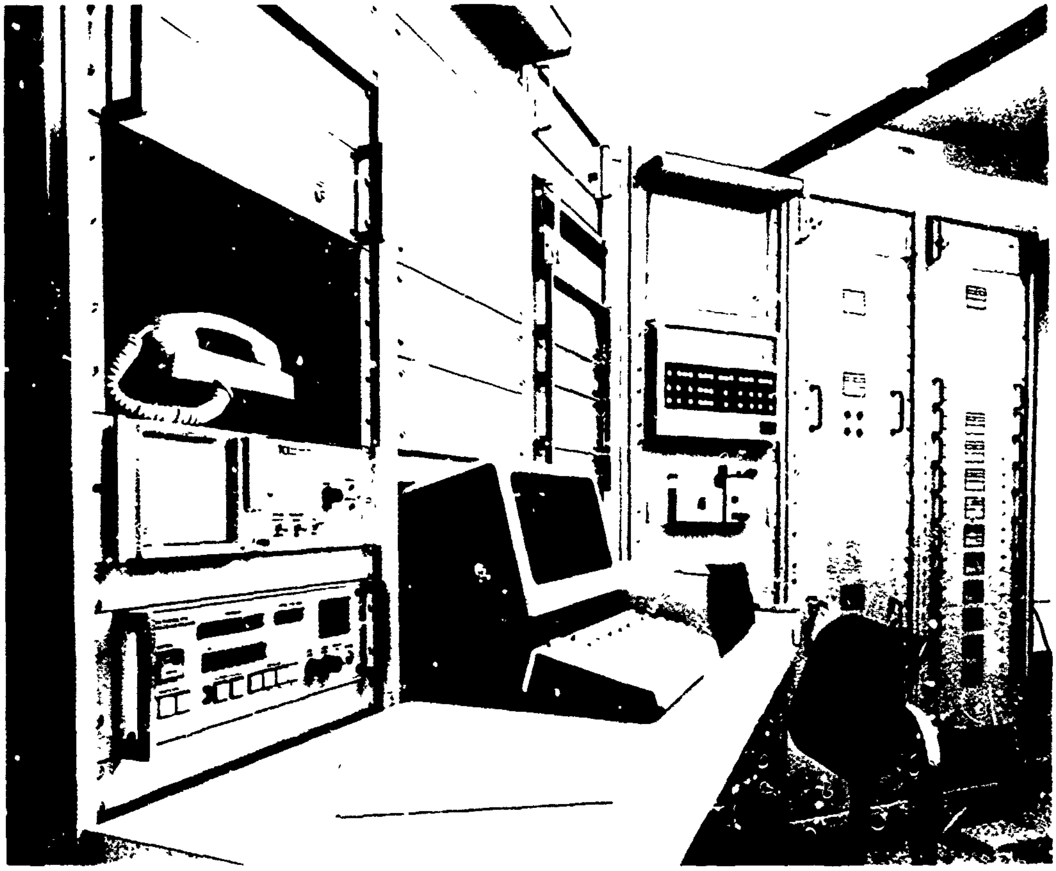


Fig. 2 Electronic Equipment at Typical DF Site

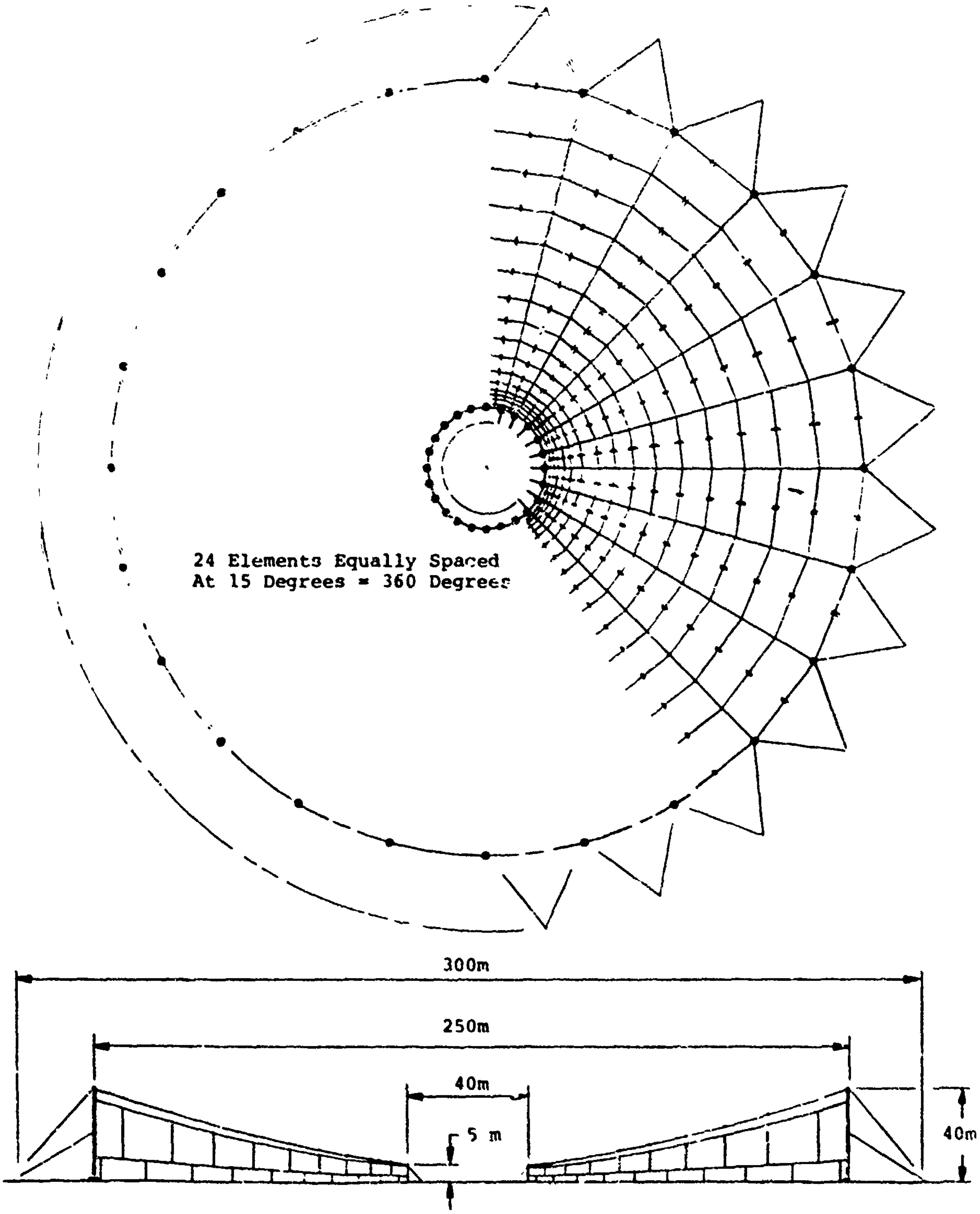

24 Elements Equally Spaced
At 15 Degrees = 360 Degrees
300m
250m
40m
5 m
40m

Fig. 3 Plan and Elevation Views of Antenna Array

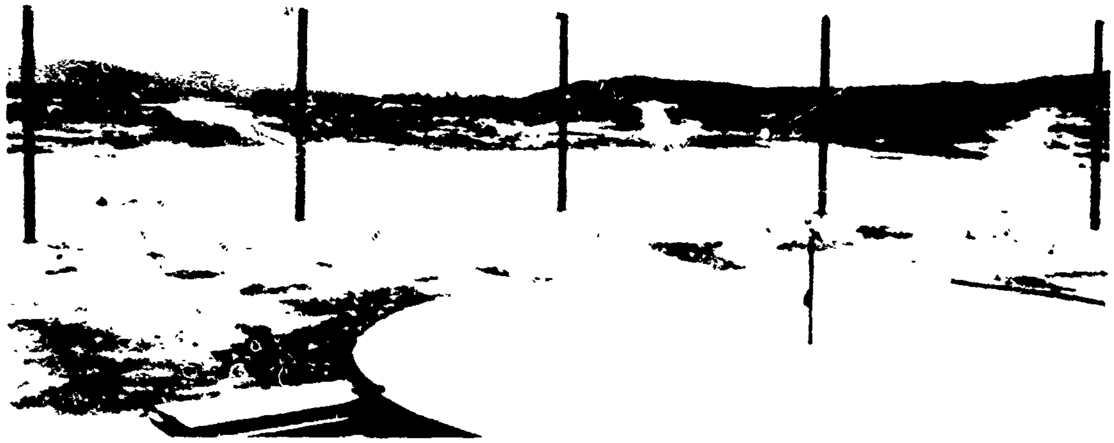


Fig. 4 Photograph of TCI Model 402 Horizontally-Polarized Array

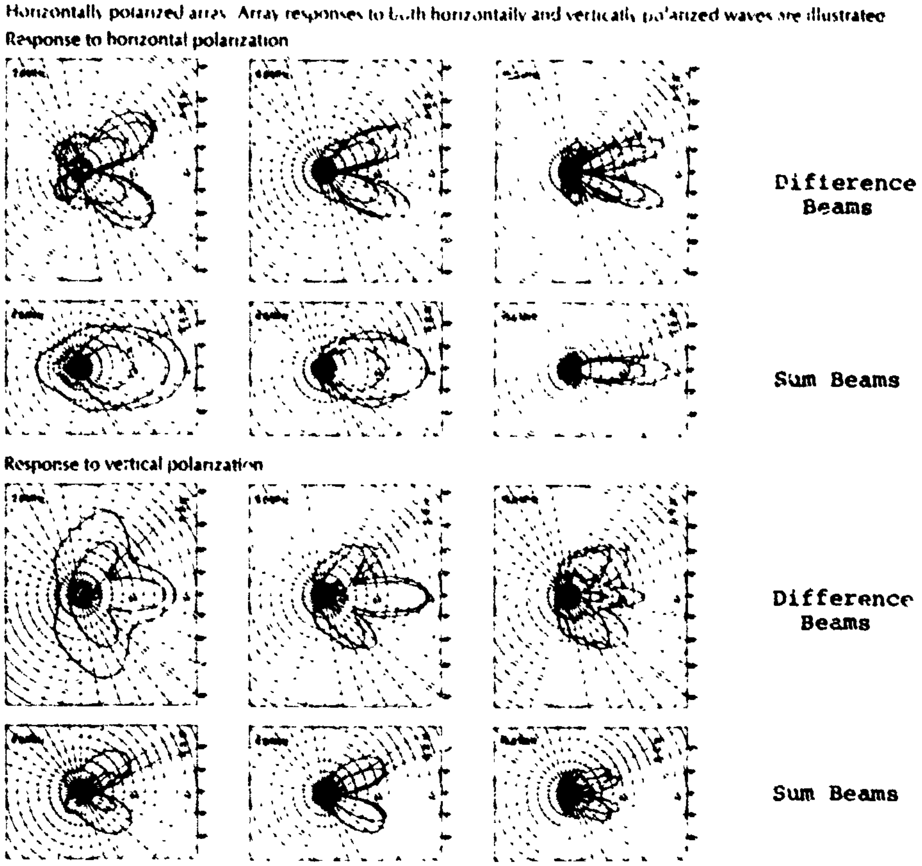


Fig. 5 Representative Patterns of 18 Element 216m Horizontally-Polarized Array. This Array was Designed to Optimize Response to High Angle Waves.

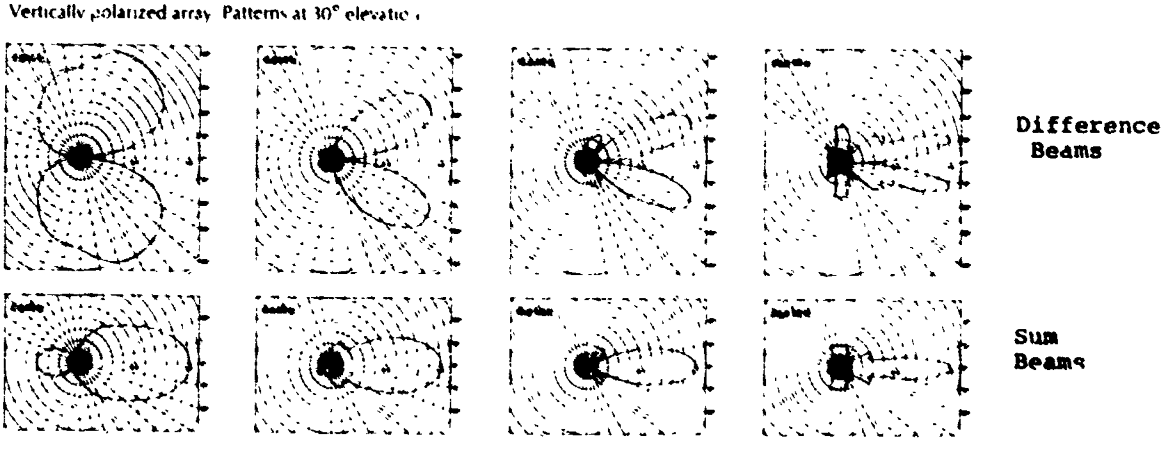


Fig. 6 Representative Patterns of 24 Element 330m Vertically-Polarized Array.

```
04/30/77  11:48:43   3.056 MHZ  AZ = 259.8  EL =  39.2  C =  .991
04/30/77  11:48:52   3.056 MHZ  AZ = 261.0  EL =  41.4  C =  .930
04/30/77  11:49:00   3.056 MHZ  AZ = 261.7  EL =  37.8  C =  .978
04/30/77  11:49:09   3.056 MHZ  AZ = 259.7  EL =  36.2  C =  .989
04/30/77  11:49:17   3.056 MHZ  AZ = 258.6  EL =  36.7  C =  .994
04/30/77  11:49:26   3.056 MHZ  AZ = 256.4  EL =  36.1  C =  .987
04/30/77  11:49:34   3.056 MHZ  AZ = 260.0  EL =  37.5  C =  .986
04/30/77  11:49:43   3.056 MHZ  AZ = 259.0  EL =  39.0  C =  .983
04/30/77  11:49:51   3.056 MHZ  AZ = 261.4  EL =  39.5  C =  .981
04/30/77  11:50:00   3.056 MHZ  AZ = 262.7  EL =  39.1  C =  .983
04/30/77  11:50:09   3.056 MHZ  AZ = 260.2  EL =  39.5  C =  .991
```

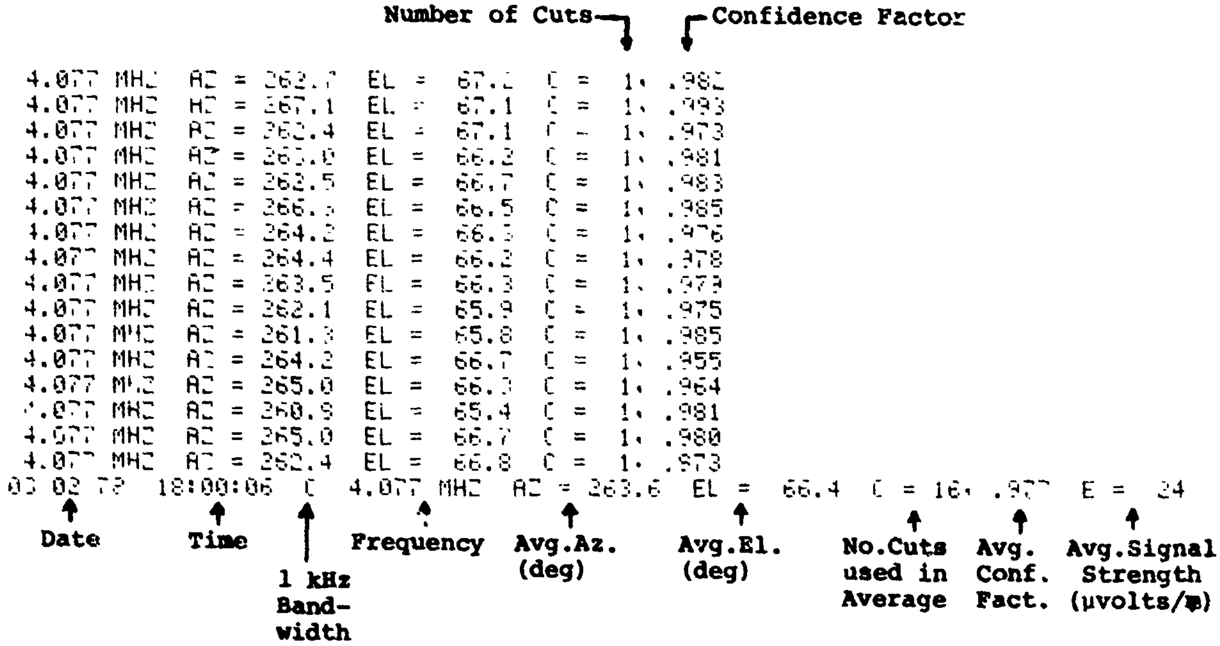
```
04/30/77  12:07:20   5.191 MHZ  AZ = 260.3  EL =  69.1  C =  .976
04/30/77  12:07:29   5.191 MHZ  AZ = 233.1  EL =  67.5  C =  .990
04/30/77  12:07:37   5.191 MHZ  AZ = 253.8  EL =  73.9  C =  .900
04/30/77  12:07:45   5.191 MHZ  AZ = 259.5  EL =  63.8  C =  .982
04/30/77  12:07:54   5.191 MHZ  AZ = 262.5  EL =  70.0  C =  .903
04/30/77  12:08:02   5.191 MHZ  AZ = 265.4  EL =  71.9  C =  .894
04/30/77  12:08:19   5.191 MHZ  AZ = 261.5  EL =  66.4  C =  .984
04/30/77  12:08:28   5.191 MHZ  AZ = 259.4  EL =  57.7  C =  .951
```

Fig. 7a Typical DF Results with TCI Transmitter Locator System. True Bearing of Signal Source is 261.0 Degrees. E-Layer and F-Layer Propagation.

```
04/30/77  12:48:53   5.295 MHZ  AZ = 320.0  EL =  57.5  C =  .976
04/30/77  12:49:01   5.295 MHZ  AZ = 322.9  EL =  54.1  C =  .974
04/30/77  12:49:10   5.295 MHZ  AZ = 319.0  EL =  64.1  C =  .984
04/30/77  12:49:18   5.295 MHZ  AZ = 320.0  EL =  60.9  C =  .993
04/30/77  12:49:26   5.295 MHZ  AZ = 317.5  EL =  60.7  C =  .979
04/30/77  12:49:35   5.295 MHZ  AZ = 320.2  EL =  60.8  C =  .994
04/30/77  12:49:43   5.295 MHZ  AZ = 319.7  EL =  60.4  C =  .990
04/30/77  12:49:51   5.295 MHZ  AZ = 319.8  EL =  59.6  C =  .951
04/30/77  12:50:00   5.295 MHZ  AZ = 318.5  EL =  59.9  C =  .990
```

```
04/30/77  12:51:51   7.299 MHZ  AZ = 322.3  EL =  60.3  C =  .987
04/30/77  12:51:59   7.299 MHZ  AZ = 323.2  EL =  59.7  C =  .992
04/30/77  12:52:08   7.299 MHZ  AZ = 322.5  EL =  60.0  C =  .988
04/30/77  12:52:16   7.299 MHZ  AZ = 323.5  EL =  60.1  C =  .987
04/30/77  12:52:24   7.299 MHZ  AZ = 322.8  EL =  60.2  C =  .984
04/30/77  12:52:32   7.299 MHZ  AZ = 323.3  EL =  58.3  C =  .987
04/30/77  12:52:41   7.299 MHZ  AZ = 325.1  EL =  58.7  C =  .977
04/30/77  12:52:50   7.299 MHZ  AZ = 325.0  EL =  58.0  C =  .979
04/30/77  12:52:58   7.299 MHZ  AZ = 325.9  EL =  57.9  C =  .986
04/30/77  12:53:07   7.299 MHZ  AZ = 323.0  EL =  56.7  C =  .984
```

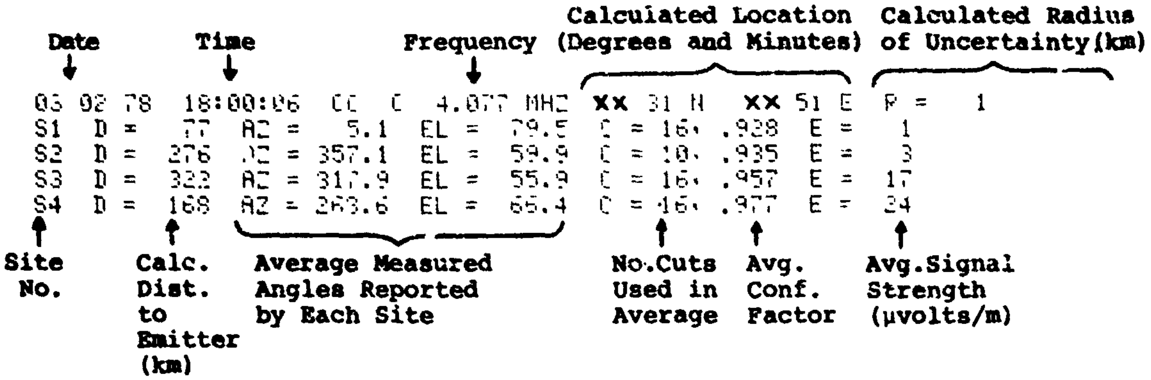
Fig. 7b DF Results at Two F-Layer Frequencies. True Bearing is 319.9 Degrees.



```
4.077 MHZ  AZ = 262.7  EL =  67.2  C =  1, .982
4.077 MHZ  HZ = 267.1  EL =  67.1  C =  1, .993
4.077 MHZ  AZ = 262.4  EL =  67.1  C =  1, .973
4.077 MHZ  AZ = 263.0  EL =  66.2  C =  1, .981
4.077 MHZ  AZ = 262.5  EL =  66.7  C =  1, .983
4.077 MHZ  AZ = 266.5  EL =  66.5  C =  1, .985
4.077 MHZ  AZ = 264.2  EL =  66.3  C =  1, .976
4.077 MHZ  AZ = 264.4  EL =  66.2  C =  1, .978
4.077 MHZ  AZ = 263.5  FL =  66.3  C =  1, .979
4.077 MHZ  AZ = 262.1  EL =  65.9  C =  1, .975
4.077 MHZ  AZ = 261.3  EL =  65.8  C =  1, .985
4.077 MHZ  AZ = 264.2  EL =  66.7  C =  1, .955
4.077 MHZ  AZ = 265.0  EL =  66.3  C =  1, .964
4.077 MHZ  AZ = 260.9  EL =  65.4  C =  1, .981
4.077 MHZ  AZ = 265.0  EL =  66.7  C =  1, .980
4.077 MHZ  AZ = 262.4  EL =  66.8  C =  1, .973
03 02 78  18:00:06  C  4.077 MHZ  AZ = 263.6  EL =  66.4  C = 16, .977  E =  24
```

Number of Cuts
Confidence Factor
Date
Time
1 kHz Bandwidth
Frequency
Avg.Az. (deg)
Avg.El. (deg)
No.Cuts used in Average
Avg. Conf. Fact.
Avg.Signal Strength (μvolts/m)

Fig. 8a Typical Hard Copy Printout of DF Results at One Site of a Four Site Network.



```
03 02 78  18:00:06  CC  C  4.077 MHZ  XX 31 N   XX 51 E  R =   1
S1  D =   77  AZ =   5.1  EL =  79.5  C = 16, .928  E =   1
S2  D =  276  AZ = 357.1  EL =  59.9  C = 10, .935  E =   3
S3  D =  322  AZ = 317.9  EL =  55.9  C = 16, .957  E =  17
S4  D =  168  AZ = 263.6  EL =  65.4  C = 16, .977  E =  24
```

Date
Time
Frequency
Calculated Location (Degrees and Minutes)
Calculated Radius of Uncertainty(km)
Site No.
Calc. Dist. to Emitter (km)
Average Measured Angles Reported by Each Site
No.Cuts Used in Average
Avg. Conf. Factor
Avg.Signal Strength (μvolts/m)

Fig 8b Emitter Location Printout at Command Center of Four DF Network.

DISCUSSION

A. Larcher Brinca, Po

(1) Can this DF technique be used in VLF, namely for magnetospheric whistlers recognizing the practical limitations of VLF antennas?

(2) When solving the current integral equations for the antenna array do you use a variation of the method of moments? How do you take into consideration the effect of the ground?

Author's Reply

(1) In principle yes, but practically the required antenna would be too large and expensive to be used at VLF frequencies.

(2) The method is closely related to what is now called the "Method of Moments". The original program which was the progenitor of the present TCI program was developed considerably before the term "Method of Moments" was invented. Finitely conducting ground effects are taken into account by assuming the ground to be a uniform lossy (conducting) half space and solving for the fields of a current element of finite extent in the presence of this "ground". The resulting fields are used as the Green's function in the solution of the integral equation.

DIGITAL ON-LINE PROCESSING AND DISPLAY OF MULTIPARAMETER HF TRANSMISSION DATA

K. Bibl and B. W. Reinisch
University of Lowell
Center for Atmospheric Research
Lowell, Massachusetts, U.S.A.

SUMMARY

Amplitude, phase, angle of arrival, polarization, frequency shift and signal travel time of a radio wave are simultaneously measured as functions of time and frequency in a new digital ionosonde, the Digisonde 128PS.

For the study of the structure and motions of the ionosphere and its influence on Over-the-Horizon detection systems, on sea-state measurements, on direction and location finding methods and on communication systems a selection of the measurable parameters is recorded and displayed in real time with the necessary resolution. Two main modes of operation are described, the ionogram mode using a Maximum Amplitude scheme for data compression and the Doppler-Drift mode using preselected frequencies and range bins for data compression.

Many applications in new fields are anticipated in addition to the thorough understanding of the ionosphere, its sources and its forces.

1. INTRODUCTION

Radio waves are characterized by amplitude, phase, angle of arrival, polarization, frequency shift and signal travel time. All these wave parameters are measured as functions of time and frequency and displayed in real-time in a new digital ionosonde, the Digisonde 128PS. For the study of the structure and the motion of a dispersive medium, like the ionosphere, and for a full analysis of the influence of the ionosphere on Over-the-Horizon detection systems, on sea-state measurements, or direction and location finding methods and on communication systems a digital radio sounder is employed with pulsed carrier frequencies scanning in steps through a large frequency band. Frequency scanning adds an important parameter to the wave characteristics of a conventional Radar system.

2. SYSTEM DESCRIPTION

As a scientific instrument, but useful for many operational tasks, we have developed a digital ionosonde which uses a large antenna array in a scanning or in a beam-forming mode. The Digisonde 128PS (K. Bibl and B. W. Reinisch, 1978) simultaneously integrates the echo amplitudes in a phase coherent mode at many Doppler frequencies, and for all chosen antennas or beam-directions. Echoes from 24 antennas can at once be spectrum analyzed in 128 coherent channels, or 16 spectral lines can be monitored for 128 range bins. Many other combinations of range windows, frequencies, polarizations, spectral channels and incidence angles can be processed simultaneously. Using the time between the transmitted pulses to measure complex spectrum amplitudes in up to 256 range bins (time sectors) requires a very fast on-line spectrum analyzer. The Processing Controller of the Digisonde forms the products between the sine and cosine samples of the high frequency (HF) signal and the trigonometric spectral function (of three averaged spectral lines) at a rate of 1 MHz. But such a large amount of data can not be handled on a continuous basis. We therefore operate the system in one of two data compression schemes: the Ionogram Mode with the Maximum-Amplitude Method, and the Doppler-Drift Mode with limited frequencies and ranges.

2.1 Maximum Amplitude Mode

The Maximum Amplitude Mode is used for recording of digital ionograms. For each of the 128 range bins of the more than 100 carrier frequencies the largest amplitude of the 24 possible channels is selected. The amplitude is recorded together with a status character, indicating the selected channel number. Since the signals arriving from different directions with different Doppler travel along different paths, they mostly have different travel time. Thus the maximum method does not suppress essential information.

In the contrary this differential method, selecting the strongest amplitudes, enhances the main features of the investigated propagation medium, as shown on the right side in Fig. 1. With frequency as the abscissa and range as ordinate the amplitudes (on the bottom) or the status characters (on top) form patterns from the numerical presentations as the third dimension (K. Bibl, 1974). Different features are clearly distinguishable simultaneously. While one area of reflection (showing large values of the status indicator on the top of the picture) moves toward the observation station, the older ionization, forming the lower pattern, with numerical values below 8, moves away. While the presence of the "4" bit (numbers between 5 and 7 or 13 and 15) indicates high Doppler speed, the numbers 1, 5, 9, 13 indicate one direction, 2, 6, 10, 14 another and the numbers 3, 7, 11 and 15 a third direction of the arriving echoes.

A different method of data presentation, the split-screen technique, is presented in Fig. 1a. Here two complementary data sets, echoes with ordinary and with extraordinary polarization, are presented in two separate ionograms. Splitting into several ionograms is possible by an on-line microcomputer (Fig. 2). On the bottom of the figure the total of the recorded data is presented first in the form of amplitude ionograms and above those as status ionograms. On top of there two ionograms the data are broken down into three practically independent amplitude ionograms which show echoes arriving from three different directions. Since the interference is split in the same way a substantial reduction in interference appearance is achieved that way. To increase the number of scanned directions the receiver antenna beam is looking into different directions in sequential ionograms which are corresponding to the left or the right side of Fig. 2. This is done to increase the number of reflection areas with the same range displayed simultaneously and to increase the sensitivity of the differential selection in the Maximum Amplitude mode which overcomes the limitations caused by the large beam width of the receiver antenna array. Thus the ionogram showing the vertical echo also includes echoes from directions not separated by the split screen method. Therefore the consecutive "vertical" ionograms appear somewhat different and must be further cleaned in the microcomputer to include only common echoes. Similarly two additional directions: north and south can be constructed by extracting echoes common in the NW and NE or SW and SE ionogram respectively. Thus it will be possible to construct false color ionograms indicating six different directions as six different colors plus overhead as white and still maintain amplitude and/or Doppler information as numbers. A third method is the suppression of unwanted echoes, as shown in Fig. 3. This method finds application mainly in automatic compression and analysis of large data series where one magneto-ionic component or oblique echoes should be suppressed (Fig. 4). But the switching between ionograms with and without suppression of specific echoes shows surprises even to experienced data analysts, specifically in the aurora region where oblique echoes are often interpreted wrongly. Thus the digital tagging of the echo properties is significant also for the manual data analysis as inputs for world-wide mapping.

2.2 Doppler-Drift Mode

In the Doppler-Drift Mode a full complex spectrum analysis is executed in real time during the measurement. The build-up of the spectra can be observed for any and all of the 24 independent channels by the programmable test features. Guided by the survey provided by the Maximum Amplitude Mode six frequencies and ranges can be chosen for the scanning of four antennas simultaneously. Rather than processing a multitude of frequencies and ranges, the signals from up to 24 antennas are spectrum analyzed simultaneously. This interlaced spectrum analysis is necessary if low Doppler frequencies have to be studied and the propagation conditions provide barely consistent data during the integration period. In contrast to the Fast Fourier Transform procedure used in most computers our Direct Discrete Fourier Transform algorithm with Hanning weighting by averaging the spectral phase function provides a continuous consistency check of the data. Fig. 5 shows that the complete multi-dimensional spectrum analysis is executed while the data are digitized as quadrature samples of the Intermediate Frequency (IF).

3. APPLICATION

A wide field of actual and possible applications for this instrument can be envisioned. The first system has been used to study the equatorial spread-F phenomenon in the ionosphere which heavily affects satellite-ground communication and navigation up to several GHz during many nights. Similar phenomena are studied in the aurora region of the ionosphere. We further applied the Digisonde to measurements of the sea-state of the oceans by direct and ionospherically reflected scatter and bistatic radio propagation experiments.

3.1 Gravity Waves and Fine Structure in the Ionosphere

Incidence angle measurement on spectrum-analyzed monostatic and bistatic pulse radio signals have shown that the structure of the ionosphere is multidimensional. Only rarely can the ionosphere be considered a plane reflector with an effective reflecting surface of the size and shape of the first Fresnel zone. In most cases the surface of constant electron density is two-dimensionally curved.

Therefore it is not always possible to find an area of perpendicularity but sometimes more than one perpendicular area can be found. The requirement of perpendicularity is related to the requirement of constant phase by the Fresnel condition which permits coherent integration of all contributions within the Fresnel zone.

Although spectrum analysis diminishes the area which can contribute to a coherently integrated signal it does simultaneously ease the condition of perpendicularity. Under the assumption of an almost constant drift of all irregular structure over the surface of constant electron density the spectrum analysis cuts out a band from this surface. The width of the band is determined by the drift speed and by the spectral resolution to comprise all reflection points with the same Doppler frequency. The length of the band is given by the Fresnel condition and is only extensive if the surface is perpendicular at least in the direction along the band (which is oriented perpendicular to the drift velocity).

But the perpendicularity requirement across the band is substantially relaxed because a change by one-half wavelength in phase path can be admitted for the width of the reflecting surface band if the roughness of the surface is sufficient to produce specular reflection. Under reasonable assumptions of drift speed and spectral resolution the reflecting surface band has a width in the order of 500 m (for E-region heights) which permits coherent addition of returns from all points of the surface if the inclination of the surface is less than 6° at 3 MHz probing frequency. Similar deviations from perpendicularity are permitted in the F-region since some of the changes in parameters compensate each other.

This Doppler-induced quasi-perpendicularity condition makes a large part of the constant electron density surface visible simultaneously and permits determination of its three-dimensional structure and motion at least in the time sequence of the "sky maps" which indicate the locations of all the reflection areas for all Doppler frequencies present at the sampling time (Fig. 6).

In transionospheric transmissions from satellites and reflections from targets fading, scintillation and positioning errors are caused by multimode propagation. Although the spacing of the different propagation paths is small compared with those of the vertical or oblique reflection case, effects of the multimode propagation might become important since higher accuracy for the incidence angle determination is required. Not always does the higher operation frequency compensate for the more stringent accuracy requirements. Therefore a good three-dimensional model is necessary for the correction of incidence angle errors. It can be produced by either multi-antenna complex spectral analysis of known satellite signals or by extrapolation from vertical sounding experiments not too far (<500 km) from the subionospheric point of the expected satellite propagation path.

3.2 Sea-Surface Waves

In March 1973 a Digisonde was used for the radio Doppler probing of the ocean surface in Eglin, Florida (Fig. 7). At a radio frequency around 6 MHz the energy scattered back showed Doppler offsets of ±0.25 Hz. These observations are in good agreement with the theoretically predicted Bragg scatter lines, but show many additional features and unexpected events. Nevertheless even the motion of the water parallel to the wave motion can be resolved.

3.3 Future Applications

The precision of the digital direct frequency synthesis, and the large antenna array permit accurate angular measurements for radio-interferometry in the frequency range of 20 to 40 MHz. Solar, planetary and galactic sources can be studied with this system. A combined acoustic-electromagnetic sounding of the non-ionized atmosphere is very promising. Acoustic scanning in medical and material research requires only little modification. Scanning of the sea-bottom (after conversion to acoustical waves) and underground radio communication and structure research should also be possible with this system.

We are developing a prototype for a topside digital ionosonde with very low output data rate (360 bits/sec). Such a system can also be used as a unmanned station on an island or a buoy in the ocean, on-board of a ship or in the arctic or antarctic. This sounder will broadcast the significant information of the preceding ionogram during the following ionogram to another Digisonde station in a distance of up to 3000 km for remote recording via an ionospheric propagation path (K. Bibl and B. W. Reinisch, 1976). This method has the advantage that the ionospheric conditions at the location of the ionospheric reflection can be monitored simultaneously. Data from several stations can be recorded simultaneously. Such a system can form an emergency radio propagation prediction network in case that satellite communication is interrupted. But for normal conditions simpler and more reliable satellite links can be used for collecting the low rate data from remote unmanned Digisonde stations.

ACKNOWLEDGEMENTS

The development of this instrument has been supported in part by DNA under Contract No. DNA001-77-C-0187 and by AFGL under Contract No. F19628-77-C-0024. The first on-line spectrum analyzer was integrated into a Digisonde for the U.S. Army Electronics Command under Contract No. DAAB07-75-C-A176.

REFERENCES

K. Bibl and B. W. Reinisch (1978), The Universal Digital Ionosonde, Radio Science, 13, No. 13, pp. 519-530.

K. Bibl (1974), Font of Digital, or Other, Characters and Method for Pattern Printing Thereof, U.S. Patent No. 3,810,095; DBP Patent No. 21 36 331.

J. Patenaude, K. Bibl, and B. W. Reinisch (1973), Direct Digital Graphics - The Display of Large Data Fields, American Laboratory, Sep. 73, pp. 95-101.

K. Bibl and B. W. Reinisch (1976), Method and Apparatus for Transferring Messages to and From Remote Locations, U.S. Patent No. 4,030,033.

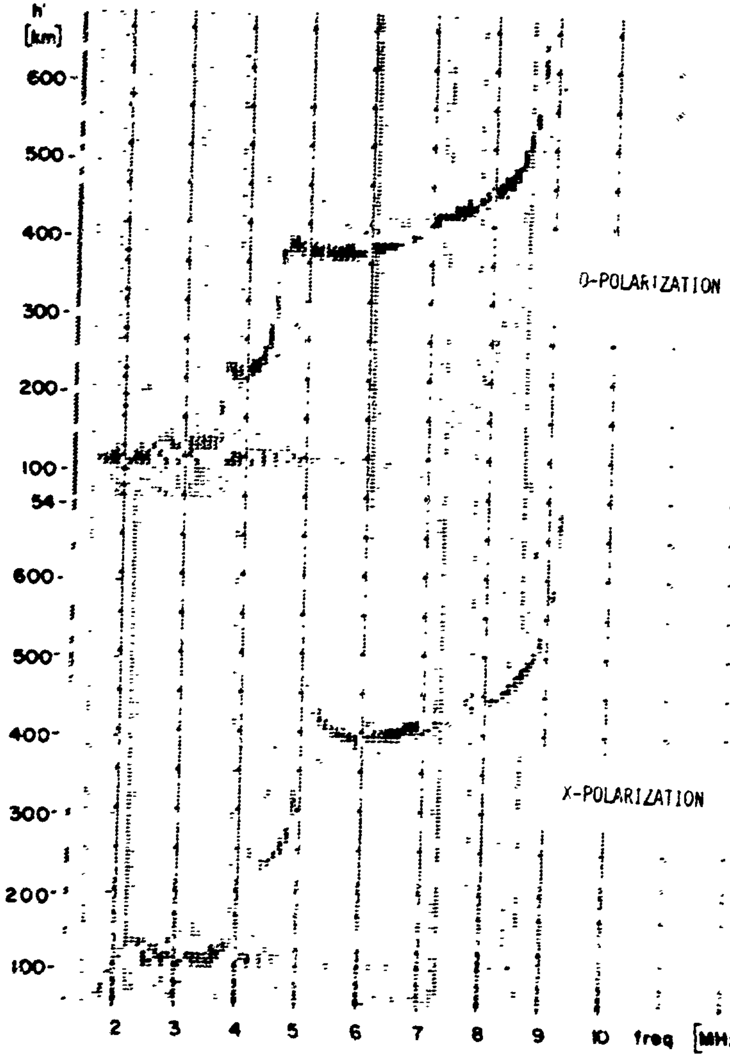


Fig. 1a SPLIT IONOGRAM

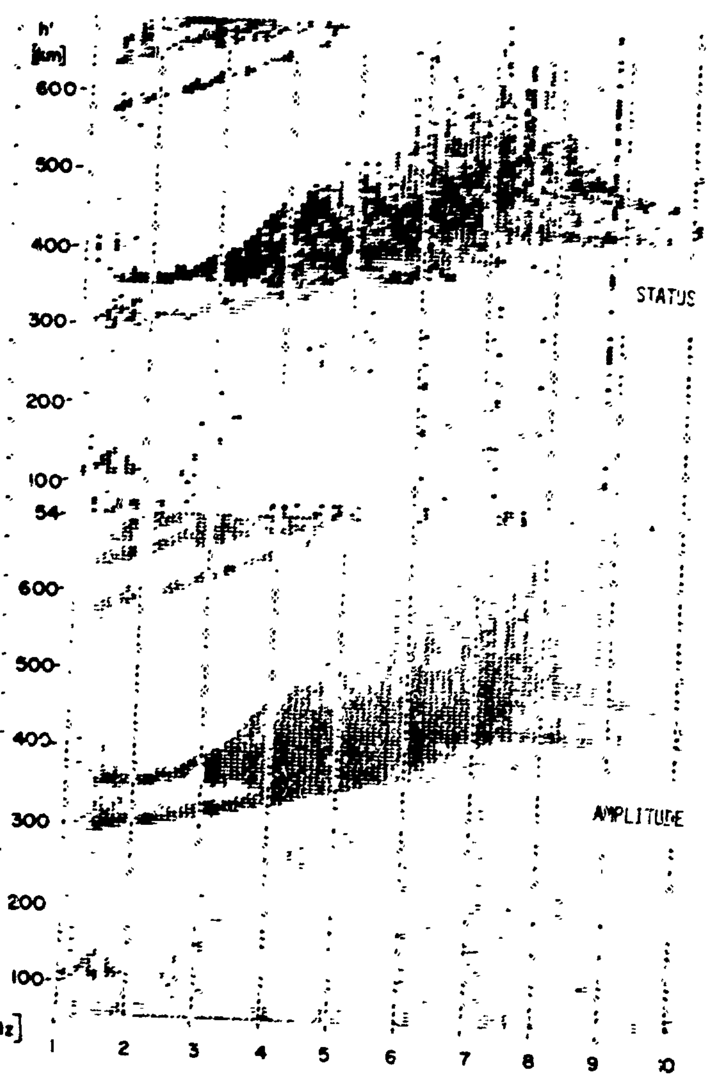


Fig. 1b AMPLITUDE, AZIMUTH AND DOPPLER

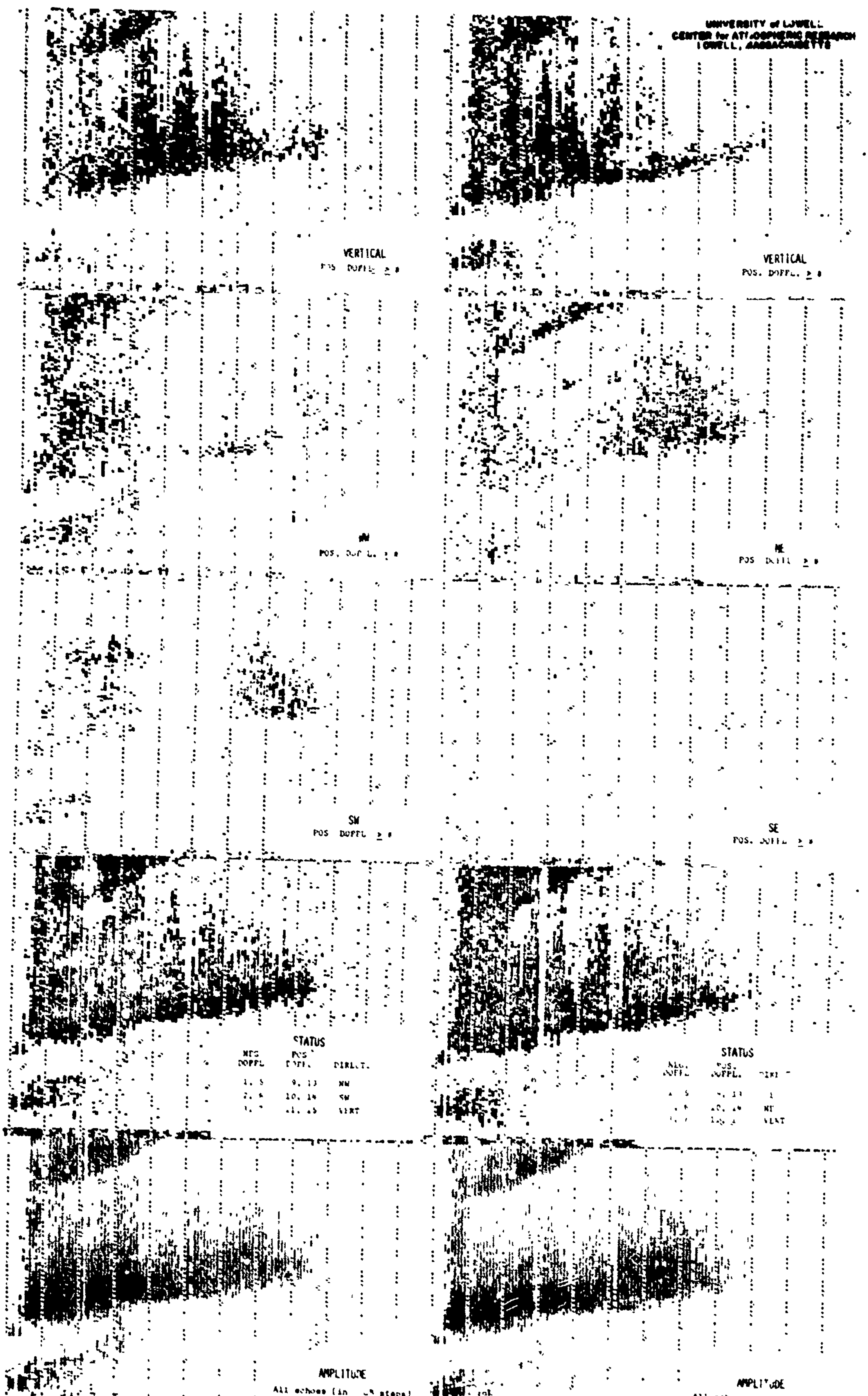


Fig. 2 COMPUTER-SPLIT DIRECTIONAL IONOGRAMS

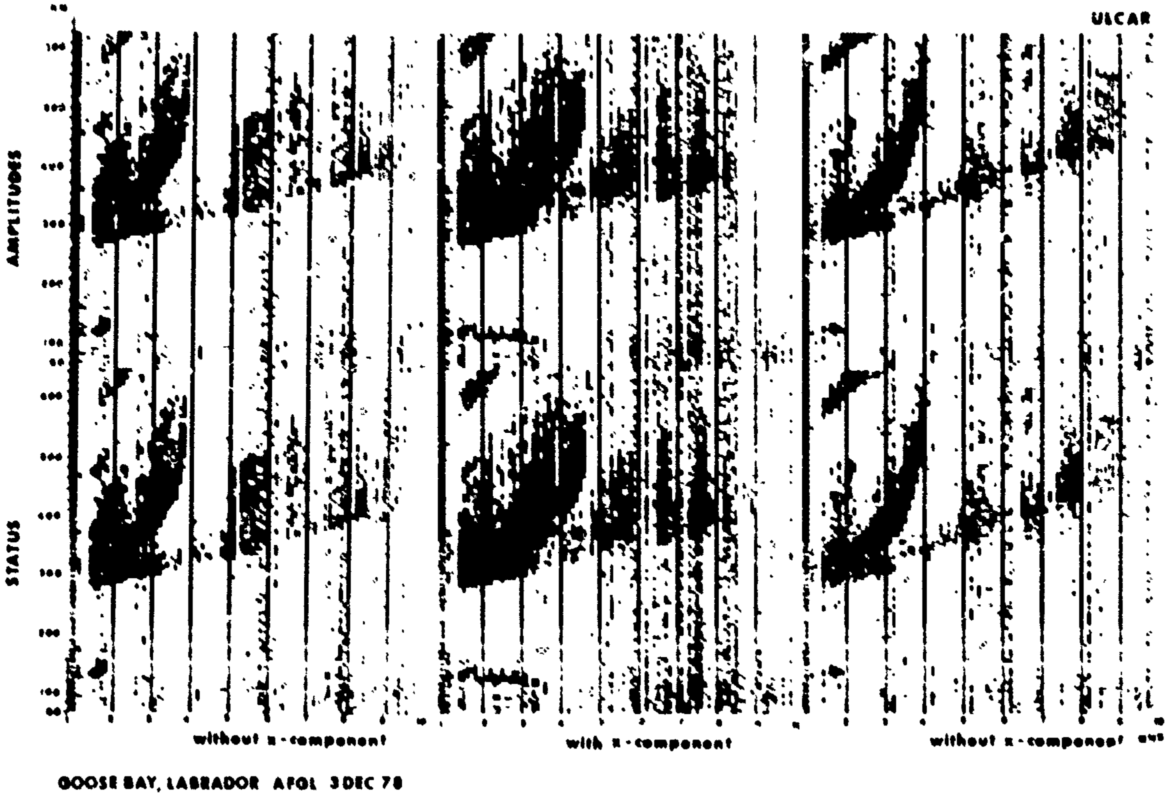
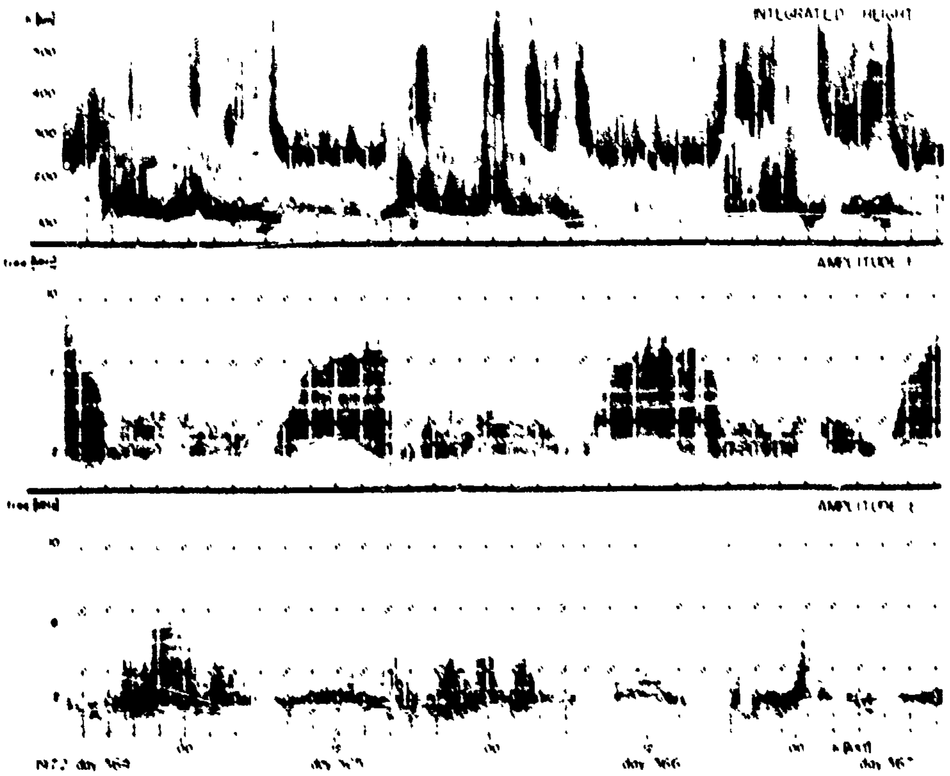
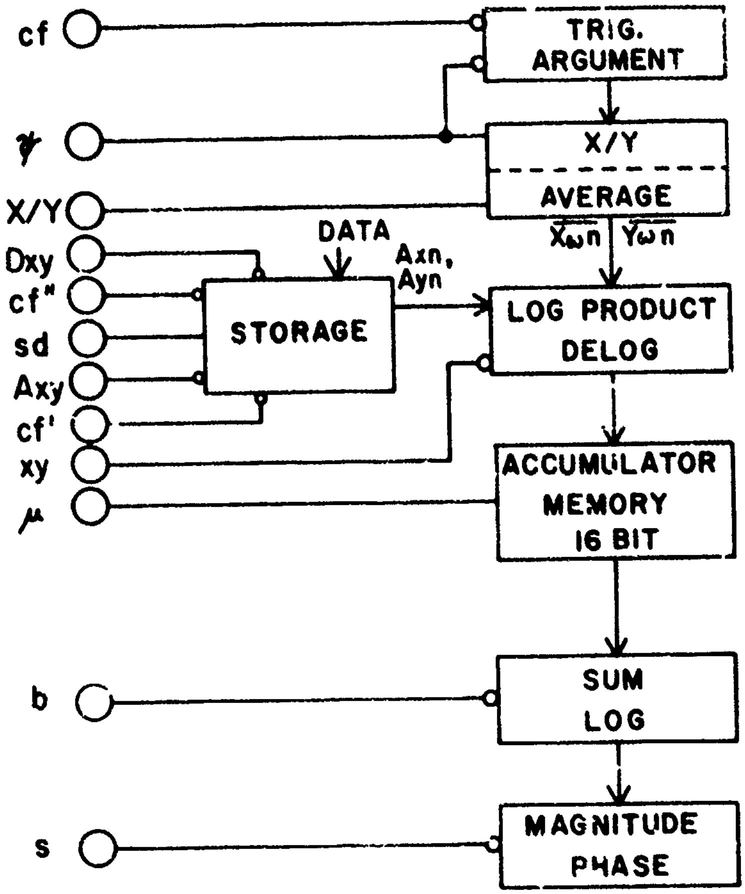


Fig. 3 DIGITAL SUPPRESSION OF THE EXTRAORDINARY COMPONENT



AIR6
VERTICAL IONOGRAMS
GOOSE BAY, LABRADOR

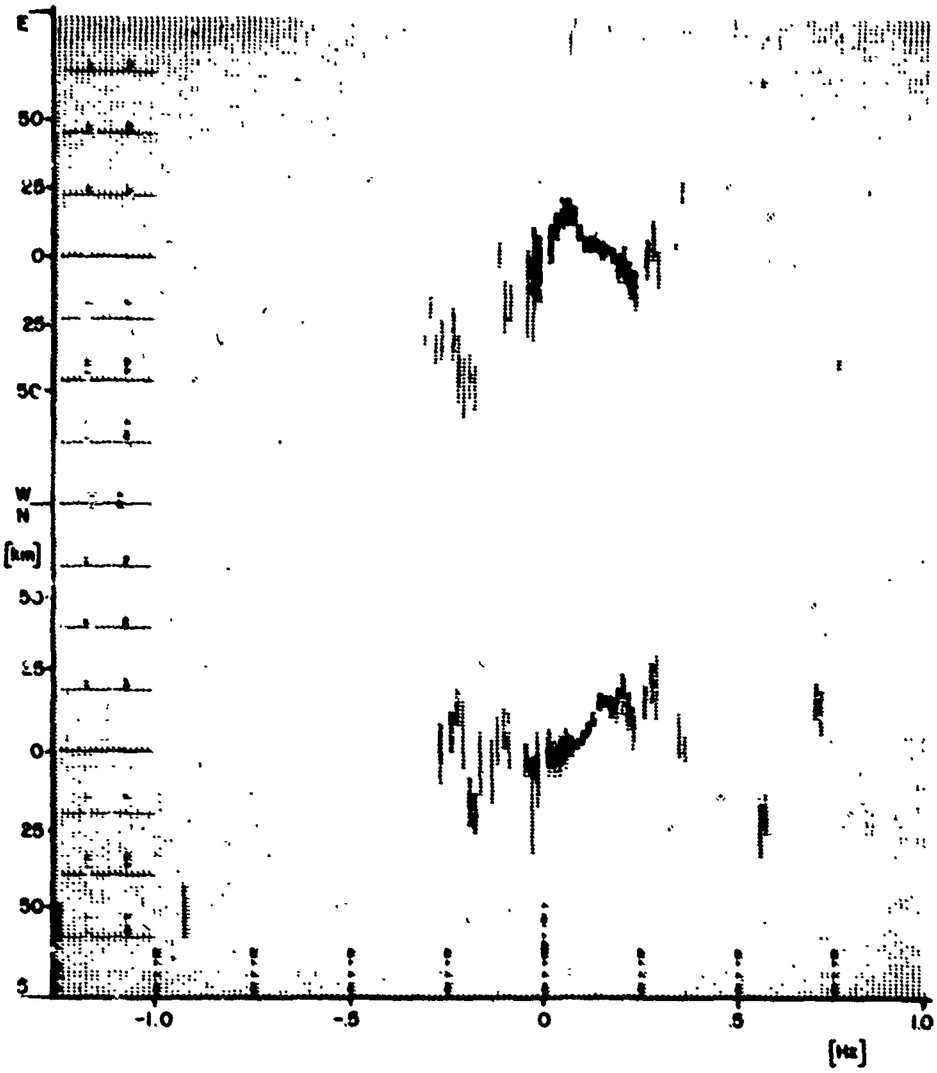
Fig. 4


cf
γ
X/Y
Dxy
cf"
sd
Axy
cf'
xy
μ
b
s
TRIG.
ARGUMENT
X/Y
AVERAGE
$\overline{X\omega n}$ $\overline{Y\omega n}$
DATA
Axn,
Ayn
STORAGE
LOG PRODUCT
DELOG
ACCUMULATOR
MEMORY
16 BIT
SUM
LOG
MAGNITUDE
PHASE

$$\sum^{n} (Axn \cdot \overline{X\omega n} + Ayn \cdot \overline{Y\omega n}) = R^{+}\omega$$
$$\sum^{n} (Ayn \cdot \overline{X\omega n} - Axn \cdot \overline{Y\omega n}) = I^{+}\omega$$
$$\sum^{n} (Axn \cdot \overline{X\omega n} - Ayn \cdot \overline{Y\omega n}) = R^{-}\omega$$
$$\sum^{n} (Ayn \cdot \overline{X\omega n} + Axn \cdot \overline{Y\omega n}) = I^{-}\omega$$

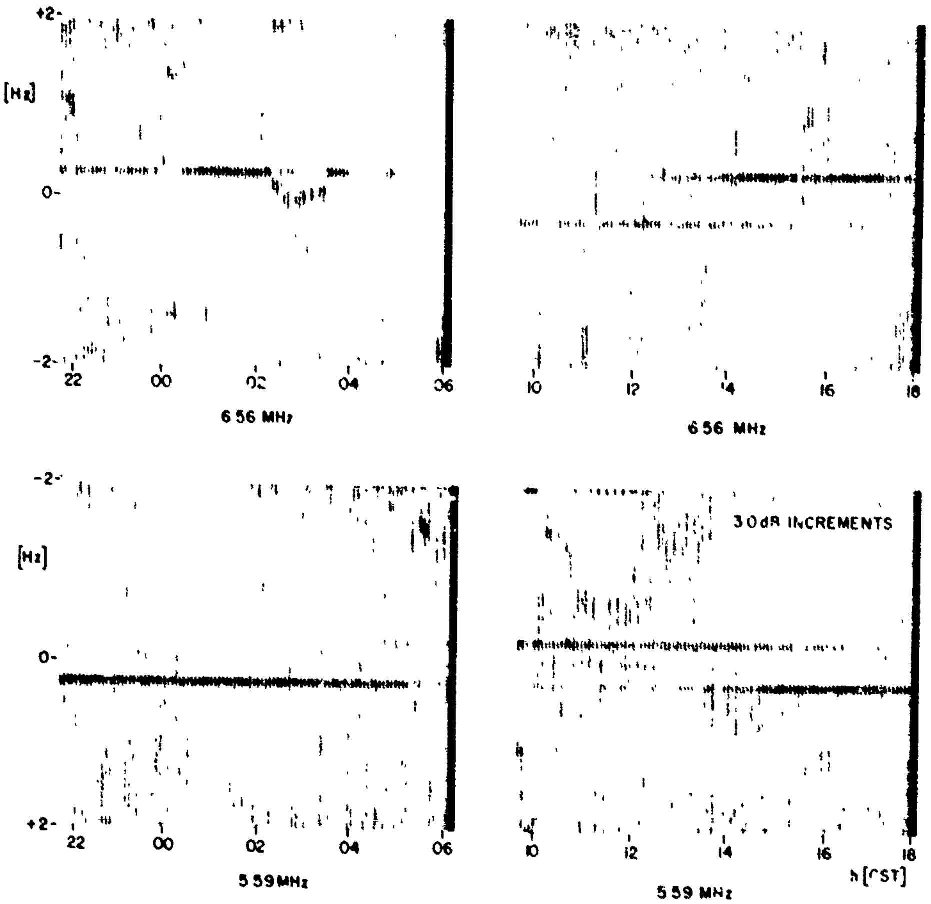
REAL-TIME SPECTRUM ANALYSIS

Fig. 5



DAASM DISPLAY
4 APRIL 72
EGLIN, FLA.
2.01 MHz-105 km

Fig. 6



SEASCATTER DOPPLER SPECTRUM
17-18 MARCH 1973
EGLIN, FLA

Fig. 7

NEW TECHNOLOGY TO IMPROVE HF CIRCUIT RELIABILITY AND AVAILABILITY FOR REMOTE REGIONS

J.S. Belrose and L.R. Bode
Communications Research Centre
Department of Communications
Ottawa, Ontario

ABSTRACT

New technology offers several approaches and opens the door to the development of improved HF-SSB radio communication systems. These can range from simple improvements, employing new equipment and techniques, to sophisticated integrated systems employing microprocessor technology to reduce operator workload through increased automation. The concept of an HF system with automatic connectivity to other radio systems, to private lines or to the switched public telephone system is an important feature of these new developments. Selective calling relieves the user of the need for radio listening skills and the fatigue caused by the manual monitoring of HF channels . . in addition it improves HF circuit reliability and extends circuit availability to 24 hours per day, for emergency use. This paper will describe R&D underway at the Communications Research Centre, Ottawa, and in Canadian Industry.

1. INTRODUCTION

High Frequency (HF) radio communications has been used for decades to provide economical, easily transportable non-line-of-sight communications. While there is an increasing trend away from HF to the higher frequency bands and to satellite communications because of increasing demand for higher information rates and message reliability, there are still important applications in which HF is expected to remain the principle communications technique for the foreseeable future. Some of these are: communications between naval vessels and between naval vessels and shore-based commanders; aircraft to ground communications; NATO military exercises, communications in remote and northern regions of the world; and important domestic communications in developing countries, where HF continues to provide a good service, and more important, HF communications is cost-effective for the low information rate and reliability required. In fact with today's emphasis on mobile self-contained independent point-to-point communications, HF communications is more essential than before. The IEE Conference on Recent Advances in HF Communication Systems and Techniques, held in London, 27-28 February, 1979 made this fact clear (c.f Ref. 1, 2).

A fundamental problem with HF communication links is that the radio channel is inherently variable, due to propagation vagaries and noise and interference, so much so that a particular frequency (or channel) may deteriorate during the period of sending the message, from being very good to useless. In general, these changes are only partly predictable. Users normally attempt to reduce the extent of fluctuations by adapting the operation of the transmitting and receiving equipments to varying channel conditions. Efficient adaptation requires flexible equipment and reliable and up-to-the-minute information on the cause of the degradation. Traditionally, HF links have been controlled by experienced radio operators. Unfortunately, the operator does not necessarily have adequate information for efficient control, but relies heavily on experience and subjective assessment of channel quality. Also the user is not necessarily able to make these decisions, nor should there be a need for him to do so (Ref. 3).

New technology offers several approaches and opens the door to the development of improved systems, which can range from simple to very complex systems employing microprocessor technology to reduce operator workload through increased communications automation. The concept of an HF system with automatic connectivity to other radio systems, to private lines or to the switched public telephone system is an important feature of such systems, since selective calling relieves the user of the need for radio listening skills and the fatigue caused by continuous manual monitoring of HF channels. In addition, for communications from northern and remote regions, where conventionally a radio operator monitors only during certain hours of the day, selective calling, without the assistance of a radio operator, means that the HF communications link is potentially available 24-hours a day for emergency use. Selective calling is particularly important when a mobile unit is operating with low battery powered equipment, so that the operator does not waste time and battery power uselessly calling when propagation over the link is disrupted. The user will know instantly that the channel is available or not available, and if not, he will call later when propagation conditions may have improved.

Direct automatic access to the switched public telephone system from portable or mobile radio transceivers is not new. Radio amateurs have been operating VHF autopatch repeaters for a number of years and sophisticated systems are being evaluated currently for general public avail; mobile radio telephones will probably utilize short range UHF radio frequencies employing a cellular type of communication system (Ref. 4). While Industry and in our interest Canadian Industry was responding to the anticipated requirements for repeater-autopatch or radio telephone interconnect (RTI) for FM radio systems operating at VHF and UHF frequencies, Communications Research Centre researchers saw a need for system studies and hardware development for systems employing HF-SSB suppressed carrier radio equipment, not only for the signal and control interface elements of the system, but also for a new kind of HF-SSB transceiver. The CRC has been engaged for some 4 years now in two projects that involve the development of new, improved and automated HF radio-telephone technology: (1) a rather sophisticated system intended to provide basic telephone and message service to small communities in the northern regions of Canada; and (2) a system which could be integrated with the first, or operated as a completely separate system, aimed at providing improved communications from the trail and from remote camps in northern and remote regions. The need for this latter communications requirement arose from a brief presented to the Federal Government, by the Northern Quebec Inuit Association in October, 1974, and it has led to the installation of a prototype system at Koartak on the northwest tip of Ungava Bay in Northern Quebec. This paper will comment on features and results of this project, and on research and development presently being done at CRC and in Canadian Industry.

2. TRAIL-REMOTE CAMP RADIO COMMUNICATIONS: A SYSTEMS APPROACH

HF-SSB trail radios have been used by Northern people for several years now, for communications from remote camps to their home base community. Battery operated 10 watt PEP transceivers are employed, using either the Spilsbury Communications SBX-11A or the Marconi CP3. Sometimes these radios are used with short whip antennas, which at frequencies of 3-4 MHz, have typical radiation efficiencies of -20 dBd (dB with respect to a dipole); consequently, communications is poor since negligible power is radiated and the launch angle is not suitable for short to medium distance links. Furthermore, communications is between the portable unit and a similar, or higher power base station radio-telephone. That is, an operator in the community must listen for a call from the remote camp. It seemed that there was a need for improvement, not only to the radio equipment, and the associated component elements such as antennas, batteries, etc., but also to the whole system concept.

Because of the success achieved with VHF transceivers operating through a remote repeater in land-mobile communications in cities, and the experience of radio amateurs who employ repeaters for communications between mobiles and from distant locations, such as hunting and fishing camps, etc., VHF-FM repeater systems seemed to be a good approach for trail radio. Community VHF radios were set up in peoples' houses, in places where they worked, etc. for communications through a remote repeater, with VHF portables operated on the trail and from remote camps (Ref. 5). The communications requirement however was for portable communications from distances greater than could be reached by VHF equipment, even when these were operated through a remote repeater which greatly extended their range, to 70-100 km. HF-SSB transceivers would still be needed for communications at the greater distances, 60-200 km. HF radio telephones should, therefore, be a part of the communications system, and the VHF community radio should be able to hear a call from the trail or from a remote camp whether the call originates from an HF or VHF portable.

A communications system was designed and developed, employing tone signalling for access, and installed in September 1977 at Koartak in Northern Quebec for use in a pilot project by northern people for their own natural communication needs. The system operation is illustrated in Figures 1 and 2. In Figure 1, communication is via VHF, from a VHF portable unit to a VHF community radio employing a remote battery operated repeater. Communications between an HF portable unit and a VHF community radio is illustrated in Figure 2. The HF portable unit must "access" the system, by tone signalling. When he calls his signal is received by the HF base station receiver, relayed by a UHF link to the remote VHF repeater, so the community VHF radio hears the call even though it originated from an HF portable. An important feature of the system, was that it had to be initially accessed by sending a dual-tone-multi-frequency (DTMF) burst over the link. If this signal were correctly received by the base station receiver-decoder, a 1 second tone burst was sent back over the link by the base station transmitter, so that the mobile knew he had achieved interconnect. He could therefore press his microphone button and call.

The communication system described above is an all radio one, and since no base station radio operator is involved, it is potentially available 24 hours per day. With the deployment of a number of VHF community radios in peoples' houses, instead of one radio under operator control, there is less chance of an emergency call going unheard. Clearly, since tone signalling is employed as an integral part of the system, with appropriate design, it could be possible for the mobile to call his home directly, through a radio link interconnection to the community telephone system. Some of these ideas, and various hardware developments through contract with Canadian Industry are described below.

3. SUB-SYSTEM CHARACTERISTICS

3.1 Antennas for Base Station and Trail Use

Various antennas were considered for use at the HF base station, in remote camps and on the trails such as the horizontal dipole, the inverted-V, various forms of full-wave delta-loops, a horizontal full-wave quad-loop and, for trail radio application, short centre-loaded whips. The requirements for the antenna system were: high angle radiation (for short distance communications); adaptability for multi-frequency operation; non-directionality in the azimuthal plane, inexpensive; and, in particular, for portable use, the antenna had to be lightweight and easily erectable.

The inverted-V dipole antenna was the antenna type that best met these various requirements. Essentially, it is a half-wave dipole antenna with drooping ends mounted on a single centre support mast. The ends of the dipole are attached by insulated cable to anchors at ground level on opposite sides of the mast (c.f. the sketch in Figure 3). While this antenna is very popular with radio amateurs, it seems to have had limited, if any, use in commercial applications.

As is evident in Figures 3 and 4, the important parameters are the height of tower (measured in fractions of a wavelength), and the included angle Λ (degrees) between the dipole arms. Obviously, if $\Lambda = 180^0$, the inverted-V becomes a horizontal dipole, and for $\Lambda = 0^0$ we have a very inefficient radiator. For antenna mast heights, h, of $0.1 - 0.25\lambda$ the input impedance of the inverted-V is about 50 ohms, and a resonant antenna can therefore be fed directly with 50 ohm coaxial cable. However, it is necessary to utilize a 1:1 balun transformer for unbalanced to balanced impedance transformation

The antenna radiates very well at high elevation angles, and the azimuthal response of the antenna for communications is essentially non-directional as will be described below. The various curves in Figure 3 show the vertical polar diagrams in the equatorial and meridian planes (the plane broadside to the antenna and the plane containing the dipole) as measured at a scale model frequency of 200 MHz on an antenna pattern range. The radial scale is linear in field-strength. The inverted-V radiates a horizontally polarized signal in the plane at a right angle to the dipole, and a vertically polarized signal in the plane in line with the dipole. The azimuthal patterns (not shown) are typical figure-8 patterns, which, naturally, are orthogonally directed for the two polarizations Thus, on transmit, the antenna radiates either horizontal or vertical polarization, dependent on direction of propagation. On receive the down-coming sky-wave at middle to high latitudes is circularly or elliptically polarized, and so the wave couples well into the receive antenna regardless of its orientation.

Other advantages of the inverted-V are: several antennas can be fed in parallel, with little interaction if the frequencies are non-harmonically related. If two frequencies are used, the antennas should be arranged orthogonal to each other; for more than two frequencies, the antennas would be arranged symmetrically about the supporting mast.

For portable application, a sectional tripod arrangement consisting of a bundle of 1.8 meter lengths of 2.54 cm diameter aluminum tubing for the centre support was devised (see Figs. 5 and 6). The height of the centre support could be 1.8 meters or 5.5 metres. When elevated at its centre to a height of only 1.8 meters above poorly conducting arctic terrain, the inverted-V dipole antenna was found to provide usable communications over paths up to a few hundred kilometers with 10 watt PEP transceivers. For portable use, the antenna is wound on bobbins and the user unrolls the correct resonant length to a coloured marker on the wire corresponding to the channel marker on his radio (see Figure 8)

A final comment: the elevation pattern for vertical polarization (in the plane containing the inverted-V) is very sensitive to any unbalance in the dipole, especially if the tower height is near to a resonant height of $\lambda/4$, which is typical for installations employing frequencies near the lower end of the HF band. Any unbalance in the dipole results in currents flowing in the tower, or in the sheath of the coaxial feeder-cable, which results in the pattern being non-symmetrical, c.f. Figure 4(b), curve 1. A balun was used for the model measurements, however a balanced feed was apparently not exactly achieved. The diagrams in Figure 3(b) have been corrected for this unbalance. The ripple on the patterns in Figure 4 are due to a range imperfection. Patterns were not measured for $h/\lambda > 0.25$, since for the intended application (short to medium distance communications) the resulting polar diagrams would be unsatisfactory.

3.2 HF-SSB Transceiver

A new HF-SSB transceiver is being developed as an offshoot of the trail remote camp radio project. Special features of the transceiver are: (1) modular construction so the 10 or 20 watt trail radio can be attached to a battery pack at the rear of the radio for portable use, or a 100 watt linear amplifier for use in camp plugs in place of the battery pack; (2) digital channel control is utilized, so that the 100 watt linear can be attached without motors or switch extensions; (3) automatic load control; i.e., the transmitter drive adjusts to mismatched antennas and avoids distortion; (4) Speech processing for increased talk-power that works automatically on louder speech (the transmitter operates as a linear amplifier on low level voice); (5) simplified channel programming, no wires to solder or filters to tune; (6) built-in voice operated squelch; (7) noise suppression to combat ignition interference; (8) protection against over-voltage or reverse polarity power supply; and (9) insulated crystal oven for reduced power consumption on the prototype transceiver. Development work on a temperature compensated crystal oscillator, with channel crystal derived frequencies locked to the master oscillator or synthesized from it is under development, since a crystal oven is not power conservative for battery operated equipment.

The prototype trail-remote radio is still under development, however the contractor, Spilsbury Communications, Vancouver, has developed a new series of SSB tranceivers for the domestic and export market utilizing technology acquired as a result of their work on the trail radio transceiver. This will a 100 watt PEP transceiver, designed for land-mobile and land-fixed use, operating from a 12 volt vehicle battery (or equivalent). A photograph of the new transceiver (Spilsbury Communications Model SBX-100) is shown in Figure 9.

3.3 SSB Tone Decoder

The problem with using tone signalling on SSB with suppressed carrier has been that during detection, the audio frequencies are shifted if there is any difference between the transmitter carrier frequency and the receiver local oscillator frequency. A conventional tone signalling system would require a wide band width to accept this difference, and would thus suffer from lack of reliability and poor noise immunity. The SSB tone signalling system developed by Challenger Electronics, Vancouver, utilizes a pilot tone mixed with an information tone; the pilot tone is 6 db higher than the information tone. If the audio tones are shifted in frequency, the difference frequency between the pilot and information tone will remain constant and may be detected in the conventional manner. The pilot tone frequency was chosen to be 2199 Hz. The various frequencies for the information tone for a 12-digit or a 16-digit signalling system are given in a table in Appendix 1. It will be noted that all signalling frequencies are >1315 Hz and all difference frequencies are <885 Hz; therefore, a hi-pass filter ahead of the decoder reduces noise and interference at the detected frequencies. The block diagram in Figure 10 illustrates how the system operates. The audio output is first applied to the hi-pass filter which removes all products of mixing leaving only the pilot and information tones. The compression amplifier assures constant amplitude to the envelope detector. The resultant signal which contains the original tones and the sum and difference frequencies is applied to the lo-pass filter. This filter removes the original frequencies and the sum frequencies leaving only the difference frequencies to be amplified by the final compression amplifier. The decoder is a conventional active bandpass filter type, with the time constant set at about 250 ms to minimize falsing. After detection and verification, conventional AT&T touch-tone digits® are regenerated so the output tones are compatible with telephone signalling equipment. The DTMF SSB signalling touch-tone® regenerator system developed for CRC by Challenger is their Model BD-0802.

3.4 SSB-DTMF Mobile Encoder

It was, of course, necessary to develop a special mobile tone encoder for use with this system. The Challenger Model ME-0802 is a dual tone multi-frequency encoder, generating synthesized frequencies in the range 1524 to 1878 Hz, as well as the pilot tone at 2199 Hz. A standard TV colour burst quartz crystal (frequency 3.579545 MHz) provides a stabilize reference for the synthesized tones. The encoder provides a frequency stability of 0.001% over the operating range −40°C to +70°C. The mobile radio telephone interfaced with the encoder is illustrated in block diagram in Figure 11 and a trail radio equipment is shown in Figure 12. The HF-SSB radio telephone is a Spilsbury Communications, Vancouver, Type SBX-11 a 10 watt PEP portable transceiver. The modified transceiver operates off the Ni-cad batteries.

3.5 Speech Processing

The amplitude compander, known since the 1930s but only recently made practical and economical (Ref. 6), provides some significant advantages. Compressing the audio signal prior to modulation to achieve more efficient use of the transmitter power is widely used by radio amateurs. However, by compressing only the amplitude peaks, the background noise increases relative to the peaks. A significant advantage can be achieved by expanding the compressed audio in the receiver. Using full amplitude companding a 12 to 15 db improvement in signal-to-noise ratio is possible (Ref. 7). When only compression is used, the improvement is not so dramatic, about 6 - 8 db improvement (Ref. 8, 9 and 10).

The amplitude compander will improve the SNR, but only if the received signal is a few db above the noise. When the signal drops into the noise, the expander will not operate properly, since there is insufficient signal for use as a reference.

The prototype trail radio transceiver under development employs voice compression on loud speech, the more sophisticated system under development (see Section 5) employs digital signalling and a microprocessor controlled compander. Radio amateurs are currently experimenting with amplitude as well as frequency companding, employing an equipment manufactured by Henry Radio, Los Angeles, California, their VBC Model 3000, using a circuit designed by Harris and Cleveland (Ref. 11 and 12). Although initial tests to date have been successful, many more tests by a large number of radio communicators under varying conditions are desired, since only then will a sufficient data base be established as to the usefulness of the various systems developed and proposed, for the guidance of future improvements.

4. THE SYSTEM CONCEPT FOR AUTOMATIC RADIO TELEPHONE INTERCONNECT (RTI)

The block diagram in Figure 13 illustrates the system concept for a prototype system yet to be evaluated by field-trial. The most difficult part of the system turned out to be the transmit/receive switch at the HF-SSB base station. In the all-radio system, that is in operational use at Koartak (Figure 2), the switching is done by the carrier of the VHF community radio, i.e., by a carrier operated switch (COS). With VHF-FM, this provides reliable switching down to levels near to signal threshold.

Our first idea was to give control of the switching to the called telephone party, after the initial functions of access, and dialling had been completed. Thus when the called telephone party picked up his phone, his voice would activate a voice-operated transmit (VOX) switch. However, the balance of the hybrid, when connected to actual telephone lines, provided only 10 - 20 db isolation between the received radio signal and the speech of the called telephone party, and since automatic gain control (AGC) had to be used to hold levels constant, this degree of isolation was insufficient. Noise and interference received by the HF-SSB base station receiver actuated the VOX control which, of course, made the system malfunction.

The system currently adopted utilizes dual control, that is, VOX operation on the speech of the called telephone party is still employed, but when the mobile speaks, a 1 second DTMF tone burst is sent over the link which sets a gate, and when he finishes speaking, and lets his microphone button go, another DTMF tone burst is sent over the link to open the gate. The mobile transceiver therefore has two modes of operation. In one mode for ordinary HF mobile to HF base station operator control communications or for the initial functions of access and dialling, the transmitter turns on and off under the control of the microphone button. No tone signalling is employed, except those tones using the keyboard encoder. When the mobile is talking or the autopatch, he presees a push-to-hold switch on his transceiver, so that each time he presses his microphone button or releases it, a 1 second DTMF tone burst is transmitted. These beeps will likely not be annoying to the user, in fact, the user could come to like to hear them because when the called telephone party hears the second beep, he knows he should talk.

The various signalling frequencies on the telephone line provide the initial control. The mobile dials * to access the system. If this is properly received, the off hook flip/flop closes the off hook relay in the voice station coupler (a unit supplied by the telephone company). The dial tone detect circuitry detects the dial tone, and turns on the HF transmitter briefly, so when the mobile lets go of his microphone button he hears dial tone. When the base station transmitter is silent, the mobile presses his microphone button again and touches out the telephone number he wishes to call. He lets go of his microphone button. Dependent on the telephone exchange, the regenerated AT&T touch tones ® can be used to dial directly, or these digits are converted to binary code and entered into a first-in-first-out (FIFO) memory, and sequentially dialled out by a dial pulse converter. If a successful dialling has been achieved, a ringing signal appears on the line. If a busy signal is detected, the system hangs itself up automatically. Each ringing signal turns on the HF base station transmitter so the mobile can hear the phone ringing. If after 10 rings, there is no answer the RTI control hangs up the phone automatically and reverts the system to standby. If the called party answers, the voice operated VOX turns on the transmitter and the mobile hears him speaking. Operation continues as described above, and when complete, the mobile sends the digit # which hangs up the phone. If propagation conditions change before he hangs up the phone and he is unable to do so, the dial tone which appears on the line 30 seconds after the called party has hung up is detected and used to perform the automatic hang up and reset to standby operation.

This prototype system is currently under evaluation by field trial. It has been designed to demonstrate the feasibility of RTI from an HF-SSB mobile radio telephone. An operational system would require further development and refinement, to incorporate user identification, etc.

5. A DIGITAL HF RADIO TELEPHONE SYSTEM: ONE OF GREATER SOPHISTICATION

The HF radio telephone system described in some detail above is basically a very simple one. Manual channel selection is employed, analogue tone signalling is utilized, and speech processing while employed is simply amplitude compression. Also supervisory functions as required by the telephone company, such as user identification, etc. was not a part of the development. This was because telephone compatible system studies, and a more sophisticated system was under development as a separate project. This system will employ digital signalling. The heart of it is the controller/interface unit (CIU) that provides the means of interfacing an HF-SSB simplex radio link to the telephone network without the aid of an operator. The interface unit monitors the operational status of the radio equipment and the associated loop from the switchboard or telephone set. It generates and receives supervisory signalling such as dial tones or busy signals, and ensures that the calling and destination subscribers also receive the appropriate signals in the course of setting up a call. Most importantly it automatically tests and evaluates the available frequency channels to determine the optimum channel for use in that link at that moment. The controller employs a microprocessor to implement the necessary control functions, so that the system software can be tailored to suit the requirements of a particular user.

The system also will utilize synchronized compression and expansion (Syncompex) of speech, to greatly improve the quality of speech obtained over a HF-SSB radio link. Syncompex operates like Lincompex, except that digital signalling rather than analogue signalling is employed. Since the control interface unit is digital, it is clear that a digital data link is required, and a digital control channel, which will accompany the voice transmissions is provided thereby linking the speech compressor in the transmitter to the complementary expander in the receiver. Selective fading due to multipath propagation characteristics of short range HF propagation has been found to be a problem with an early prototype for the Syncompex voice processor, but this has been overcome by employing inband frequency diversity (Chow, private communications).

A new HF-SSB transceiver rated at 100 watts CW has been developed under this project, by Nautical Electronic Laboratories (Nautel), Hackett's Cove, Nova Scotia, Model NXH100TR. Features of this transceiver include all solid-state construction with redundant circuitry to offer extended mean time between failures. In the power amplifier stage, a tree network of solid-state amplifiers and power combiners is employed to offer the feature of graceful degradation. If a power amplifier component fails, the remaining components can continue to operate with only a proportional reduction in output power. Other features of the Nautel transceiver include crystal stabilized, channelized, frequency agile tuning for rapid channel changes and digital control.

The Spilsbury SBX-100 is intended for use in motor vehicles, trucks, tractors, snowmobiles as well as base camps, watch towers and remote sites; and, although it can be interchanged as a mobile or base station, the Nautel NXH100TR is designed for base station use, where reliable communication is required, and the additional cost of the sophistication provided can be afforded.

6. CONCLUSIONS

The concept of an HF-SSB communications system with automatic interconnectivity is feasible, and offers a decided advantage, for the untrained user as well as time saving features for the experienced operator. The application of this feature to HF equipment under microprocessor control offers new dimensions to HF communications, and provides a basis of further advancement in automated and adaptive HF radio systems. This paper has described some research and development that has been and is being done at CRC and through collaboration with Canadian Industry. The modular concept which allows the HF system to be optimized for the particular application and the system complexity and cost desired by the customer is an important feature of today's radio equipment.

ACKNOWLEDGEMENTS

The authors wish to acknowledge discussions with their colleagues, G.W. Irvine, S.M. Chow and B.D. McLarnon.

REFFRENCES

1. Hebert, P. and A. Walthe, "Future Naval HF Communications", Conference on Recent Advances in HF Communication Systems and Techniques, IEE, London, 27-28 February, 1979.

2. Harmon, J.V., "Automatic Connectivity for HF Voice and Low Speed Data Communications", Ibid., 1979.

3. Steven, E.E., "The CHEC Sounding System", Ionospheric Radio Waves (Ed. Folkestad), pp. 359-369, Plenum Press, N.Y., 1968.

4. Young, W.R., "Advanced Mobile Phone Service: Introduction, Background and Objectives", The Bell System Technical Journal, 58, 1-15, January 1979 (this paper serves to introduce a series of ten papers in the same issue).

5. Belrose, J.S., "Trail and Remote Camp Radio", Research and Development 1977-78, Department of Communications Publication, Ottawa, 19[illegible].

6. Jung, W.G., 'Gain Control IC for Audio Signal Processing", Ham Radio, 47-53, July, 1977.

7. Lusignan, B., "Single-Sideband Transmission for Land Mobile Radio", IEEE Spectrum, 33-37, July, 1978.

8. The Radio Amateurs Handbook, ARRL Publication, Newington, Conn., pp. 391-397, 1978.

9. Tong, D.A., "AF and RF Clipping for Speech Processing", Wireless World, 79-82, February, 1975.

10. Tong, D.A., "Distortion Levels in RF Clipping", Ibid., pp. 77-81, October, 1976.

11. Harris, R.W. and J.F. Cleveland, "A Baseband Communications System, Pt. I", QST, 11-16, November, 1978.

12. Harris, R.W. and J.F. Cleveland, Ibid., Pt. 11, 14-21, December, 1978.

APPENDIX 1

SSB TONE SIGNALLING

FREQUENCY TABLE

PILOT TONE 2199 HZ

DIGIT	INFORMATION TONE (HZ)	DIFFERENCE FREQUENCY TONE (HZ)
1	1878	321
2	1856	343
3	1832	367
4	1806	393
5	1779	420
6	1749	450
7	1718	481
8	1684	515
9	1648	551
0	1609	590
*	1568	631
#	1524	675
A	1478	722
B	1427	773
C	1373	827
D	1315	885

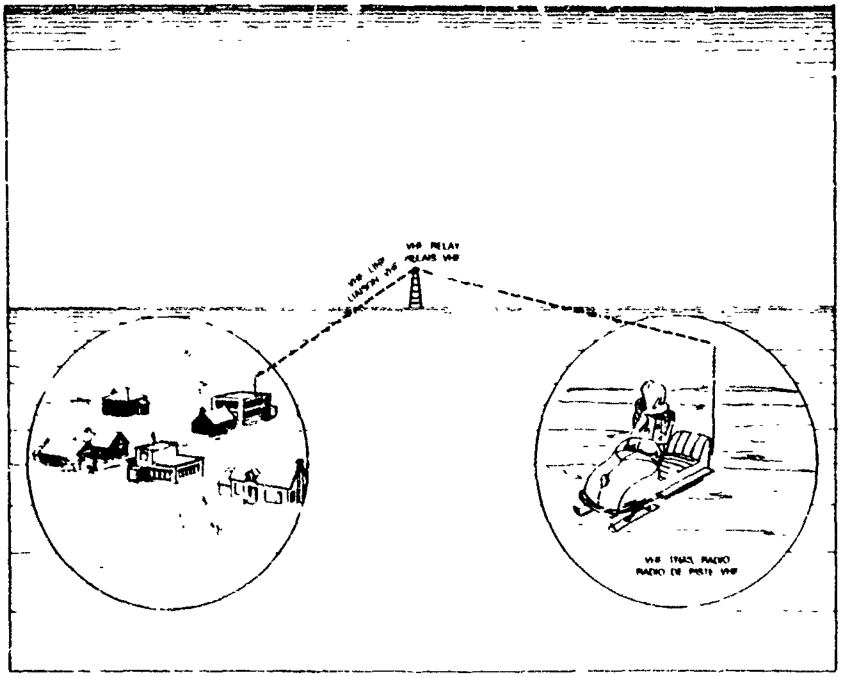

VHF RELAY
RELAIS VHF
VHF LINK
LIAISON VHF
VHF TRAIL RADIO
RADIO DE PISTE VHF

FIGURE 1. Sketch illustrating communication between a VHF portable and a VHF community radio, by relay through a remote repeater.

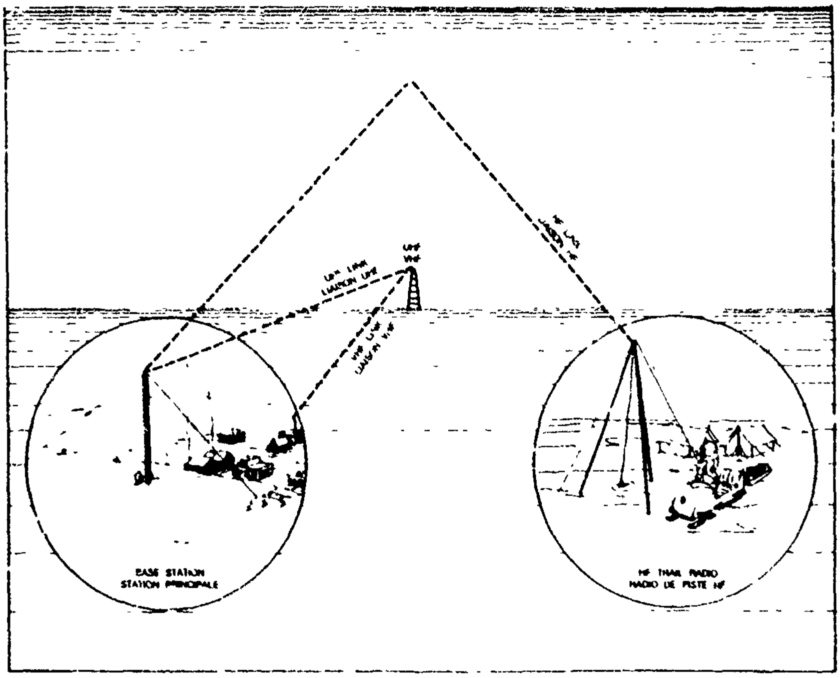

UHF
VHF
UHF LINK
LIAISON UHF
VHF LINK
LIAISON VHF
HF LINK
LIAISON HF
BASE STATION
STATION PRINCIPALE
HF TRAIL RADIO
RADIO DE PISTE HF

FIGURE 2. Sketch illustrating communications between an HF portable and a VHF community radio, employing an HF base station, a UHF link and a remote repeater.

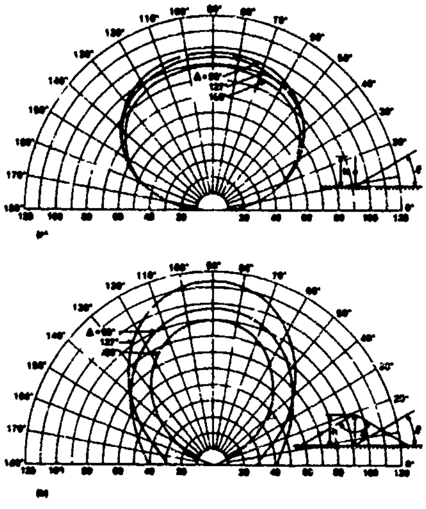
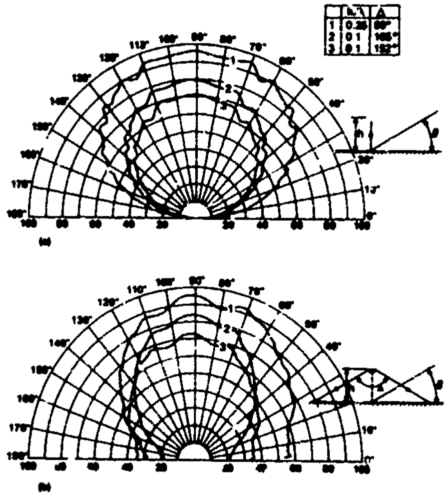


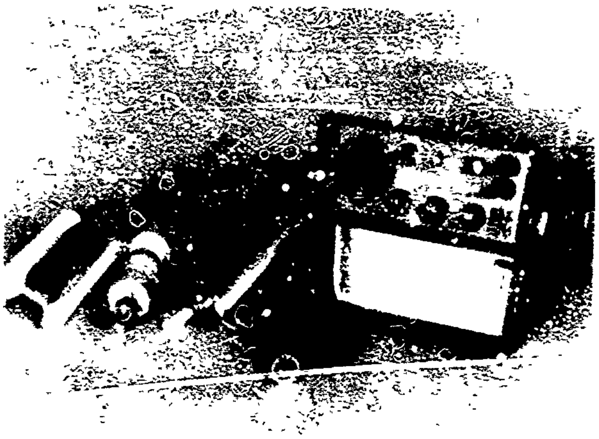
FIGURE 3. Measured vertical plane polar diagrams for inverted-V dipole antennas, measured in two orthogonal planes and polarizations. The height of the supporting tower h = λ/4.

FIGURE 4. Measured vertical plane polar diagrams for inverted-V dipole antennas measured in two orthogonal planes and polarizations, for two heights of supporting tower.

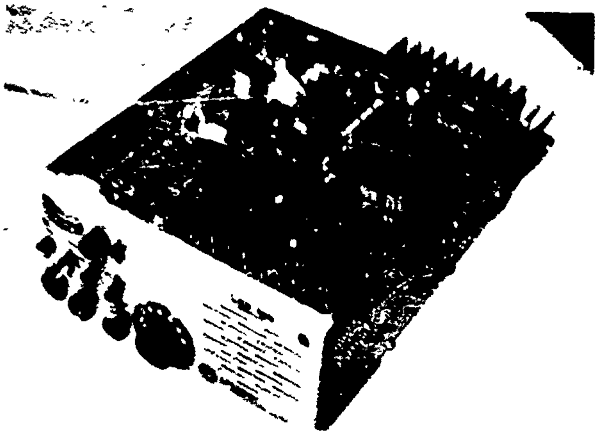


FIGURES 5 and 6. Photographs showing preparation and erection by Inuit people of a portable inverted-V dipole antenna.

[illegible]

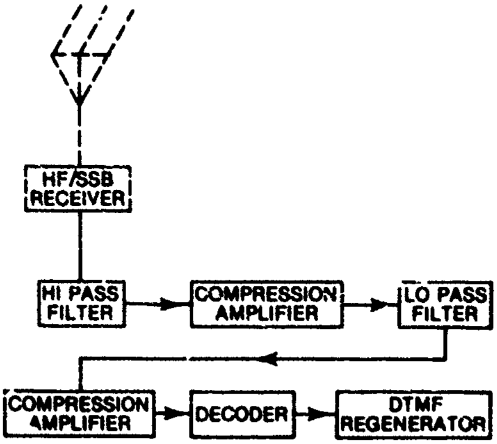


[illegible]



[illegible]

FIGURE 10. Block diagram illustrating the principle of operation for SSB tone signalling decoder.


HF/SSB RECEIVER
HI PASS FILTER
COMPRESSION AMPLIFIER
LO PASS FILTER
COMPRESSION AMPLIFIER
DECODER
DTMF REGENERATOR

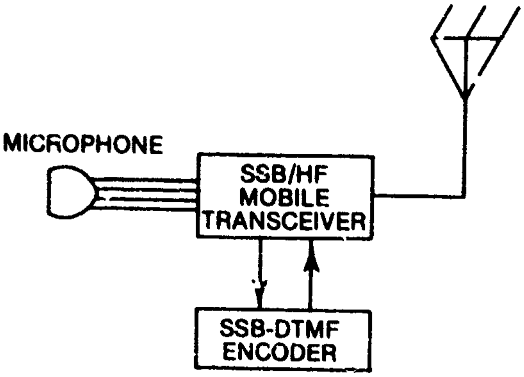
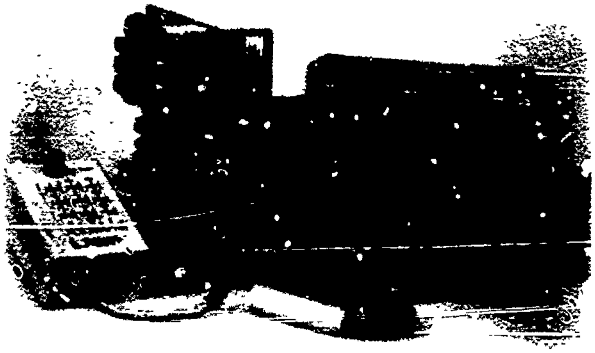

MICROPHONE
SSB/HF MOBILE TRANSCEIVER
SSB-DTMF ENCODER

FIGURE 11. Block diagram illustrating an HF transceiver interfaced with a DTMF encoder.

FIGURE 12. Photograph showing a portable trail radio (Spilsbury Type SBX-11), a SSB-DTMF encoder (Challenger Model ME-0802 , Ni-Cad batteries and carrying case for transceiver and batteries.



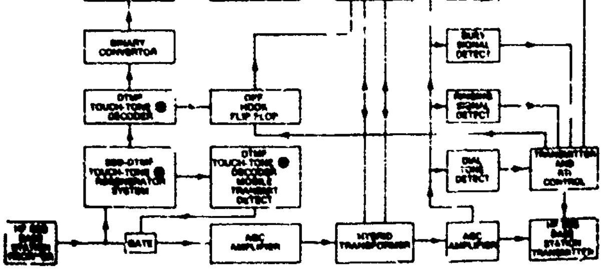

TELEPHONE LINE
FIFO MEMORY
DIAL PULSE CONVERTOR
VOICE STATION COUPLER
VOICE OPERATED TRANSMIT (VOT)
BINARY CONVERTOR
BUSY SIGNAL DETECT
DTMF TOUCH-TONE DECODER
OFF HOOK FLIP FLOP
RINGING SIGNAL DETECT
SSB-DTMF TOUCH-TONE REGENERATOR SYSTEM
DTMF TOUCH-TONE DECODER MOBILE TRANSMIT DETECT
DIAL TONE DETECT
TRANSMITTER AND RX CONTROL
HF SSB BASE STATION RECEIVER
GATE
AGC AMPLIFIER
HYBRID TRANSFORMER
AGC AMPLIFIER
HF SSB BASE STATION TRANSMITTER

FIGURE 13. Block diagram illustrating the principle of operation for an HF base station radio telephone interconnect system (prototype presently being evaluated by field trial).

DISCUSSION

E.R. Schmerling, US

How is the problem of frequency management handled? First, operating frequencies must be chosen depending on ionospheric conditions; second, the selected frequency must avoid interference with neighboring systems.

Author's Reply

In the simple systems I have described, channel evaluation is carried out on assigned frequencies, or channels, so interference is no different from normal channel usage in fact interference will probably be less since automatic channel selection takes less time than manual channel selection.

N.C. **Gerson**, US

For the relatively short distances involved, could you not use a horizontal whip?

Author's Reply

Yes, but for reasons given in the paper the inverted-V was chosen.

ELECTRICALLY SHORT H.F. AERIAL SYSTEMS

M. N. Sweeting and Q. V. Davis

University of Surrey
Guildford
Surrey GU2 5XH
England

SUMMARY

The communications efficiency of electrically short h.f. aerials in a tactical field environment remains unacceptably poor. This paper describes some of the results from a fresh, detailed, practical study of the problem, performed at full-scale and discusses the effect on the aerial performance of several factors shown by the experiments to be important but not previously sufficiently recognised.

1. INTRODUCTION

The theoretical base radiation resistance and reactance of a vertical radiator over an infinite and perfectly conducting ground are well known (SMITH C.1947) and are illustrated, for the case of a 4 m whip aerial, in figures 1 and 2. This paper is concerned with the behaviour of such a whip aerial at frequencies below the electrical quarter-wave where the radiation resistance is decreasing rapidly and the reactance is increasing at a corresponding rate as aerial is shortened.

In order to radiate power effectively, the aerial impedance must be 'tuned' and 'matched' to that of the source by means of a conjugate reactance and transformer. Although the theoretical distant radiated ground-wave field intensity induced by a vertical aerial of infinitesimal height is only 4·25% less than that induced by a $\lambda/4$ aerial radiating the same power in a lossless system (BROWN G.1937) shown in figure 3, in a real environment the unavoidable losses associated with the impedance matching device and an imperfect ground reduces performance drastically (figure 4). In small h.f. aerial systems of this type the dominant loss mechanism is generally associated with the imperfect nature of the groundplane. Figure 4 also illustrates the diminishing effect of a fixed ground loss as the aerial approaches quarterwave resonance.

The characteristics of large, complex ground systems have been studied in some detail (MEISSNER A. 1922) however, there is little data concerning the performance of small transportable systems in field conditions.

2. MEASUREMENTS

A large number of field measurements have been carried out on a standard 4 m whip mounted vertically over various ground systems throughout the frequency range 1 - 20 MHz, in order to evaluate the effectiveness of limited size systems.

In order to assess the performance of the different systems, it would ideally be necessary to carry out a series of protracted distant field intensity comparisons throughout the frequency range of interest. In this instance, however, aerial base impedance measurements were used instead, providing a practicable method for a rapid but reliable indication of relative performance of a set of similar symmetrical aerial systems. This practice is justified later.

The base impedance of the aerial system was measured using a GR 1606A series impedance bridge and a Wayne-Kerr SR 268 source/detector.

The first configuration to be measured was simply the 4 m whip mounted over a short 12" earth spike (½" diameter) driven 11" into the ground co-axially with the whip. The base resistance and reactance characteristic are shown in figures 5 and 6 and are much as anticipated. Although the theoretical radiation resistance continues to decrease below 5 MHz, an increase occurs in the overall resistance indicating a rising loss factor.

The ground system was then gradually enlargened in order to investigate the engineering trade-off between improved performance and increased size. An operational system must be simple, physically compact and quick to erect and dismantle - these criteria generally eliminate buried systems, leaving sheets, wires or meshes mounted on or near the earth.

Since the return earth current density is greatest around the base of the aerial (figure 7) a small 3 ft diameter 16 swg aluminium disc was placed on the ground beneath the aerial and was found to give a slightly lower resistance characteristic below 10 MHz, but again demonstrated the rapid increase in losses at the lowest frequencies (figure 8). The addition of six 8 ft insulated radial wires lying on the ground around the 3 ft disc further improved the resistance below 10 MHz (figure 9)

Enlarging the disc to a 6 foot square sheet reduced the resistance still further (figure 10) however care had to be exercised with the measurements if spurious resonances were to be avoided (figure 11). A six foot sheet of metal was becoming unwieldy and so a series of radial wire only ground systems were examined, commencing with six 8 ft. radials. Figure 12 shows these to exhibit a lower resistance between 5 - 15 MHz but generally increased below 5 MHz when compared to the 3 ft disc with 8 ft radials. The ground conditions had changed considerably (very dry to very damp) between the two tests and may account for the difference. Increasing the radial length to 16 ft made little general difference, Figure 13, and interspersing the six 8 ft radials in between the 16 ft wires gave only a marginal improvement at low frequencies (figure 14).

Amongst the very small ground systems, it appears that the 6 ft square aluminium sheet gave the best performance followed by the 3 ft disc with six 8 ft radials. In some instances it may be practicable to extend the groundplane somewhat further and so a series of larger systems were investigated.

It was found that some improvement below 5 MHz could be gained by increasing the radial length to 32 ft, however a broad peak in resistance centred around 7 MHz (figure 15) occurred which appeared to be independent of the equipment geometry. The reactive component shows no disturbance, probably due to the broadband nature of the resistive peak (figure 16).

When the radial wires were lengthened to 50 ft the resistive peak was moved down in frequency to around 5 MHz, however in this instance, perhaps due to the extremely dry prevailing ground conditions, the peak appears to be narrow band (figure 17). Further increases in radial length to 100 ft and then 150 ft brought the peak down to around 2·5 and 2MHz respectively (figure 18 and 19). Both the 100 ft and 150ft radial systems however exhibited multiple peaks, particularly the latter system which was laid on extremely dry ground. Reducing the number of 100 ft radials from six to four resulted in the same peak frequencies but a general increase in resistance (figure 20).

The multiple nature of these peaks and the close relationship between radial length and centre frequency strongly suggested a resonance mechanism associated with radial length. The resonant frequency appeared to correspond to the $^{n\lambda}/4$ 'free space' wavelength on the line, however the resonance clearly resulted in a high impedance transformation characteristic of a $^{n\lambda}/2$ resonance. In fact, using the impedance bridge, $^{n\lambda}/4$ resonances are almost impossible to detect whereas $^{n\lambda}/2$ resonances are easy to discern as they result in a high series impedance in the aerial system. Before investigating the radial resonance phenomena further the characters of a number of larger ground systems are presented.

In an attempt to reduce the effect of the radial resonances short and long radial wires were interspersed and figures 21 and 22 show that the resonances are considerably reduced in amplitude. This reduction can also be effected by earthing the centre common point or, preferably, the end of each radial wire as can be seen in figures 23 and 24. The aerial was erected over a 12" earthspike driven into a level pebble beach covered by 2" of sea water and figure 25 illustrates that only a modest general improvement was obtained, a fact that might be explained by the distribution of fertilizing salts over the home test range. To confirm that the increased resistive component over the theoretical predicted radiation resistance was in fact almost entirely due to losses in the ground system, the whip was mounted over the largest metal groundplane available - an aluminium sheet 300' by 80' by 50 thou thick. The resistance now falls very rapidly, following the radiation resistance, with decreasing frequency (figure 26) however the finite size of the ground system can still be detected by the rise in resistance at the lowest frequencies measured. The measured resistance around 12 MHz is slightly less than that predicted for the radiation resistance which may be due to the influence of the operator or limitations in the absolute accuracy of the measurement.

Even disregarding the resonance effects there is little general improvement shown by the extended ground systems (except perhaps at the lowest frequencies) apart from the large metal sheet. In fact the 6 ft aluminium sheet is still very competitive.

3. RADIAL RESONANCES

The current standing wave pattern was plotted along a resonant radial wire confirming the resonances to be functions of $^{n\lambda}/2$ (figure 27). On a suggestion by James Wait an outer wire ring was placed over but disconnected from the ends of the radial wires but figure 28 shows no appreciable modification to the ground system performance, however when the ends of the radial wires were linked to the outer wire circle the resonance dropped in frequency (figure 29). When four symmetrical radials were end-linked in a square, the resonance dropped still further (figure 30). An examination of the current standing wave at resonance confirmed the resonance mechanism shown in figure 31. Although the resonant frequency dropped, there was little apparent improvement in performance, probably due to the end-linking wires not being parallel to the return earth currents. Measurements on two spiral ground systems providing a large amount of conductor perpendicular to the return earth currents clearly demonstrate the necessity for continuous radial conductors - it is not purely the amount of good conductor in the vicinity of the aerial base, figures 32 and 33, where the general resistance characteristic is very close to that of a simple earthspike. The low frequency resonances appear to be spiral mode resonances and are again reduced in amplitude when the common point is earthed (figure 34).

The $^{n\lambda}/2$ current standing waves on an approximately $^{n\lambda}/4$ 'free space' wavelength wire indicates that the effective phase velocity on the radials has been reduced by a factor of almost exactly two. The most likely cause for this reduction in phase velocity appears to be the proximity of a lossy dielectric earth.

To test this hypothesis the entire aerial system (including radials) was gradually elevated above the ground, thus reducing the influence of the adjacent 'dielectric' half-space, which should result in a corresponding increase in phase velocity on the wires and hence radial resonant frequency. Figure 35 demonstrates the increase in radial resonant frequency as a function of the height of the system above ground. It can be seen that the resonant frequency corresponds to that expected from 'free space' phase velocity when the system was 5 ft above ground. It is interesting to note that as the height of the system above ground increases there is a corresponding increase in resonant Q, and that the rate of increase of phase velocity decreases.

Let the phase velocity on a wire in free space be V_o, if the wire is then placed at the interface between free space and a dielectric medium ϵ_g (earth) then the phase velocity reduces to :

$$V_{ph'} = \frac{2V_o}{\sqrt{\epsilon_g}}$$

and if the wire is buried some distance into the dielectric medium then it reduces still further to:

$$V_{ph''} = \frac{V_o}{\sqrt{\epsilon_g}}$$

The 50 ft radial system was buried 1" underground and figure 36 shows how the resonance for the radials on the surface at around 4·9 MHz has been reduced further in frequency to around 1·75 MHz. For a radial wire immersed near the junction of two media, providing the dielectric constant of the second (more dense) medium is much greater than that of the first, the effect of the first medium on the phase velocity will become negligible only a short distance into the second, dense medium - in this case, the earth. The results of the experiment yield an effective dielectric constant for the ground of about 32. An independent measurement of the effective ground dielectric constant using a sealed sample substitution method gave a value of between 30 - 34.

Analysis of the results in figure 35 shows the phase velocity on the radial wires to depend upon height above or below the interface according to figure 37. It was not possible to confirm the value exactly in the interface due to ground roughness on the range. It is expected that this curve will depend on the effective ground dielectric properties in the manner illustrated in figure 38, however further work is required to verify this.

This phenomenon raises the possibility of providing a relatively simple method for determining the effective dielectric constant of the ground over a wide range of radio frequencies. One straight forward measurement to determine the $\lambda/2$ resonant frequency, and hence phase velocity, on a shallow buried wire appears to provide an accurate bulk effect value.

Relative distant field intensity comparisons between resonant and non-resonant systems were conducted to confirm that the high impedance $n\lambda/2$ radial resonances did indeed result in a reduction in radiated power at and around the resonant frequency. Several systems were examined and figures 39 and 40 show the distant field intensities for two systems each normalised to a non-resonant earthspike system. These measurements illustrate very clearly the substantial reductions in distant field intensity, that occurred around the radial resonant frequencies, and thus in system efficiency.

When the predicted relative field intensity based on the impedance measurements are plotted on the same graphs, (figures 41 and 42) then the close correlation between the base impedance measurements and the relative distant field intensities is quite evident, establishing base impedance comparisons as a rapid and useful method of assessing this type of simple, symmetrical groundplane.

During the measurements on elevated systems a substantial reduction in resistance was apparent when the system was lifted clear of the ground, either due to reduced effective ground loss or to modification of the radiation characteristics of the aerial system (figure 43). Groundwave field intensity comparisons shown in figure 44, where a 16 ft radial system on the ground is compared to the same system raised 5 ft away from the earth, shows quite clearly a significant improvement in radiated field intensity for the elevated system. This may be relevant to vehicular mounted system.

4. CONCLUSIONS

This practical study of electrically short h.f. aerials has provided fresh insight into their behaviour in a field environment, and has demonstrated that :

- Aerial base impedance measurements appear to provide a straight forward but rapid means for the evaluation of symmetrical aerial systems.
- Small ground systems can provide a significant improvement over a simple earthspike, and are competitive with larger radial systems when wideband operation is required.
- Larger ground systems may provide improved narrow-band performances, however radial resonances at previously unexpected frequencies can give rise to substantial losses of efficiency.
- If narrow-band operation is desired then the radials should be designed to be $n\lambda/4$ resonant, taking into account the phase velocity reduction due to the proximity of the earth.
- Small elevated radial systems appeared to exhibit a 2 - 7 dB groundwave advantage over the same system on the ground.
- Radial resonance frequencies can be predicted easily from a knowledge of the effective dielectric constant of the ground and the system geometry.
- A straightforward measurement of the $n\lambda/2$ resonant frequency of a shallow buried radial wire can be used to determine the effective dielectric constant of the ground at radio frequencies with minimal mathematics.

5. ACKNOWLEDGEMENTS

The authors are grateful to Mike Underhill and Bill Painter for many invaluable discussions concerning the aerial behaviour. The study was carried out at the University of Surrey supported by the M.E.L. Co. Ltd., England.

References

ABBOT, F.R., July 1952, "Design of optimum buried-conductor R.F. ground systems", Proc. IRE, Vol. 40. pp.846-852.

ALBRECHT, H.J , May 1965, "On the relationship between electrical ground parameters", Proc. IEEE (correspond.) , Vol.53. p.544.

ALBRECHT, H.J., March 1966, "Climatic influence upon skin depth in earth", Proc. IEEE, Vol.54. pp.394.

BOUWKAMP, C.J., June 1947, "Calculation of the input impedance of a special antenna", Philips Res.Rep. Vol.2. pp.228-240.

BROWN, G.H., 1935, "The phase and magnitude of earth currents near radio transmitting antennas", Proc. IRE, Vol.23. p.168.

BROWN, G.H., LEWIS, R.F., EPSTEIN, P., June 1937, "Ground systems as a factor in antenna efficiency", Proc. IRE, Vol.25 pp.753-787.

DE SANTIS, C.M., CAMPBELL, D.V., SCHWERING, F., Nov. 1973, "An array technique for reducing ground losses in the H.F. range", Trans. IEEE AP-21, No.6. pp.769-773.

ELIASSEN, K.E., 1957, "A survey of ground conductivity and dielectric constant in Norway within the frequency range 0·2 - 10 MHz", Geo. Pub. Vol.19. No.11.

GRIESE, R.H., 1974, "Improving the performance of low-frequency vertical antennas', Ham Radio, pp.54-57.

HARRISON, C.W., KING, R.P.W., Sept. 1962, "On the impedance of a base-driven vertical antenna with a radial ground system", IRE Trans. Vol.AP-10. No.5. p.640.

HILL, D.A., WAIT, J.R., "Calculated pattern of a vertical antenna with a finite radial-wire ground system", Radio Sci. 8, pp.81-86

HILL, D.A., WAIT, J.R., 1973, "Effect of edge reflections on the performance of antenna ground screens", Trans. IEEE. AP-21. pp.230-231.

HOESTENBACH, R , Dec.1976, "Improving earth-ground characteristics", Qst. pp.16-17.

KING, R.J , Oct. 1952, "Theory of electrically short transmitting and receiving antennas", Journ. Appl. Phys., Vol.23. No.10. pp.1174-1187.

KIRKSCETHER, E.J., 1960, "Ground constant measurements using a section of balanced two-wire transmission line", IRE Trans. A & P, pp.307-312.

KNUDSEN, H.L., June 1959, "Earth currents near a top loaded monopole antenna with special regard to electrically small L - and T antennas", J. of Res. N.B.S. 62. No.6. pp.283-296.

LARSEN, T., March-April 1962, "The E-field and H-field losses around antennas with a radial wire ground system", J.Res. B.V.S. Vol.66D. pp.189-204.

LEITNER, A., SPENCE, R.D., Oct.1950, "Effect of circular groundplane on antenna radiation", Journ. Appl. Phy. Vol.21. pp.1001-1005.

MALEY, S.W., KING, R.J., March 1962, "Impedance of a monopole antenna with a radial-wire ground system on an imperfectly conducting half-space - Part I", J. of Res. N.B.S., Vol.66D. No.2. pp.175-180.

MALEY, S.W., KING, R.J., 1964, "Impedance of a monopole antenna with a radial-wire ground system on an imperfectly conducting half-space - Part II", J. of Res. N.B.S., Vol.68D. No.2. pp.157-163.

MALEY, S.W., KING, R.J., March 1961, "Impedance of a monopole antenna with a circular conducting-disc ground system on the surface of a lossy half-space". J.Res. N.B.S., Vol.65D. No.2. pp.183-188.

MEIER, A.S., SUMMERS, W.P., 1949, "Measured impedance of vertical antennas over finite groundplanes", Proc. IRE, Vol.37. pp.609-616.

MEISSNER, A., 1922, "Earthing resistance of Antennae", Yearbook of Wireless Teleg. and Teleph, pp.1235-1244.

MONTEATH, G.D., 1958, "The effect of the ground constants and of the earth system on the performance of a vertical medium-wave aerial", IEE (London) Monograph No. 279R. pp.1-15 Proc. IEE Vol. 105C.

NORTON, K.A., 1937, "PartII: Propagation from vertical, horizontal and loop antennas over a plane earth of finite conductivity", Proc. IRE, 25, pp.1203-1222.

POPE, A.E., 1957, "Report on impedance measurements on various types of aerials in the 2-20 Mc/s range", S.R.D.E. Tech. Memo No. RAD 4/57.

SCHIEFER. G., "Physical limitation of small aerials", Phillips Zeatrauaboratoruim Laborbericht Nr 28.

SMITH, A.N., DEVANY, T.E., 1959, "Fields in electrically short ground systems: an experimental study", J. Res. N.B.S., Vol.63D. No.2. pp.175-180.

SMITH-ROSE, R.L., "The electrical properties of soil for alternating currents at radio frequencies", Proc.R.S., Vol.140.A. pp.355-377.

SMITH-ROSE, R.L., BARFIELD, R.H., 1925, "On the determination of the directions of the forces in wireless waves at the earth's surface", Proc.R.S., 107. p.587.

STORER, J.E., Aug. 1951, "The impedance of an antenna over a large circular screen" Journ.Appl.Phys. Vol.22. pp.1058-1066.

WAIT, J.R., 1962, "Effective impedance of a wire grid parallel to the earth's surface", IRE Trans.A-P, pp.536-542.

WAIT, J.R., Oct. 1969, "Impedance characteristics of electric dipoles over a conducting half-space", Radio Sci., Vol.4. No.10. pp.971-975.

WAIT, J.R., POPE, W.A., 1955 "Input resistance of L. F. unipole aerials", Wireless Engineer, 32, p.131.

WAIT, J.R., POPE, W.A., 1954, "The characteristics of a vertical antenna with a radial conductor ground (screen) system", Applied Scientif Research B.4. p.177.

WELLS, N. 1942, "Aerial characteristics", J.IEE, Vol.89. Pt.III. pp.76-99.

WHEELER, H.A., 1975, "Small antennas", Trans. IEEE, AP-23.NO.4. pp.462-469.

Fig.1

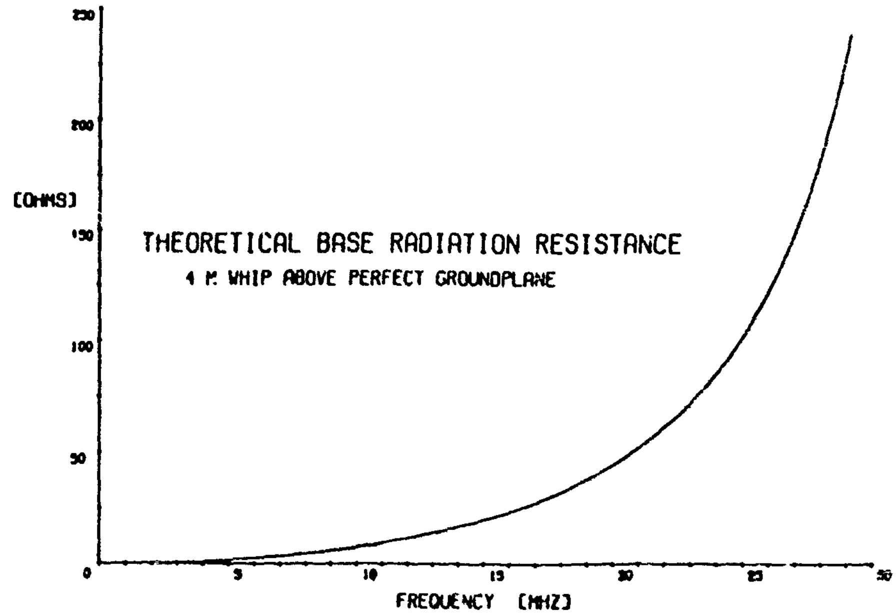


Fig.2

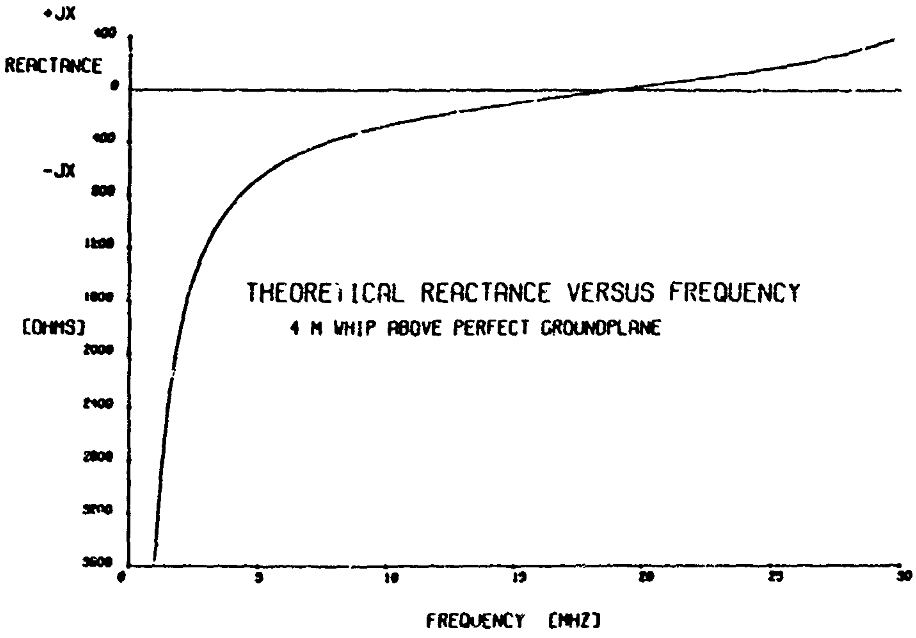
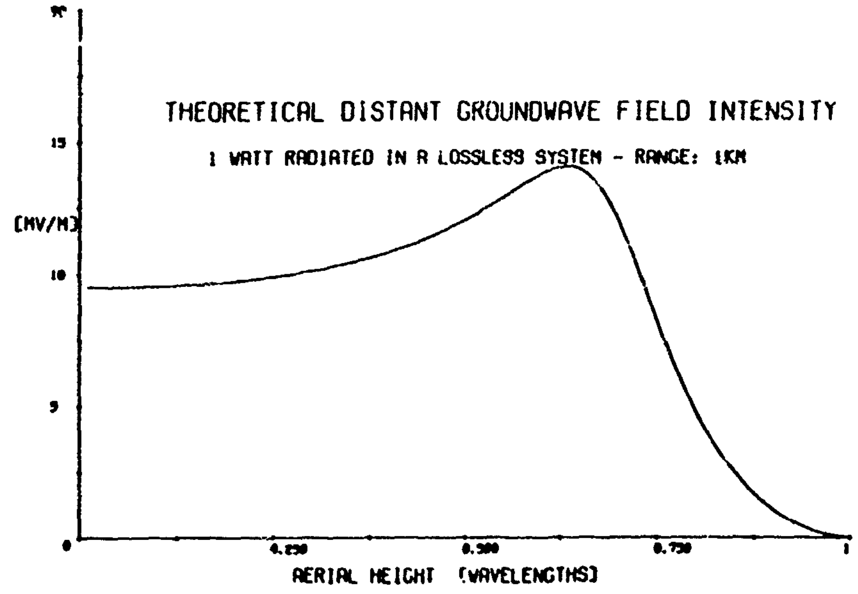


Fig.3



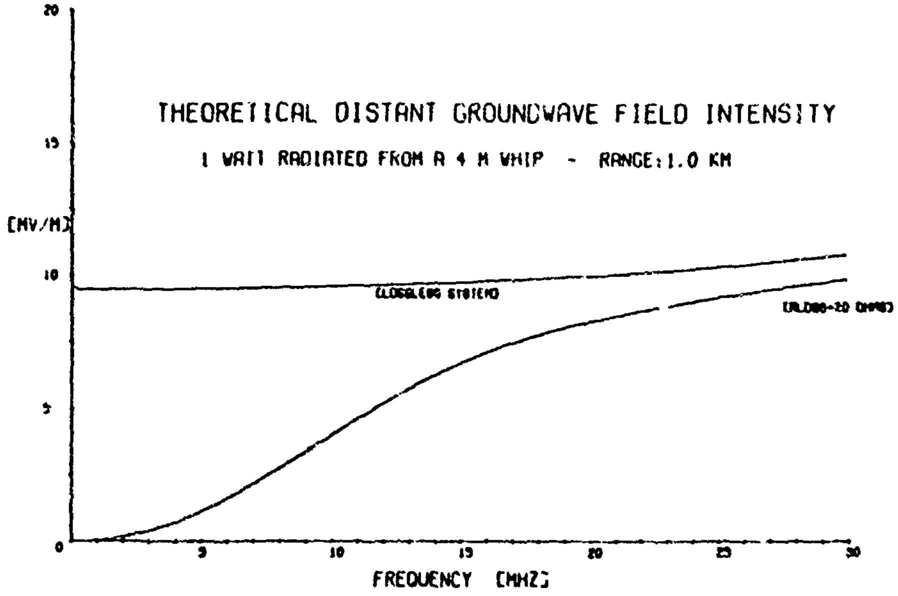


Fig.4

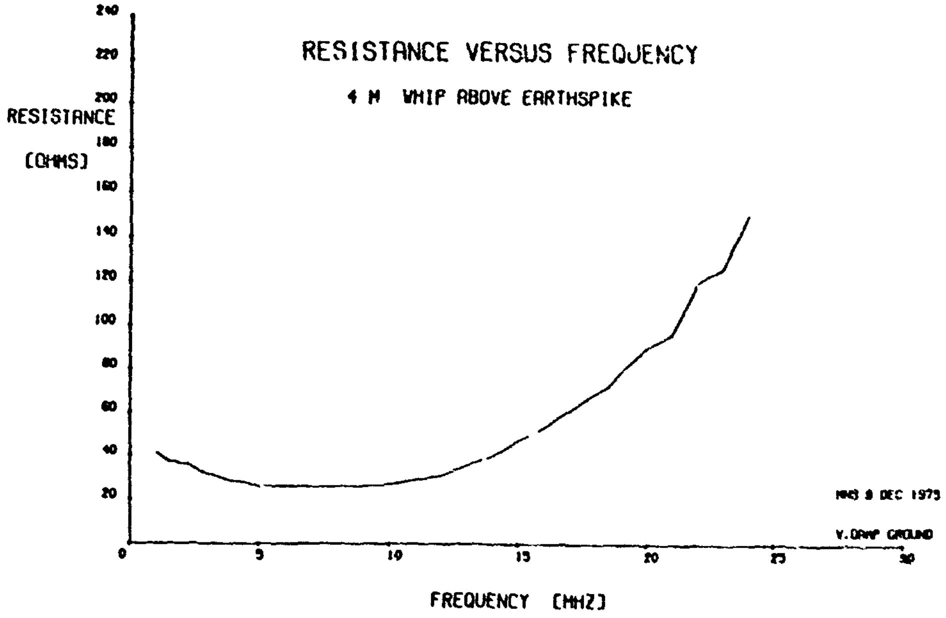


Fig.5

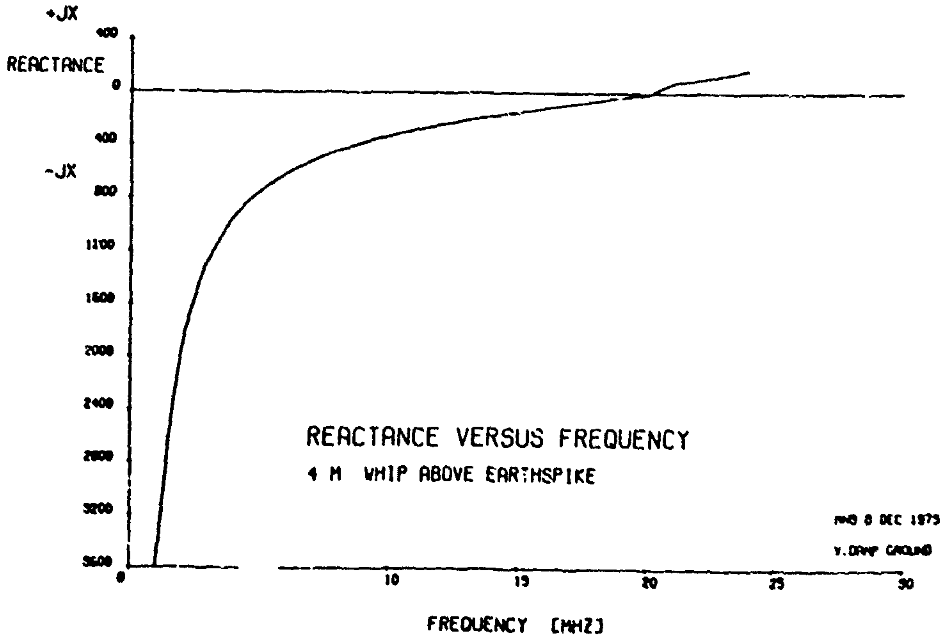


Fig.6

Fig.7

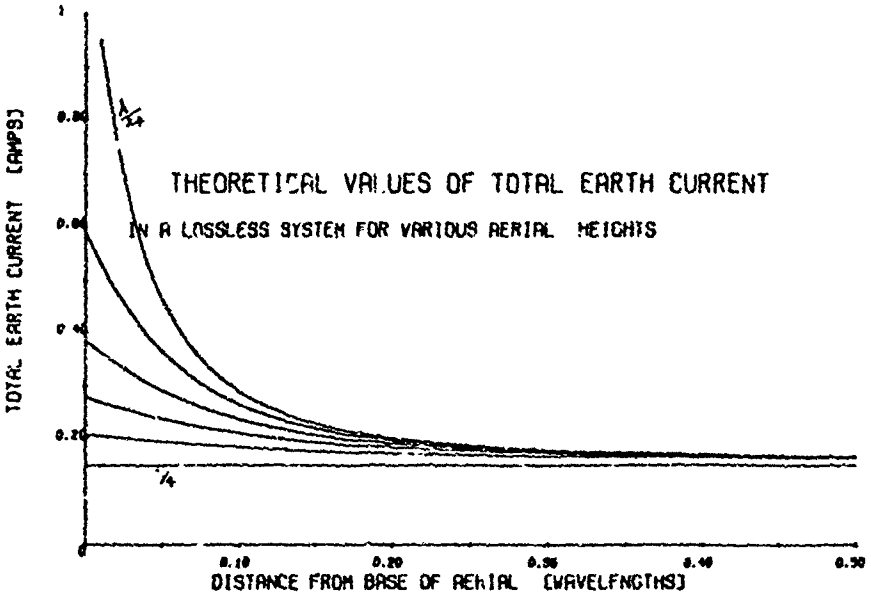


Fig.8

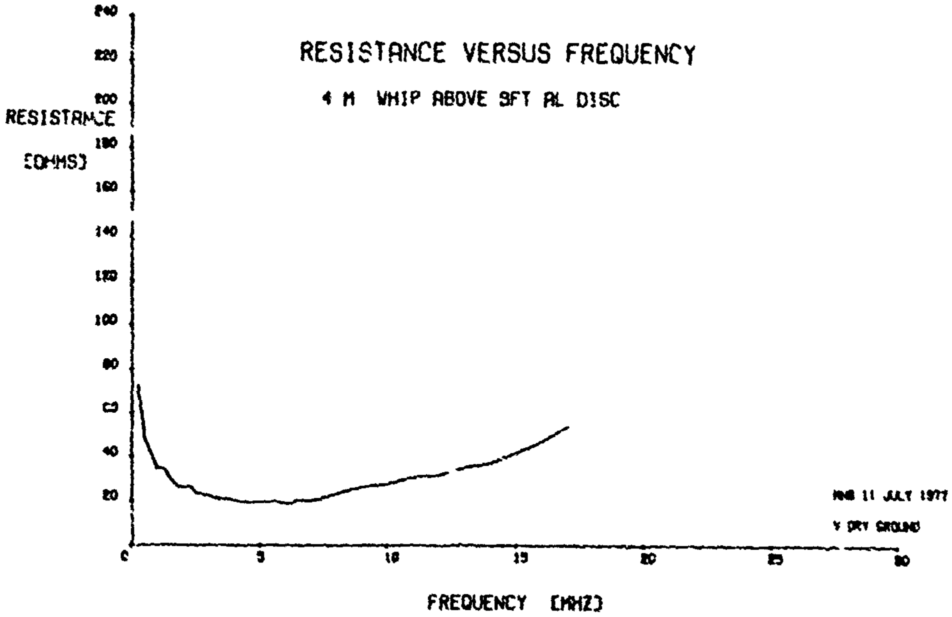
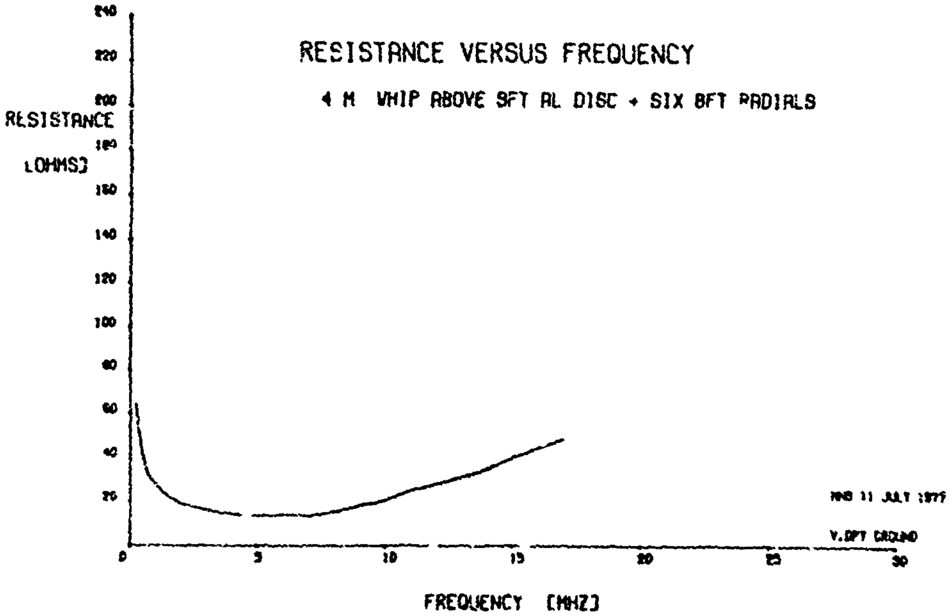


Fig.9



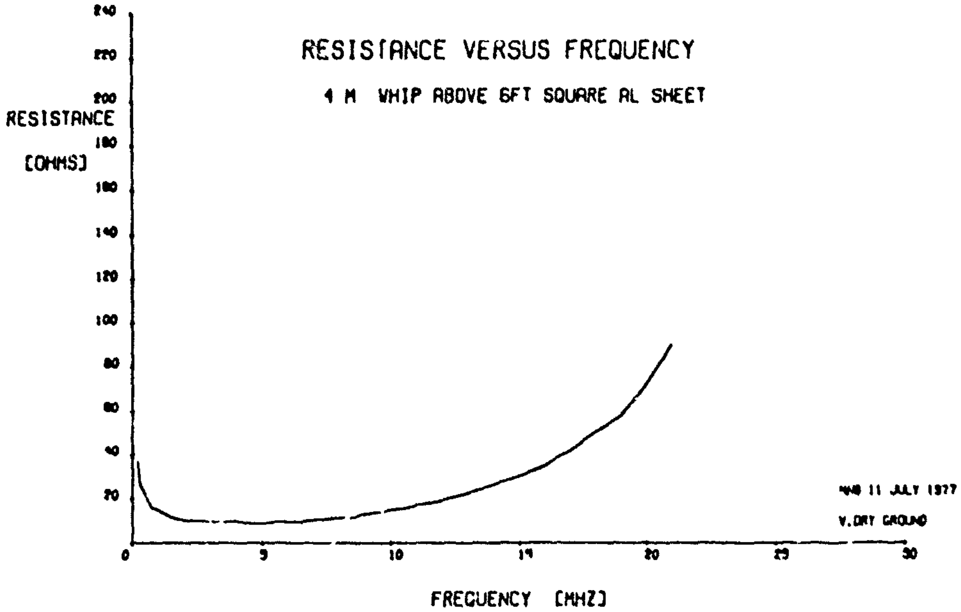


Fig.10

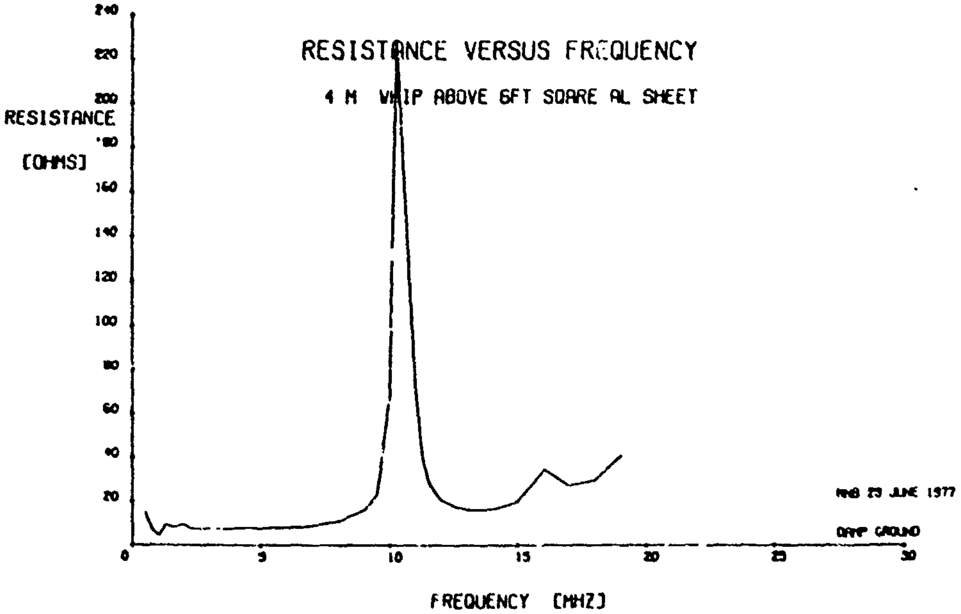


Fig.11

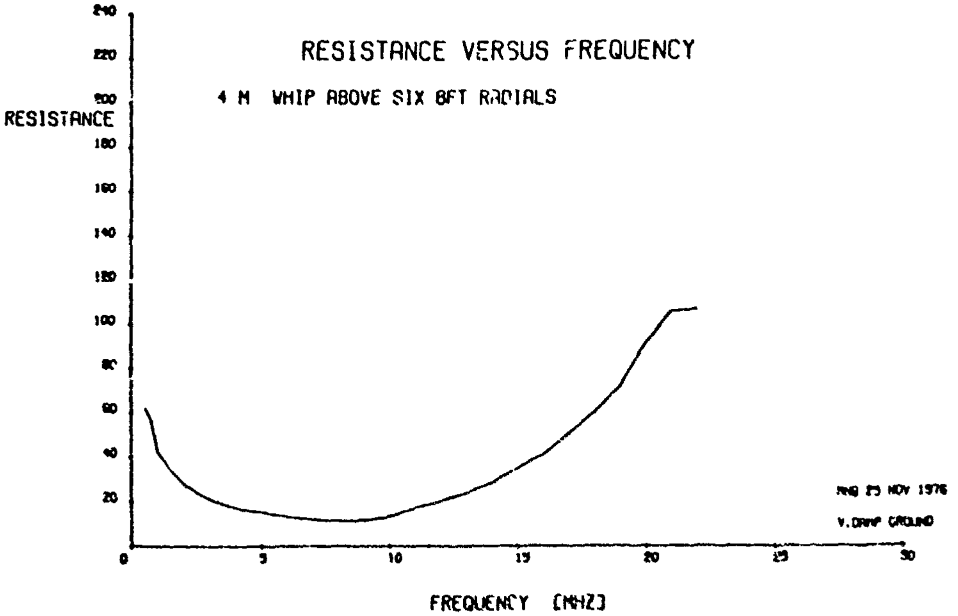


Fig.12

Fig.13

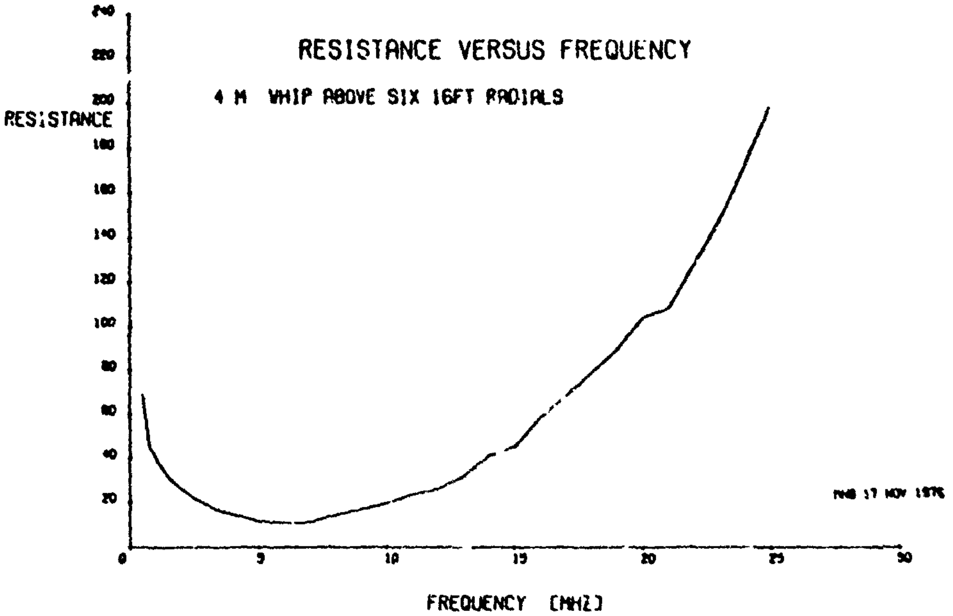


Fig.14

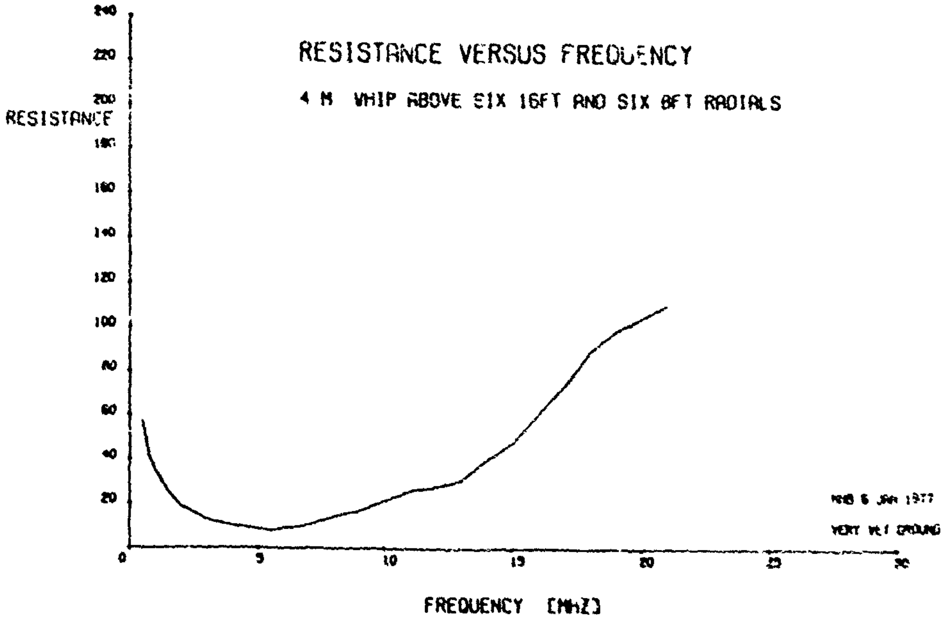
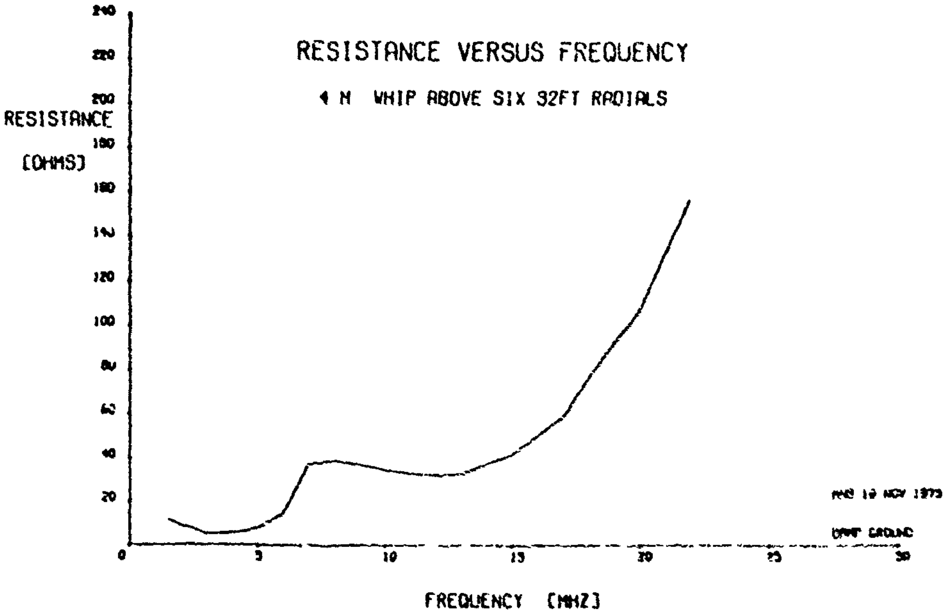


Fig.15



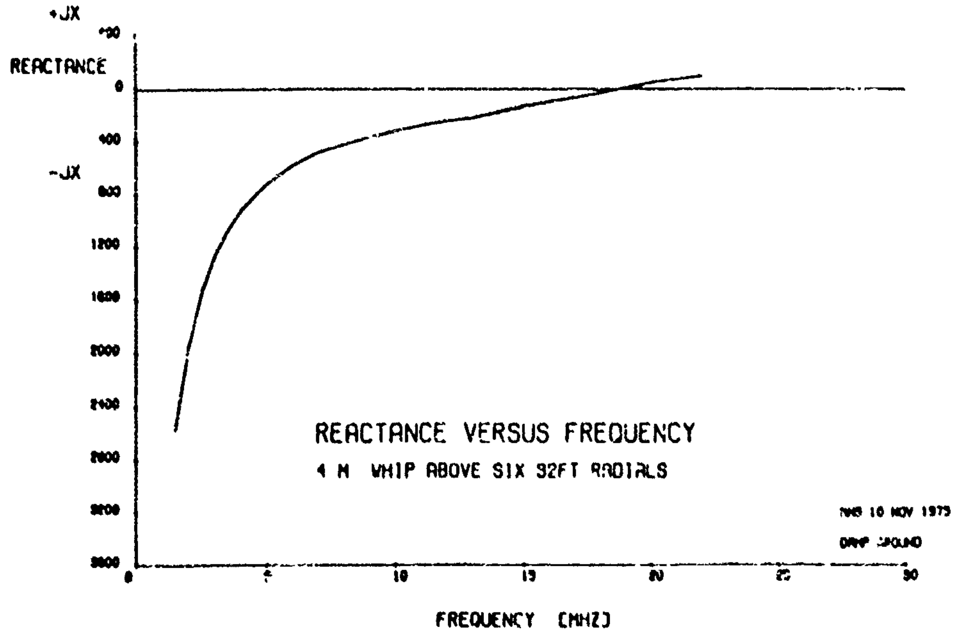


Fig.16

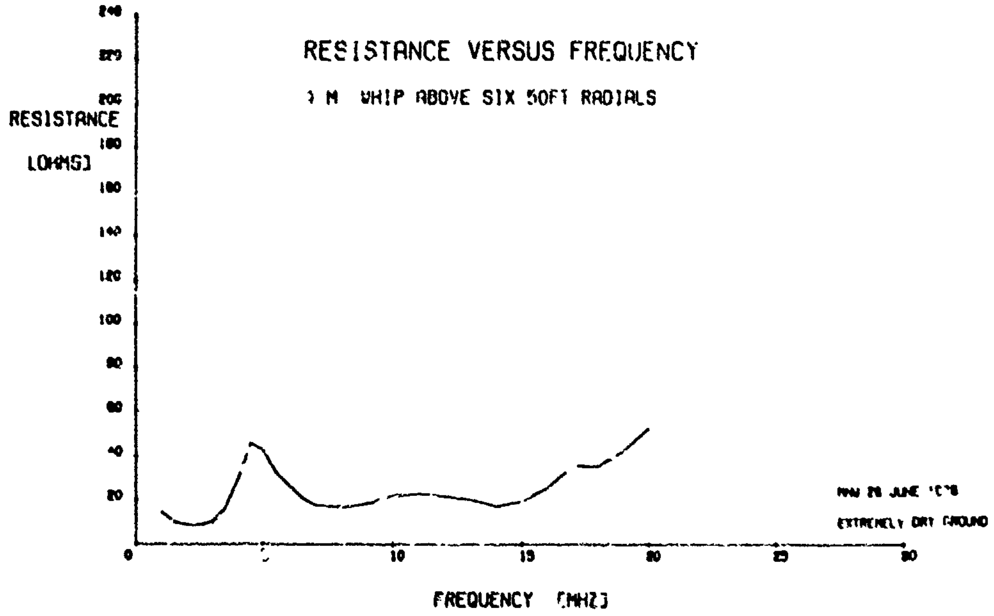


Fig.17

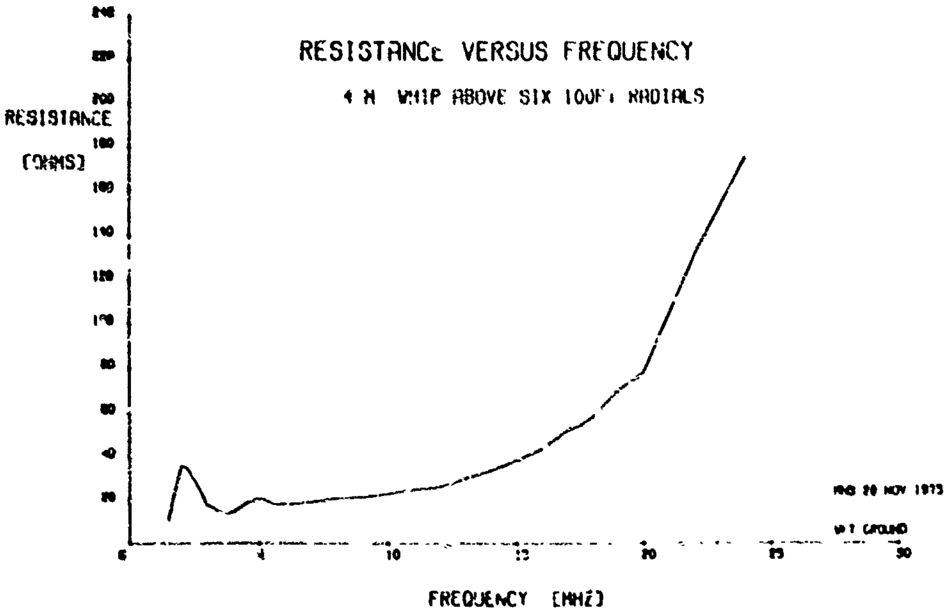


Fig.18

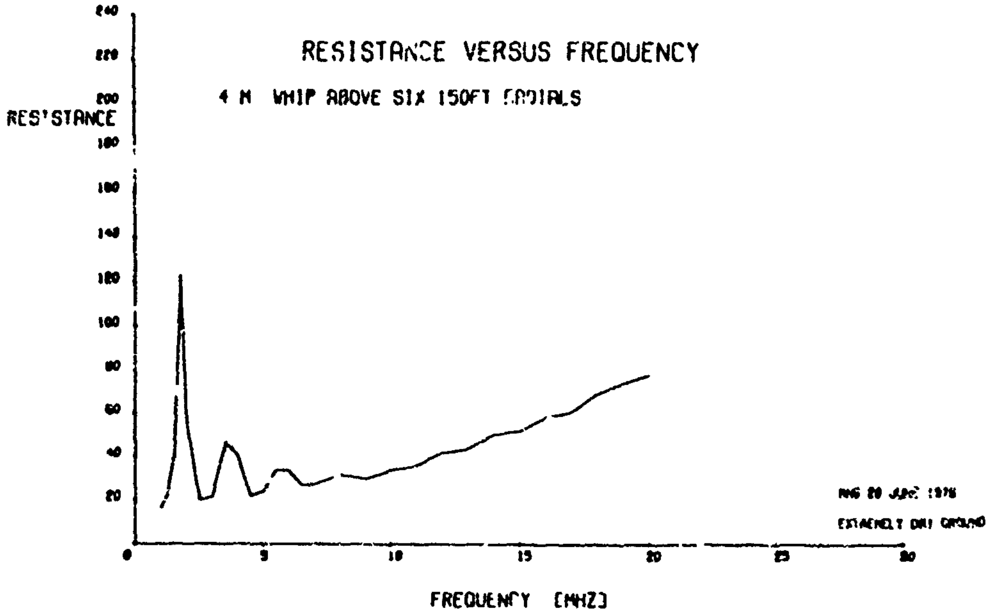


Fig.19

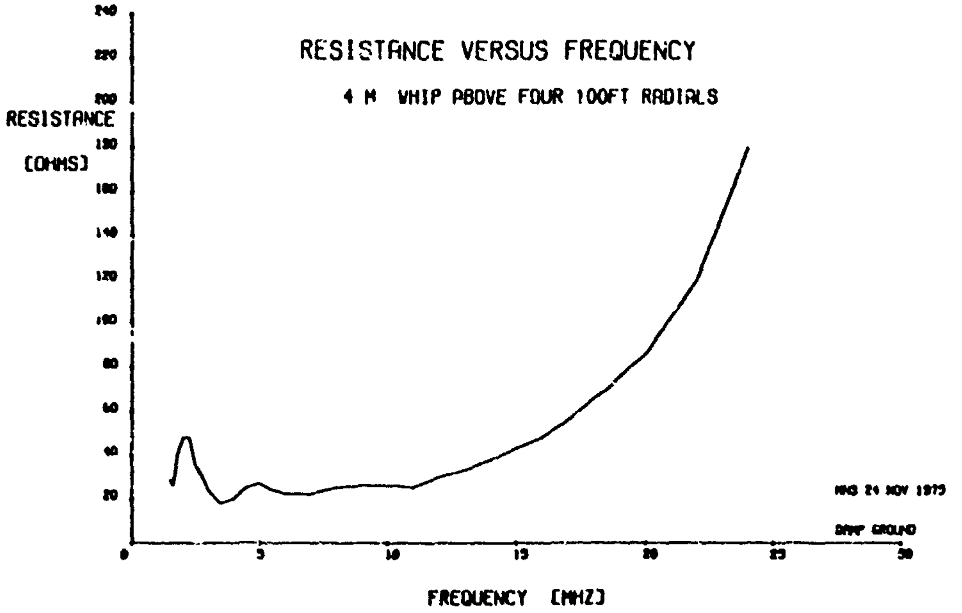


Fig.20

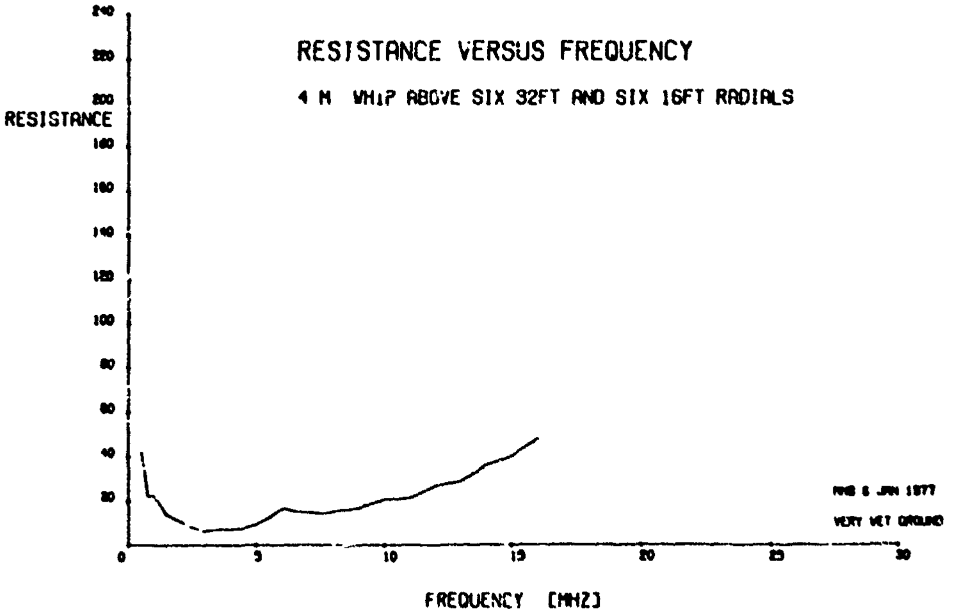


Fig.21

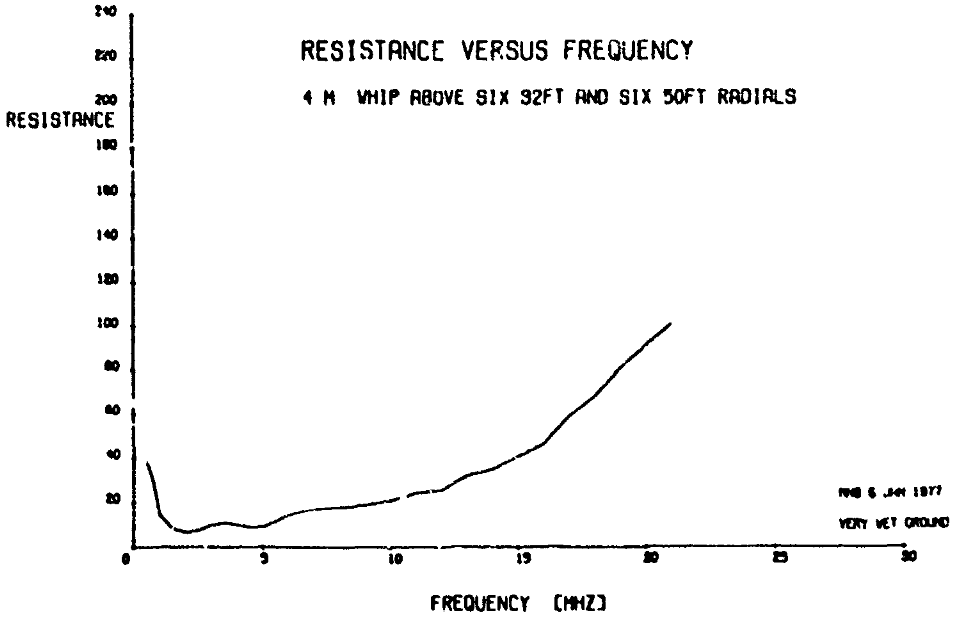


Fig.22

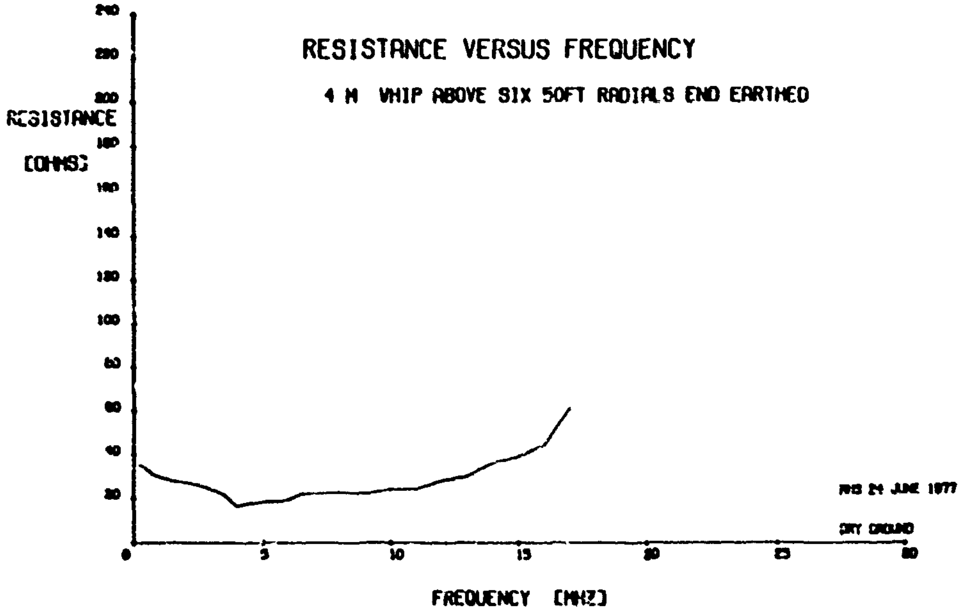


Fig.23

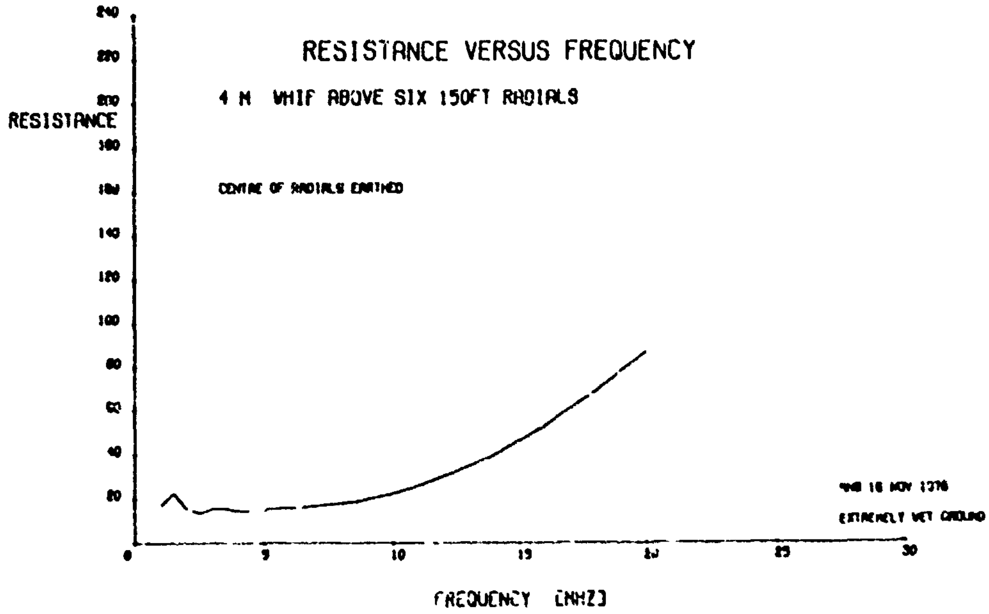


Fig.24

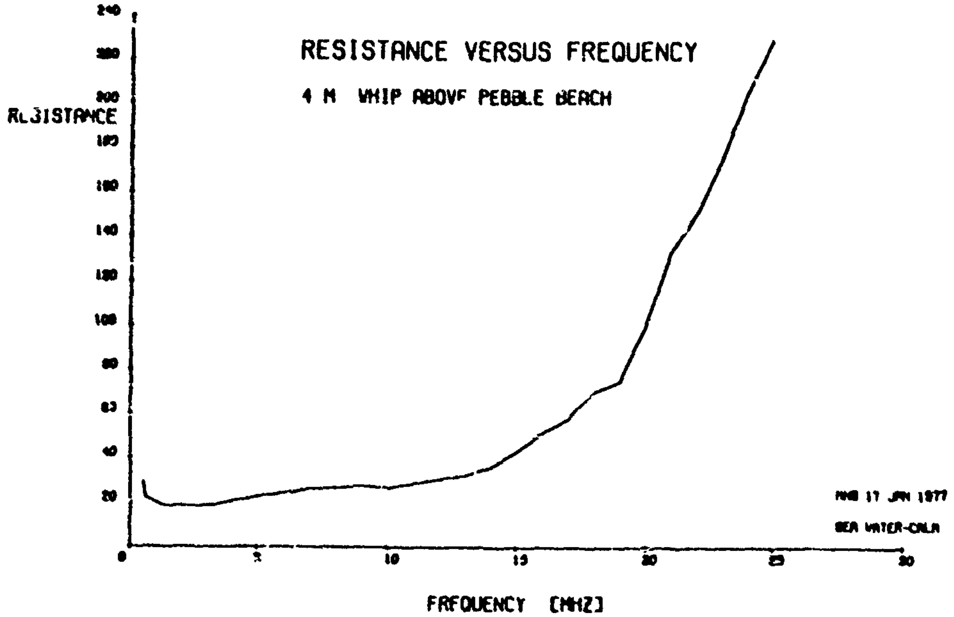


Fig.25

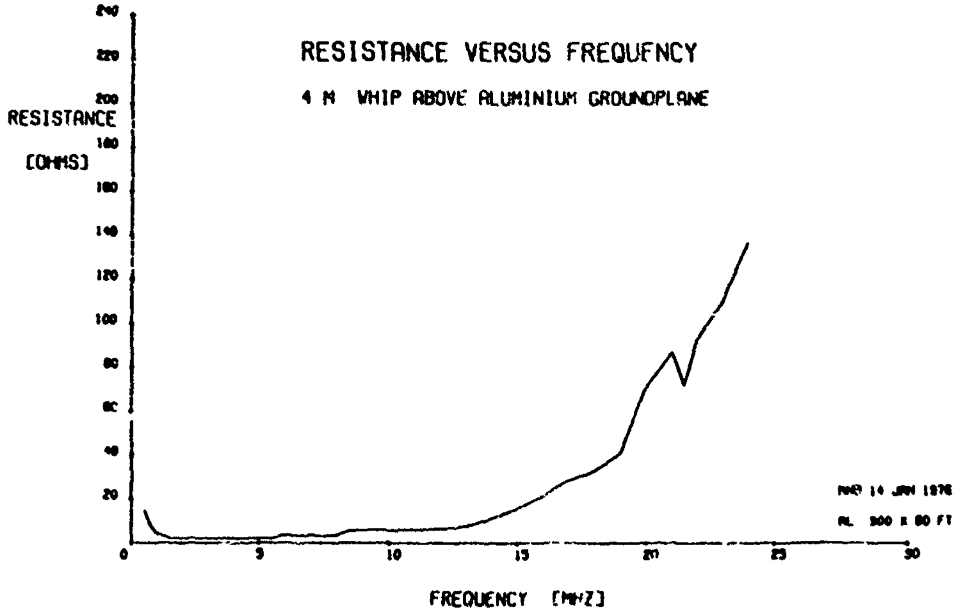


Fig.26

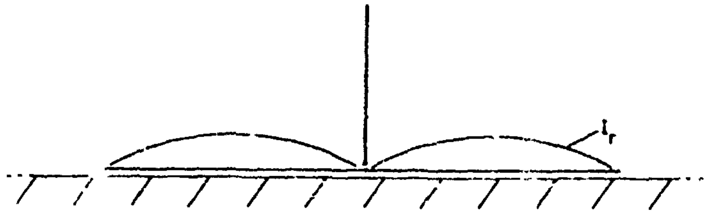


Fig.27

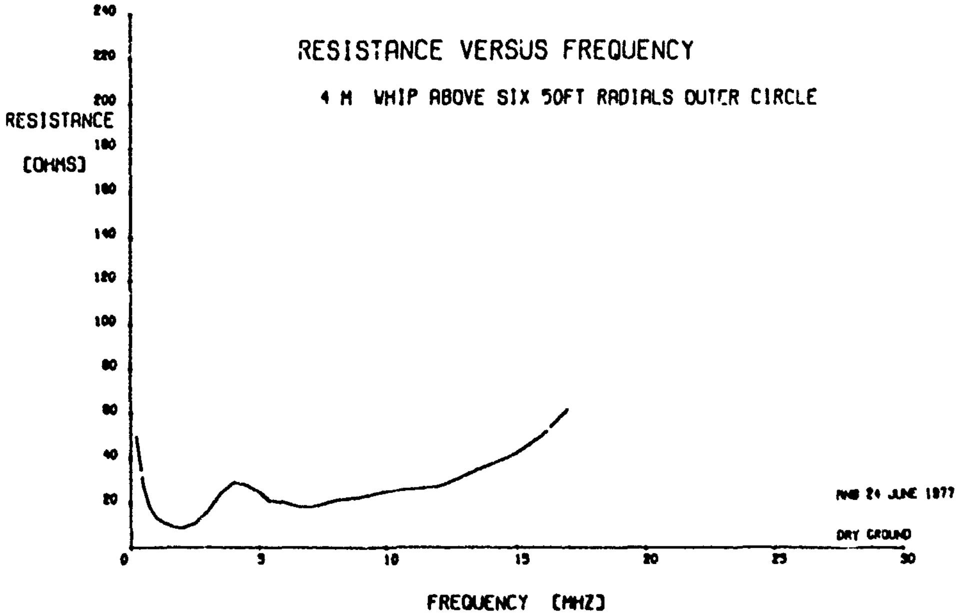


Fig. 28

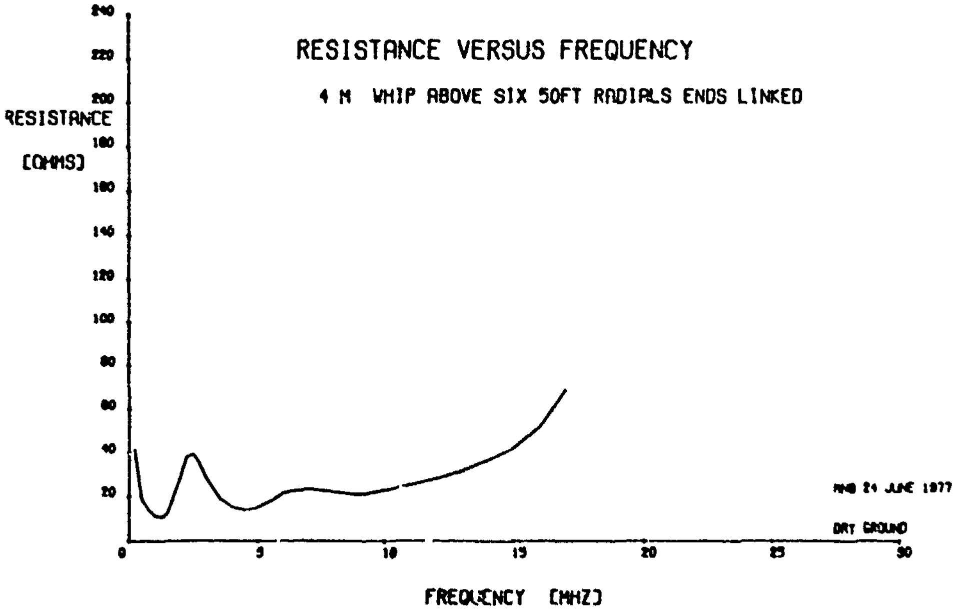


Fig. 29

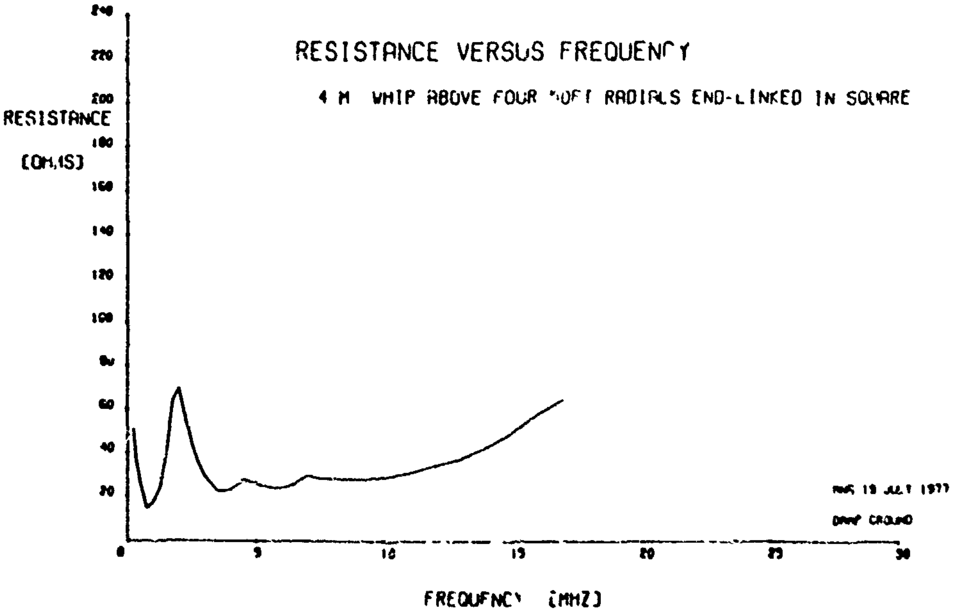


Fig. 30

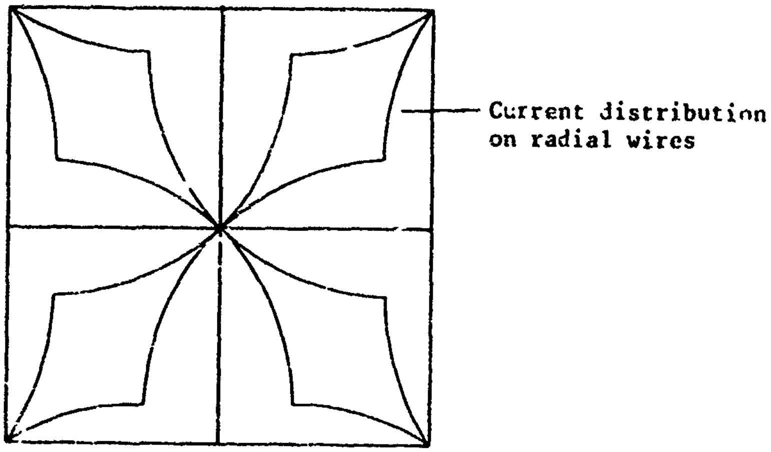

Current distribution on radial wires

Fig.31

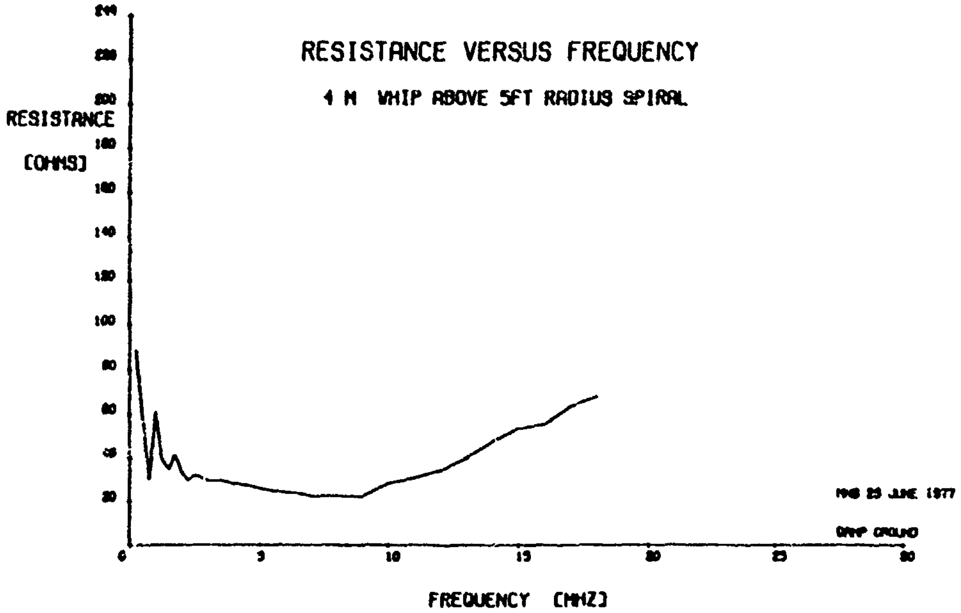


Fig.32

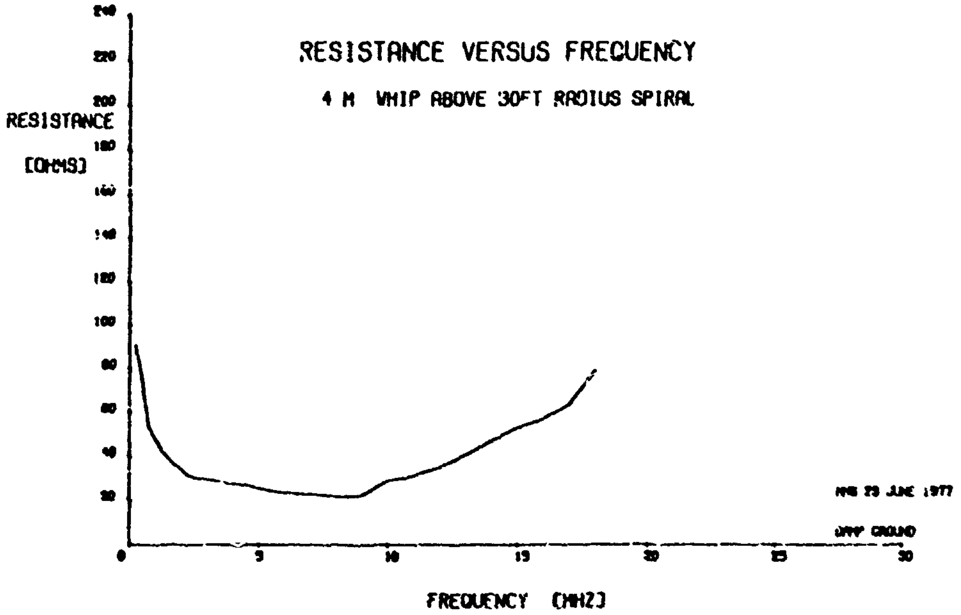


Fig.33

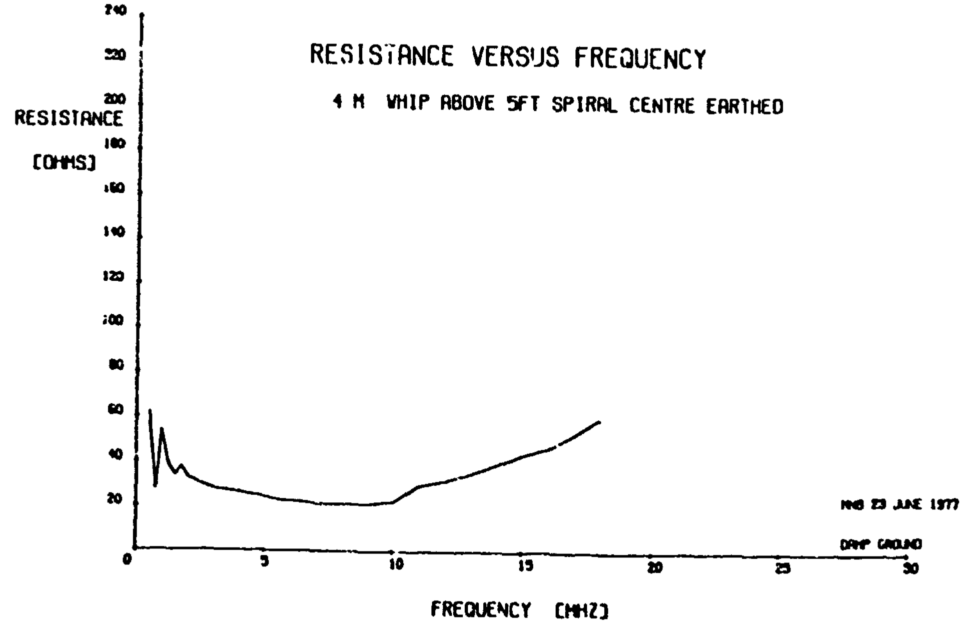


Fig.34

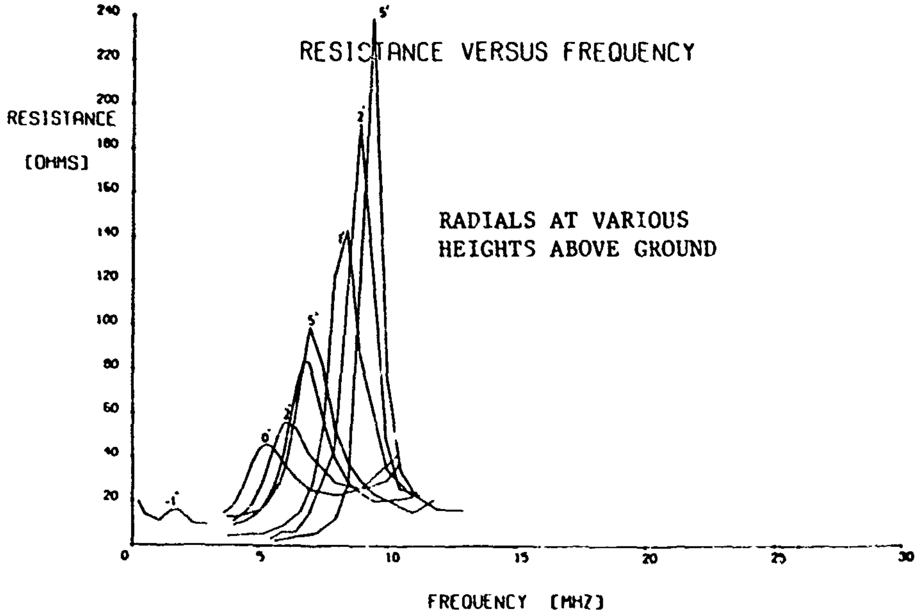


Fig.35

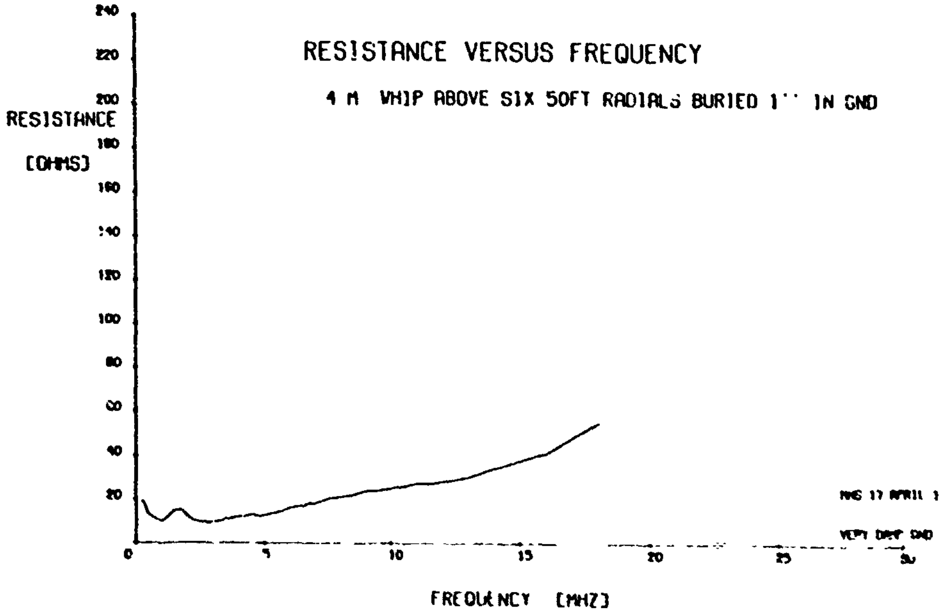


Fig.36

Fig.37

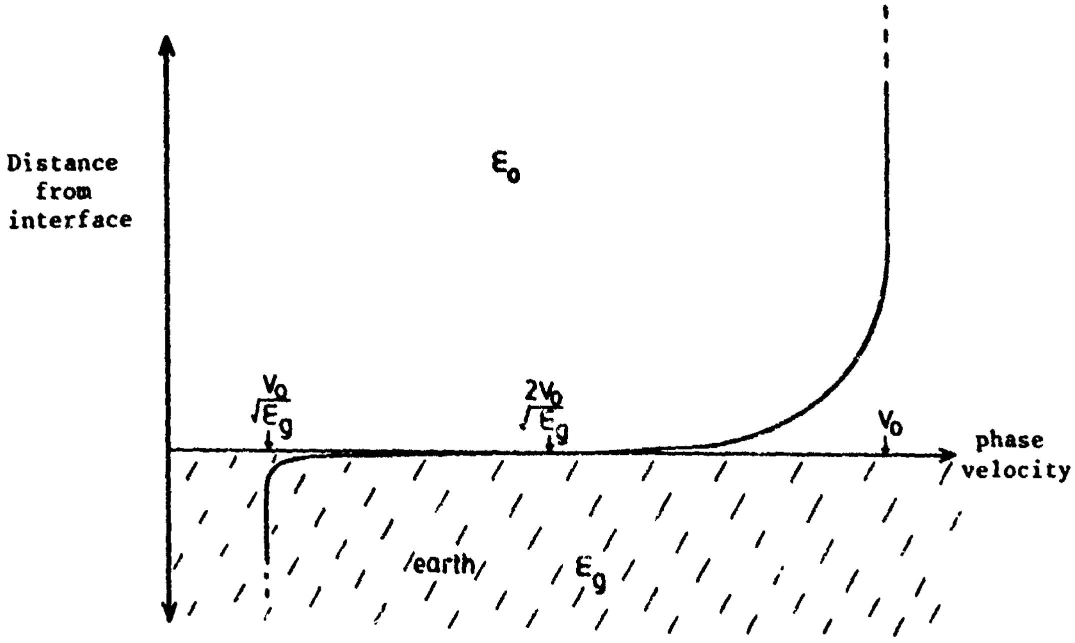
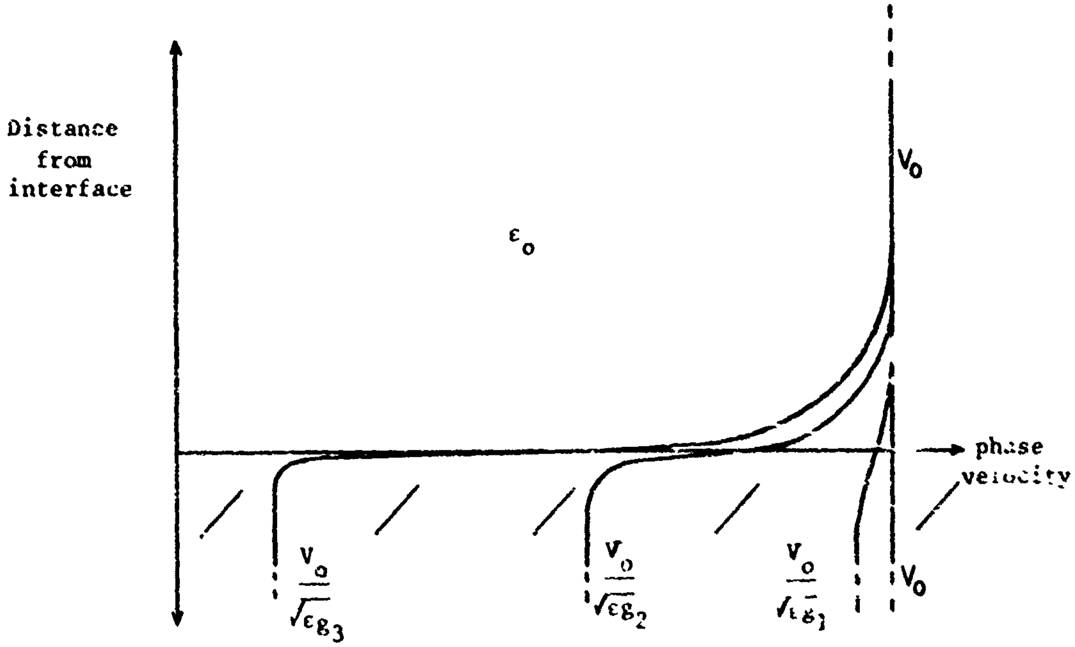

Distance
from
interface
ε_0
$\frac{V_0}{\sqrt{\varepsilon_g}}$
$\frac{2V_0}{\sqrt{\varepsilon_g}}$
V_0
phase
velocity
earth
ε_g

Fig.38


Distance
from
interface
ε_0
V_0
phase
velocity
$\frac{V_0}{\sqrt{\varepsilon g_3}}$
$\frac{V_0}{\sqrt{\varepsilon g_2}}$
$\frac{V_0}{\sqrt{\varepsilon g_1}}$
V_0

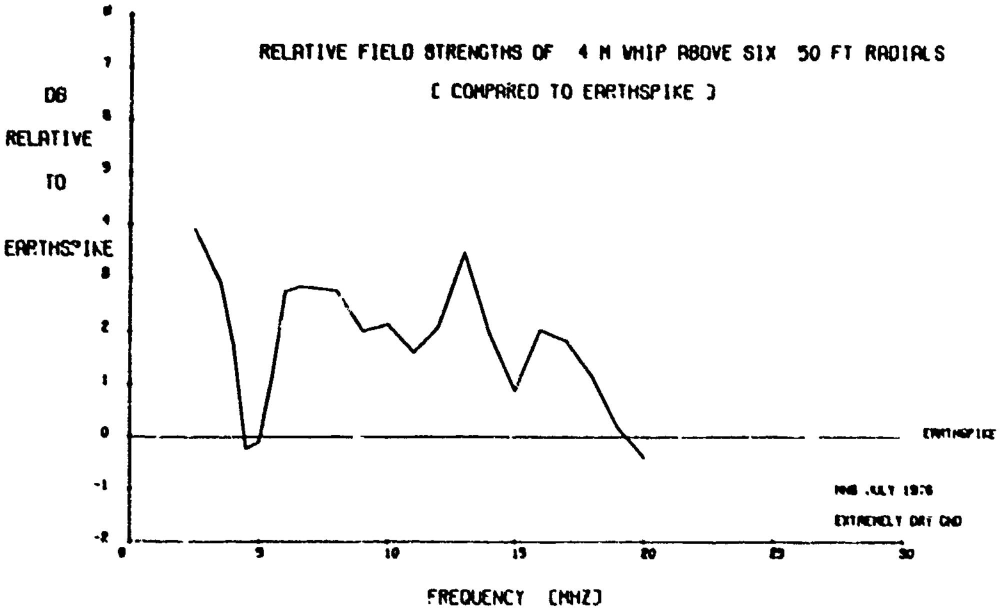

RELATIVE FIELD STRENGTHS OF 4 M WHIP ABOVE SIX 50 FT RADIALS
[COMPARED TO EARTHSPIKE]
DB
RELATIVE
TO
EARTHSPIKE
8
7
6
5
4
3
2
1
0
-1
-2
0
5
10
15
20
25
30
FREQUENCY [MHZ]
EARTHSPIKE
NHB JULY 1978
EXTREMELY DRY GND

Fig.39

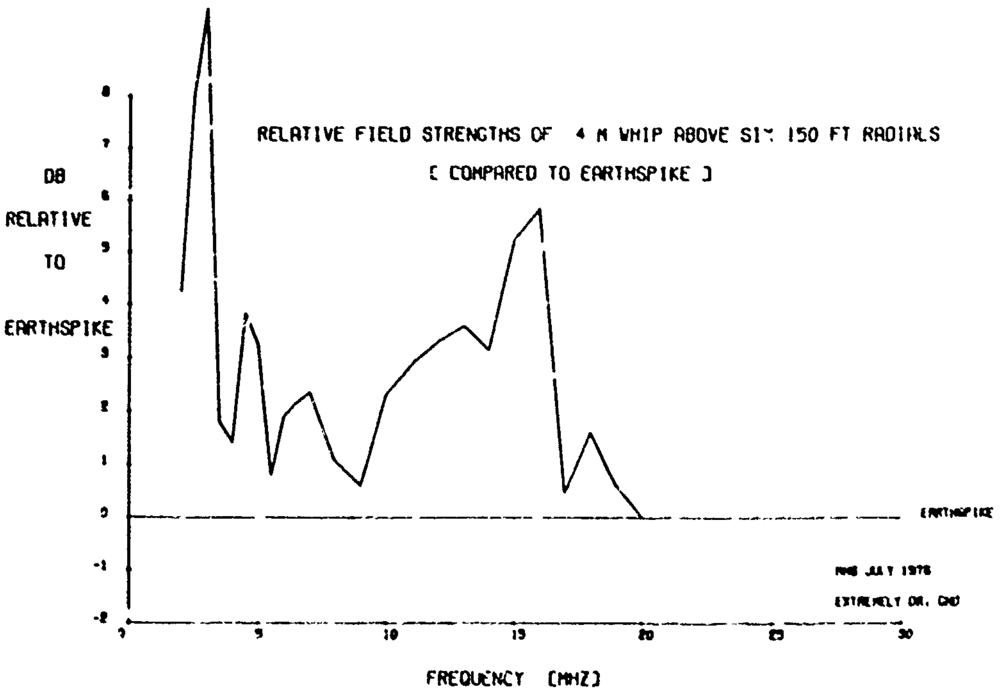

RELATIVE FIELD STRENGTHS OF 4 M WHIP ABOVE SIX 150 FT RADIALS
[COMPARED TO EARTHSPIKE]
DB
RELATIVE
TO
EARTHSPIKE
8
7
6
5
4
3
2
1
0
-1
-2
0
5
10
15
20
25
30
FREQUENCY [MHZ]
EARTHSPIKE
NHB JULY 1978
EXTREMELY DRY GND

Fig.40

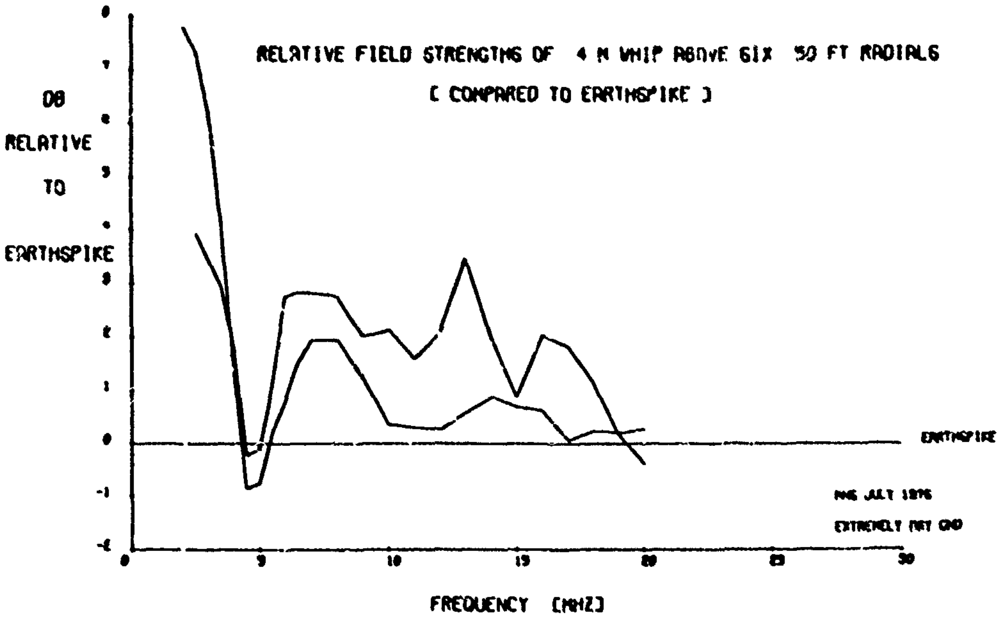

RELATIVE FIELD STRENGTHS OF 4 M WHIP ABOVE SIX 50 FT RADIALS
(COMPARED TO EARTHSPIKE)
DB
RELATIVE
TO
EARTHSPIKE
8
7
6
5
4
3
2
1
0
-1
-2
0
5
10
15
20
25
30
FREQUENCY (MHZ)
EARTHSPIKE
JULY 1976
EXTREMELY DRY GND

Fig.41

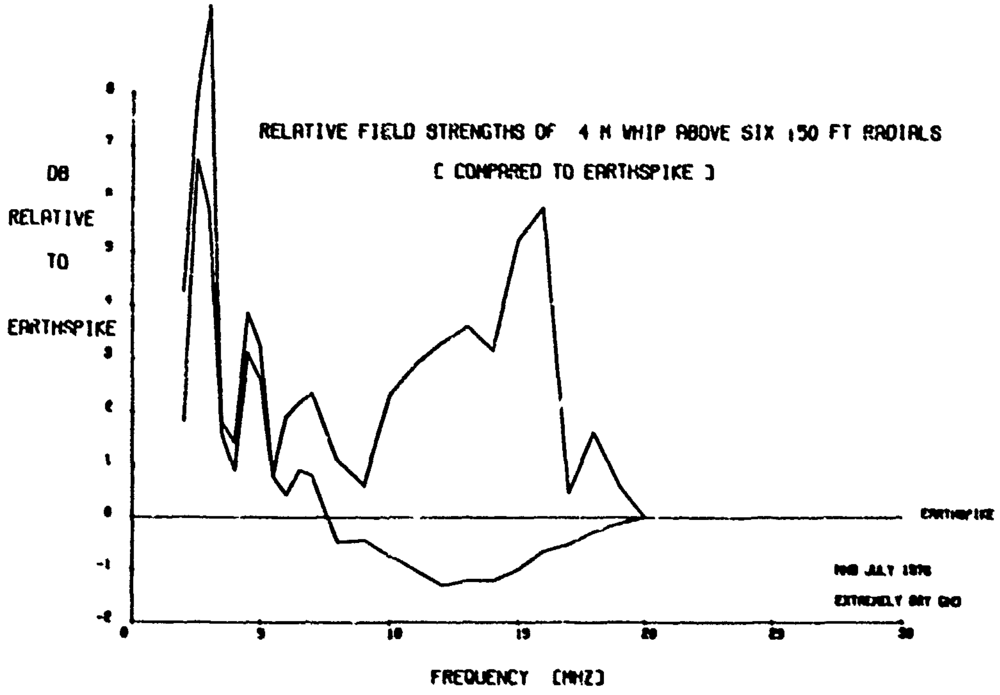

RELATIVE FIELD STRENGTHS OF 4 M WHIP ABOVE SIX 150 FT RADIALS
(COMPARED TO EARTHSPIKE)
DB
RELATIVE
TO
EARTHSPIKE
8
7
6
5
4
3
2
1
0
-1
-2
0
5
10
15
20
25
30
FREQUENCY (MHZ)
EARTHSPIKE
JULY 1976
EXTREMELY DRY GND

Fig.42

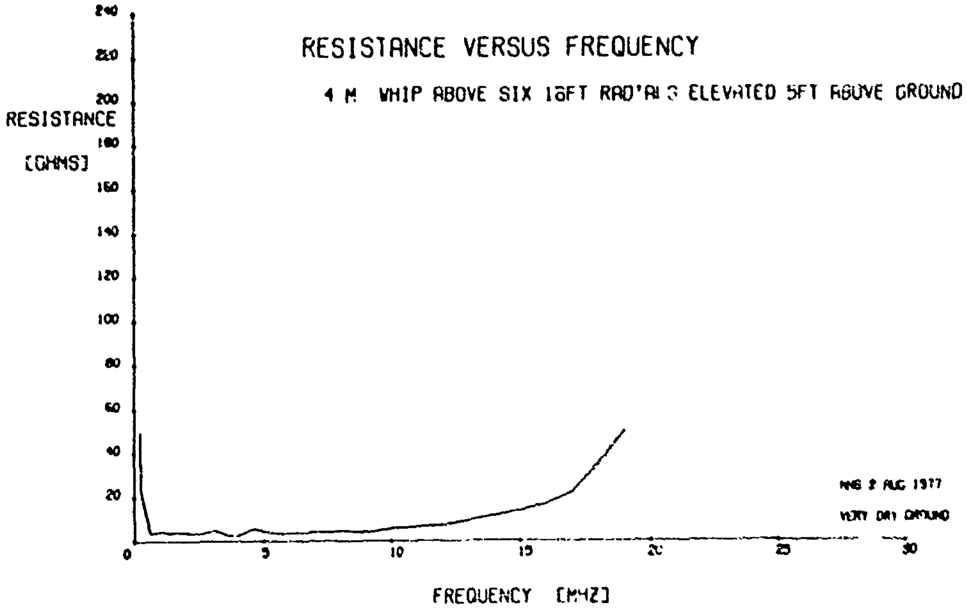

RESISTANCE VERSUS FREQUENCY
4 M WHIP ABOVE SIX 16FT RADIALS ELEVATED 5FT ABOVE GROUND
RESISTANCE
[OHMS]
240
220
200
180
160
140
120
100
80
60
40
20
0
5
10
15
20
25
30
FREQUENCY [MHZ]
MNG 2 AUG 1977
VERY DRY GROUND

Fig.43

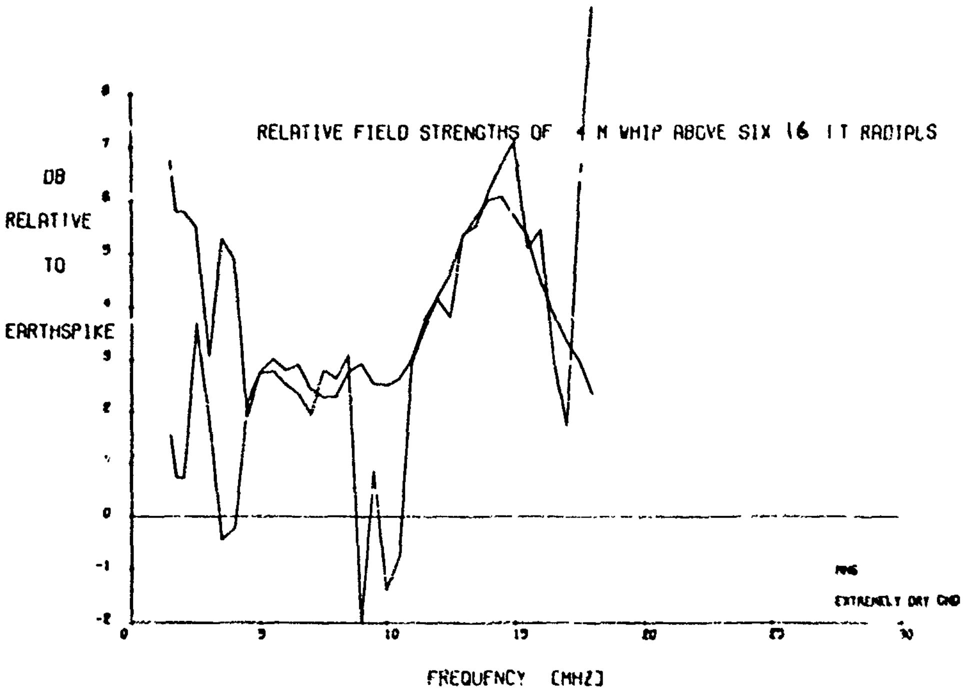

RELATIVE FIELD STRENGTHS OF 4 M WHIP ABOVE SIX 16 FT RADIALS
DB
RELATIVE
TO
EARTHSPIKE
8
7
6
5
4
3
2
1
0
-1
-2
0
5
10
15
20
25
30
FREQUENCY [MHZ]
MNG
EXTREMELY DRY GND

Fig.44

RADIO-LINK COMPUTATIONS OPTIMIZE PATTERN SHAPING OF SHORTWAVE ANTENNAS

A. Stark

Rohde & Schwarz

Munich, FRG

SUMMARY

The reliability of a radio link in the shortwave range depends to a large degree on the vertical patterns of the transmitting and receiving antennas that are used. The information necessary for dimensioning vertical radiation patterns can be derived by computer forecasts for representative HF radio links. The paper shows how statistical data can be obtained from forecast printouts to produce frequencydependent, optimized values for the elevation angles of the radiation maximum and for the vertical beamwidths.

Four kinds of antenna that are important in everyday use are then examined to see how suitable they are for applying these principles:

a) Vertical antennas, e.g. rod or whip antennas, offer no possibility of creating the vertical patterns in the desired form.
b) Horizontal dipoles, through appropriate selection of the height above ground, can be matched very well to the requirements of wave propagation. Their fixed height above ground, however, means that a compromise has to be made as regards the requirements differing with frequency.
c) Vertical logarithmic-periodic antennas permit no shaping of the [illegible] because of their construction with vertical radiators.
d) Horizontal logarithmic-periodic antennas offer ideal possibi[illegible] the elevation angle for maximum radiation plus the width of the vertical pat[illegible] very accurately matched by correct dimensioning to the values dictated by the frequency.

1. INTRODUCTION

The vertical patterns of transmitting and receiving antennas are of decisive importance for the efficiency of a shortwave radio link. The shaping of the pattern is part of the development phase and corrections to an antenna once it has been erected are only possible - if at all - at the expense of a good deal of effort, so this aspect of the design of an antenna calls for special care.

In the following it will be shown how sound statistical material for the shaping of vertical patterns is derived from the information contained in HF forecasts.

A marked advantage of the method described is that really representative data can be generated and compiled; there is no dependence on accidentally available information that could easily lead to over-assessment of special cases. That which would exceed the realm of possibility for normal measurements, is possible with a computer: in addition to times of the day and of the year, the complete sun-spot cycle is included, and radio links with distances graded simply according to aspects of propagation can be referred to in different geographical regions. The results that are obtained can be reproduced at any time and cannot of course be impaired by interference or instrumentation problems.

2. ELEVATION-ANGLE STATISTICS FROM HF FORECASTS

Standard radio forecasts as shown in Fig 1a (Süßmann, P., 1976) reveal no information about the elevation angles. Detailed computer printouts like Fig 1b and 1c, however, show the elevation angles that the radio-link computations are based on. The three blocks of figures in Fig 1b and 1c - bracketed for the sake of clarity - contain the data that are important for shaping the vertical patterns:

The uppermost block shows the frequency that is used. The middle block ANGLE gives the angle of elevation assotiated with the observed transmission path. The line REL indicates the reliability of the radio link for the particular frequency. One can at once see which combination of elevation angle and frequency leads to maximum or acceptable reliability.

These data apply for the radio link studied at a certain time. However, since the best frequencies as well as the height of the layers change with time of day (Süßmann, P., 1976), one obtains a number of combinations of frequency and elevation angle for a 24-hour period. And once one starts to consider the influence of time of the year and relative sun-spot number, it becomes clear that the relationships so important for shaping vertical patterns - even for a fixed radio link - can only be described with statistical means.

With a shortwave antenna it is generallynecessary to cover different distances. This, plus the need for designing an antenna so that it can be used in different geographical regions. means an expansion of the statistical material that is required.

Fig 2 shows how, in principle, the elevation - angle statistics that are derived then appear. If one emphasizes general-purpose use of an antenna, the median of the required elevation angles falls - as is to be expected - from high-angle radiation in the lower portion of the shortwave range andbelow to flat radiation at the higher long-haul frequencies. A characteristic feature is the initially slight dropping off that leads to maximum alteration of elevation angle with frequency in the mid-band. In the upper portion of the shortwave range the change in the median value with increasing frequency then becomes less. The width of the angular range between the limiting curves entered in Fig 2 - the lower and upper deciles - is also typical; whereas the mid-band frequencies exhibit a relatively large range of elevation angles, the range is very much constricted at higher frequencies. For practical purposes, as will be seen later, this result is by no means undesirable.

In a number of applications, e.g. with rotatable logarithmic-periodic antennas (Stark, A., 1976/1), it makes sense to limit the frequency range. This generally then also means concentrating on particular distances. Fig 3 illustrates an example for the preferred coverage of medium- to long-haul distances in round-the-clock operation with the inclusion of shorter distances during the daylight hours. Fig 4a shows elevation-angle statistics for a case such as this. As can be seen, the region of high-angle radiation was suppressed in this case by placing more emphasis on the greater distances

3. FROM STATISTICS TO VERTICAL PATTERNS

No matter whether an application calls for a radio link over any distance at any time or is tailored to a special range of distance, one always obtains from an evaluation of radio forecasts a relationship between elevation angle and frequency that looks, in principle, as in Fig 4a. The information for shaping the vertical patterns is obtained from sections through this diagram for a constant frequency. The section at f_1 entered as a broken line results in the distribution of elevation angles in Fig 4b.

For shaping the vertical pattern at this frequency f_1 one proceeds as follows: The most frequently occurring elevation angle ϑ_M is chosen as the angle for the radiation maximum. The quartile values from Fig 4b produce angles on both sides of the maximum at which a certain drop in gain, of 1 dB for example, is permissible. The decile values provide two more angles at which, for example, a drop of 3 dB as against the maximum is practicable. This is shown in Fig 4c and one can see how an ideal pattern can be sketched for this frequency.

4. PUTTING INTO PRACTICE

In the following four different antennas that are presently very commonly used in the shortwave range will be looked at to see how suitable they are for implementing the derived pattern requirements.

4.1 Vertical antennas

The vertical radiation pattern of vertical antennas with a height of $h \leq \lambda/4$, e.g. the 1.5-meter high active rod antenna in Fig 5a (Stark, A., 1977/1; Stark, A., Demmel, F., 1974), is shown in Fig 5b. Alteration of this pattern is - disregarding the conductivity of the ground - almost impossible. If the height of the antenna is increased to more than $\lambda/4$, one obtains the more strongly dipped vertical patterns of Fig 5c. Defined matching of the maximum and the pattern widths to elevation-angle statistics is not possible with vertical antennas. It can be seen that vertical antennas, besides their use for ground-wave links, are only suitable for long-haul sky-wave communications if omnidirectional characteristics are called for or higher antenna gain can be forfeited. The fact that the region of greater elevation angles cannot be covered with vertical antennas is seen, among other things, in the well-known zone of silence between the areas of ground waves and sky waves.

4.2 Horizontal dipoles

The vertical patterns of horizontal dipoles alter considerably in relation to the height of the antenna above ground. The significance of this kind of antenna in the shortwave range stems from the fact that with suitable height above ground the dependence on frequency of the patterns as dictated by propagation can be achieved relatively well. As Fig 6a shows, heights can be found at which one obtains high-angle radiation in the lower portion of the operating-frequency range and a smooth transition to the flatter angles of departure needed for the higher frequencies.

Horizontal dipoles are thus widely used in shortwave applications, for reception, for instance, in the form of active antennas (Stark, A., 1977/1; Stark, A., Demmel, F., 1974), and enabling for the first time the realization of electrically short, omnidirectional receiving antennas for horizontal polarization as shown in Fig 6b (Stark, A., 1977/2; Stark, A., 1979/1).

In the past the use of horizontal dipoles for transmitting purposes was often frustrated by the fact that the antennas were of large dimensions and had to be rigged between two or three masts. Fig 6c illustrates a new version, for transmitting power of up to 1 kW, that surmounts these drawbacks (Stark, A., 1978):

This new horizontal dipole is erected on a single supporting mast, meaning that it can be employed even when available space is extremely scarce. This design was made possible by the low overall length of the dipole of only 10 meters, the prerequisite for this being the tuning of the antenna over a tuning unit integrated into the feedpoint of the dipole. As a result of the low-loss design of the transforming circuit - enabling the preprogramming of any six frequencies in the range 2 to 30 MHz - higher degrees of efficiency are attained, despite the small dimensions, than with the loaded broadband

26 meters in length that were often used in the past (Stark, A., 1979/2).

4.3 Vertical logarithmic-periodic antennas

Being extremely broadband directional radiators, logarithmic-periodic (LP) antennas have attained unusually large significance in radio and radiomonitoring. The versions for vertically polarized waves, as shown, for example, in Fig 7a for the range 1.5 to 30 MHz, consist of monopoles enabling minimal erection heights. The vertical patterns result from those of the individual vertical antenna according to 4.1 through multiplication by a frequency-independent mathematical factor; the result in Fig 7b therefore shows, apart from the unilateral directional effect, the same principal behaviour as the vertical antenna: the large elevation angles that could not be reached with the single antenna element cannot be covered by combining to form an LP antenna either (Stark, A., 1977/3). The differences between the two patterns shown in Fig 7b are merely a result of the increase in the attenuation of the real ground with rising frequency. Shaping the vertical patterns after the derived statistics is not possible with vertical antennas for general-purpose use. This also applies to vertical LP antennas that use dipoles as individual radiators. The dipole also exhibits a null in the direction of its axis, so the regions of greater elevation angle are impractical in this case too. Shaping of the vertical pattern according to frequency is only possible for the very small elevation angles, meaning a restriction to long-haul radiocommunications.

4.4 Horizontal logarithmic-periodic antennas

For understanding the excellent possibilities for shaping the vertical patterns of this category of antenna, one has to look a little more closely at the way in which logarithmic-periodic radiators work. Fig 8 shows the familiar arrangement of dipoles, the lengths of which and the intervals between them obey well-known mathematical principles (Carell, R., 1961; Isbell, D.E., 1960). When looking at the vertical patterns it is important to note that only a group of a few radiators is effective for a particular frequency within the operating-frequency range. Such a group of generally three to five adjacent dipoles is limited toward the rear by an element of about half a wavelength. If the frequency is increased, the effective zone shifts in the direction of the antenna tip, towards shorter radiators.

Fig 9a illustrates such an LP antenna arrangement; the intended tip of the antenna is at the level of the ground. It can be shown that the height above ground in relation to the wavelength is always constant: at the lower frequency limit f_1 the dipole with a length of $L_1 \approx \lambda_1/2$ has an absolute height of h_1 above ground, i.e. a relative - referred to the operating wavelength - height above ground of, for example, $h_1/\lambda_1 = 0.25$. At a frequency that is twice as high, $f_2 = 2f_1$, the effective zone will be found to have shifted to a dipole of half the length $L_2 \approx \lambda_2/2 = L_1/2$. This dipole is at a height above ground - as a simple geometrical calculation shows - that is exactly half of h_1; referring to the smaller wavelength λ_2 one therefore has again the same relative value of $h_2/\lambda_2 = 0.25$. The free-space patterns of LP antennas are independent of frequency. One therefore obtains with this arrangement, as a result of the constant relative height above ground, vertical patterns as in Fig 9b that are constant over the entire operating-frequency range if the antenna is regarded as being above an ideally conducting plane. If one considers the damping by the real ground, which rises with frequency, the result is even of a contrary sense to the behaviour required from the point of view of propagation: the maximum of the vertical pattern shifts with increasing frequency to somewhat larger elevation angles.

Constant vertical patterns with horizontal LP antennas can be achieved in the manner described with little effort; however, they are generally contrary to the requirements of wave propagation and cannot lead to optimum reliability of a radio link.

If the tip of the logarithmic-periodic structure is raised from the ground (Fig 10a), the relative height above ground increases with frequency. The maximum of the vertical pattern is thus dropped towards higher frequencies and the width of the pattern decreases (Fig 10b). The correct alteration of the vertical patterns with increasing frequency as regards propagation is therefore achieved. Horizontal LP antennas of this kind (Fig 10c) are somewhat more elaborate than what was described beforehand because a mast at the antenna tip of appreciable height is called for (Stark, A., 1978). However, through optimization of the structural parameters τ and α and of the height above ground, vertical patterns are possible with this arrangement whose dependence on frequency corresponds exactly to the requirements of the elevation-angle statistics (Stark, A., 1977/3).

The rotatable LP antenna (Stark, A., 1976/1) shown in Fig 11a can be taken as an example. With an operating-frequency range of 5 to 30 MHz it is designed primarily for medium- and long-haul links, shorter distances upwards of about 200 km being possible during the day (Fig 3). The shaping of the vertical patterns was therefore based on the elevation-angle statistics shown in Fig 4a. The radiation patterns obtained are reproduced in Fig 11b These show that:

a) The elevation angles of the maxima for all frequencies lie exactly at the 50 % values of Fig 4a.

b) The pattern widths could be well adapted to the spacing between the 10 % and 90 % curves of Fig 4a. The corresponding region of the vertical patterns is shown by broken lines in Fig 11b.

These radiation patterns apply when the antenna is above ideally conducting ground. Investigations have shown that the alterations of the elevation angles and pattern widths occurring as a result of real ground conductivity can be neglected in genuine use (Stark, A., 1976/2).

5. CONCLUSION

For achieving high reliability of a shortwave radio link, the vertical radiation patterns of transmitting and receiving antennas are of primary importance. It is shown how statistical information can be derived from radio forecasts to form the basis of vertical patterns that are optimally shaped from the point of view of propagation.

The practicability of these data is very much dependent on the type of antenna concerned: vertical antennas and vertical logarithmic-periodic antennas are useful solutions for ground-wave links and when one is restricted to long-haul traffic. Their vertical patterns, however, cannot be adapted to the dependence on frequency, dictated by propagation, of the angle of maximum fire and the pattern width. For sky-wave communications over short and medium distances, horizontal antennas present substantially better characteristics. The horizontal dipole can be well matched to the conditions of transmission by being at a suitable height above ground. Its limitations are to be found in the constant height above ground that exists for all frequencies, this calling for a trade off between the differing requirements according to frequency. Horizontal logarithmic-periodic antennas offer ideal possibilities for shaping vertical patterns in line with propagation. A different zone of the antenna becomes effective depending on the frequency, so within certain limits it is possible to select the best height above ground in every case. By combining appropriate free-space patterns with a suitable height above ground, vertical patterns that are optimally shaped from the point of view of propagation can be achieved throughout the shortwave range.

6. ACKNOWLEDGEMENTS

The author wishes to thank the research group "Ionosphere" of the research institute of the Deutsche Bundespost for its assistance with forecast calculations.Special thanks are due to Dr. Damboldt and Mr. Süssmann for their encouragement in preparing this lecture and for the many interesting and useful hours of discussion spent with them.

REFERENCES

Carrel, R., 1961, "The Design of Log-Periodic Dipole Antennas" IRE Intern. Conv. Record pt 1, 61 - 74.

Isbell, D.E., 1960, "Log-Periodic Dipole Arrays", IRE Transactions on Antennas and Propagation, 260 - 267.

Stark, A.; Demmel, F., 1974 "Active Receiving Antennas for 1.5 to 30 MHz", News from Rohde & Schwarz 64, 14 - 17.

Stark, A., 1977/1,"Dimensioning electrically short receiving antennas for frequencies below 30 MHz", News from Rohde & Schwarz 76, 23 - 26.

Stark, A., 1977/2, "Active HF omnidirectional receiving antennas HE 004 and HE 005 for horizontal polarization", News from Rohde & Schwarz 78, 7 - 9.

Stark, A., 1977/3, "Dimensioning Log-Periodic Shortwave Antennas Using Computer Analysis", Communications international,Volume 4/Number 6/June 1977, 39 - 46.

Stark, A., 1976/1, "Rotatable and Lightweight Log-Periodic HF Antennas", News from Rohde & Schwarz 72, 8 - 10.

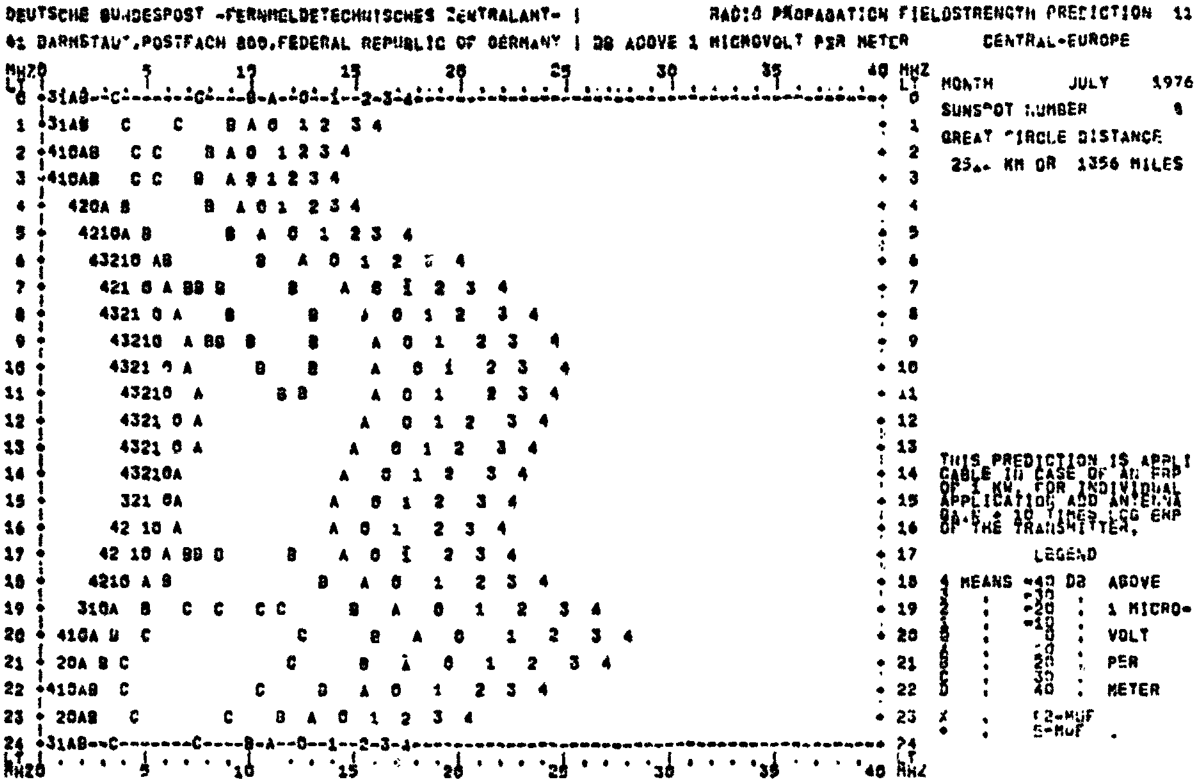
Stark, A., 1976/2, "Effect of earth's surface on vertical patterns and gain of horizontally polarized HF logarithmic-periodic antennas", News from Rohde & Schwarz 75, 15 - 19.

Stark, A., 1978, "Further developments in shortwave antennas", News from Rohde & Schwarz 81, 19 - 21.

Stark, A., 1979/1, "Reception of high-angle HF with active omnidirectional antennas HE 004" News from Rohde & Schwarz 84, 18 - 20.

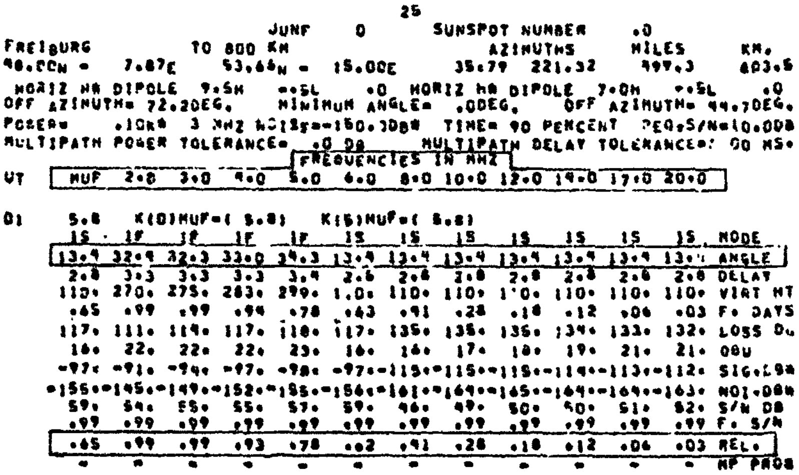
Stark, A., 1979/2, "Better Antenna Isolation with HF-Dipole HK 007", News from Rohde & Schwarz 85, 10 - 13.

Süßmann, P., "Die weltweite Ausbreitung von Kurzwellen über die Ionosphäre und ihre Vorhersage", Nachrichtentechnische Zeitschrift 29 (1976) H. 5, 394 - 399.



```
DEUTSCHE BUNDESPOST -FERNMELDETECHNISCHES ZENTRALAMT- I          RADIO PROPAGATION FIELDSTRENGTH PREDICTION  11
61 DARMSTADT,POSTFACH 800,FEDERAL REPUBLIC OF GERMANY I DB ABOVE 1 MICROVOLT PER METER      CENTRAL-EUROPE
MHZ0        5        10        15        20        25        30        35        40 MHZ
LT I                                                                                LT   MONTH        JULY    1976
 0 +31AB--C-------C----B-A--0--1--2-3-4-------------------------------------------+ 0    SUNSPOT NUMBER       8
 1 +31AB   C    C    B A 0  1 2  3 4                                               + 1    GREAT CIRCLE DISTANCE
 2 +410AB   C C    B A 0  1 2 3 4                                                  + 2     25.. KM OR  1356 MILES
 3 +410AB   C C   B  A 0 1 2 3 4                                                   + 3
 4 +  420A B       B  A 0 1  2 3 4                                                 + 4
 5 +   4210A B       B  A  0  1  2 3  4                                            + 5
 6 +    43210 AB        B   A  0  1  2  3  4                                       + 6
 7 +     421 0 A BB B       B    A  0  1  2  3  4                                  + 7
 8 +     4321 0 A    B       B    A  0  1  2   3  4                                + 8
 9 +      43210  A BB  B      B     A  0  1   2  3   4                             + 9
10 +      4321 0 A      B    B     A   0  1   2  3   4                             + 10
11 +       43210  A       B B      A  0  1    2  3  4                              + 11
12 +       4321 0 A               A   0  1  2   3  4                               + 12
13 +       4321 0 A              A   0  1  2   3  4                                + 13   THIS PREDICTION IS APPLI
14 +       43210A               A   0  1  2   3  4                                 + 14   CABLE IN CASE OF AN ERP
15 +       321 0A              A   0  1  2   3  4                                  + 15   OF 1 KW. FOR INDIVIDUAL
16 +      42 10 A              A  0  1   2  3  4                                   + 16   APPLICATION ADD ANTENNA
17 +     42 10 A BB 0       B    A  0  1   2  3  4                                 + 17   GAIN + 10 TIMES LOG ERP
18 +    4210 A B                B   A  0   1   2  3  4                             + 18   OF THE TRANSMITTER.
19 +   310A  B   C  C   C C      B   A   0   1   2   3  4                          + 19           LEGEND
20 + 410A B  C              C      B   A   0   1   2   3  4                        + 20   4 MEANS -40 DB  ABOVE
21 + 20A B C               C      B   A   0   1   2   3  4                         + 21   3   :   -30  :
22 +410AB  C             C     B  A  0   1   2  3  4                               + 22   2   :   -20  :  1 MICRO-
23 + 20AB   C         C    B  A  0  1  2  3  4                                     + 23   1   :   -10  :
24 +31AB--C-------C----B-A--0--1--2-3-4---------------------------------------------+ 24   0   :    0   :  VOLT
LT I     .   5   .    10   .    15   .    20   .    25   .    30   .    35   .    40 LT   A   :   10   :
MHZ0                                                                              MHZ   B   :   20   :  PER
                                                                                        C   :   30   :
                                                                                        D   :   40   :  METER
                                                                                        X   :   F2-MUF
                                                                                        +   :   E-MUF   .
```

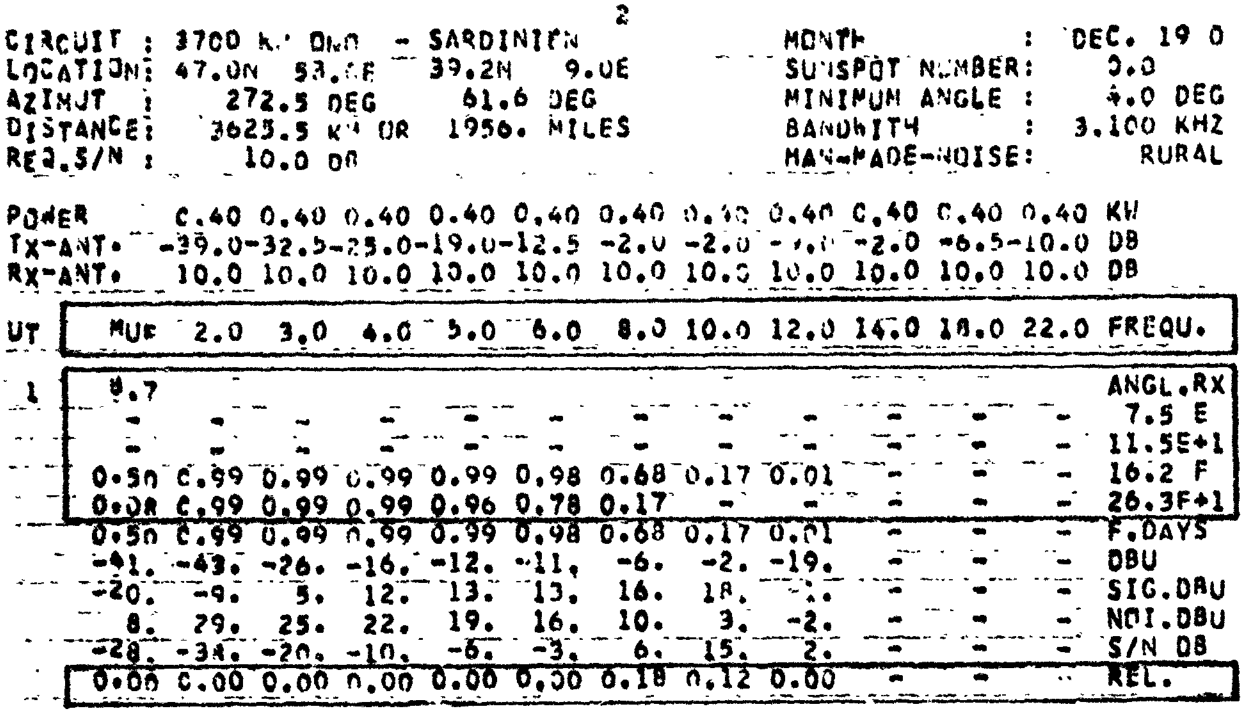
Fig.1(a) Standard radio forecast for the HF-range



```
                                   25
                          JUNE    0         SUNSPOT NUMBER      .0
FREIBURG          TO 800 KM                      AZIMUTHS      MILES       KM.
48.00N -    7.87E     53.66N -  15.00E      35.79  221.32     497.3      803.5
 HORIZ HW DIPOLE  9.5H    -.5L     .0  HORIZ HW DIPOLE  7.0H    -.5L     .0
OFF AZIMUTH= 72.2DEG.     MINIMUM ANGLE=  .0DEG.     OFF AZIMUTH= 44.7DEG.
POWER=    .10KW  3 MHZ NOISE=-150.0DBW  TIME= 90 PERCENT  REQ.S/N=10.0DB
MULTIPATH POWER TOLERANCE=  .0 DB     MULTIPATH DELAY TOLERANCE= .00 MS.
                               FREQUENCIES IN MHZ
UT     MUF   2.8   3.0   4.0   5.0   6.0   8.0  10.0  12.0  14.0  17.0  20.0

01     5.8   K(D)MUF=( 5.8)   K(S)MUF=( 5.8)
       1S    1F    1F    1F    1F    1S    1S    1S    1S    1S    1S    1S  MODE
      13.4  32.4  32.3  33.0  34.3  13.4  13.4  13.4  13.4  13.4  13.4  13.4 ANGLE
       2.8   3.3   3.3   3.3   3.4   2.6   2.8   2.8   2.8   2.8   2.8   2.8 DELAY
      110.  270.  275.  283.  299.  110.  110.  110.  110.  110.  110.  110. VIRT HT
       .65   .99   .99   .94   .78   .63   .41   .28   .18   .12   .06   .03 F. DAYS
      117.  111.  114.  117.  118.  117.  135.  135.  136.  134.  133.  132. LOSS DB
       16.   22.   22.   22.   23.   16.   16.   17.   18.   19.   21.   21. DBU
      -97.  -91.  -94.  -97.  -98.  -97. -115. -115. -115. -114. -113. -112. SIG.DBW
     -155. -145. -149. -152. -155. -156. -161. -164. -165. -164. -164. -163. NOI.DBW
       59.   54.   55.   55.   57.   59.   46.   49.   50.   50.   51.   52. S/N DB
       .99   .99   .99   .99   .99   .99   .99   .99   .99   .99   .99   .99 F. S/N
       .65   .99   .99   .93   .78   .62   .41   .28   .18   .12   .06   .03 REL.
        -     -     -     -     -     -     -     -     -     -     -     -  HF PROB
```

Fig.1(b) More detailed HF-propagation prediction



```
                                        2
CIRCUIT : 3700 KM DMO  - SARDINIEN                MONTH             :  DEC. 19 0
LOCATION: 47.0N  53.0E   39.2N    9.0E            SUNSPOT NUMBER:      0.0
AZIMUT  :     272.5 DEG     61.6 DEG              MINIMUM ANGLE :      4.0 DEG
DISTANCE:  3625.5 KM OR  1956. MILES              BANDWITH      :    3.100 KHZ
REQ.S/N :      10.0 DB                            MAN-MADE-NOISE:        RURAL

POWER      0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40 0.40 KW
TX-ANT.   -39.0-32.5-25.0-19.0-12.5 -2.0 -2.0 -[illegible] -2.0 -6.5-10.0 DB
RX-ANT.    10.0 10.0 10.0 10.0 10.0 10.0 10.0 10.0 10.0 10.0 10.0 DB

UT    MUF  2.0  3.0  4.0  5.0  6.0  8.0 10.0 12.0 14.0 18.0 22.0 FREQU.

 1    8.7                                                        ANGL.RX
       -    -    -    -    -    -    -    -    -    -    -    -   7.5 E
       -    -    -    -    -    -    -    -    -    -    -    -   11.5E+1
     0.50 0.99 0.99 0.99 0.99 0.98 0.68 0.17 0.01   -    -    -   16.2 F
     0.08 0.99 0.99 0.99 0.96 0.78 0.17   -    -    -    -    -   26.3F+1
     0.50 0.99 0.99 0.99 0.99 0.98 0.68 0.17 0.01   -    -    -   F.DAYS
     -41. -43. -26. -16. -12. -11.  -6.  -2. -19.   -    -    -   DBU
     -20.  -9.   5.  12.  13.  13.  16.  18. [illegible]   -    -    -   SIG.DBU
       8.  29.  25.  22.  19.  16.  10.   3.  -2.   -    -    -   NOI.DBU
     -28. -38. -20. -10.  -6.  -3.   6.  15.   2.   -    -    -   S/N DB
     0.00 0.00 0.00 0.00 0.00 0.00 0.18 0.12 0.00   -    -    -   REL.
```

Fig.1(c) Detailed HF-propagation prediction developed by the FTZ (FRG)

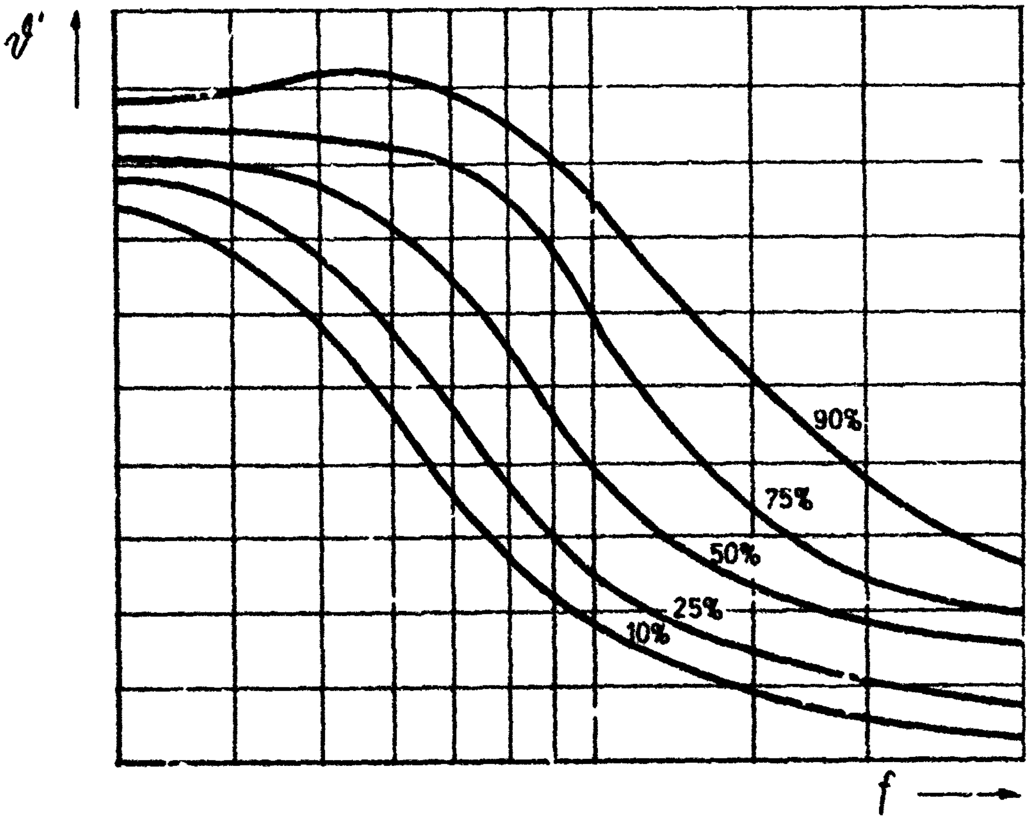


Fig.2 Elevation angle versus frequency

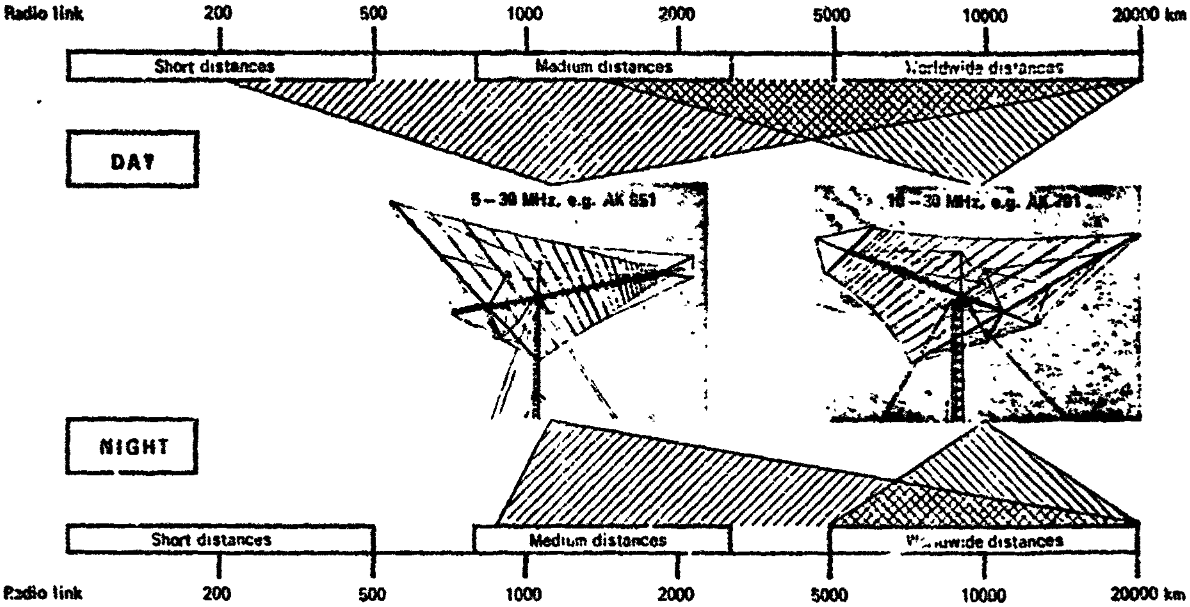

Radio link
200
500
1000
2000
5000
10000
20000 km
Short distances
Medium distances
Worldwide distances
DAY
5 – 30 MHz, e.g. AK 851
10 – 30 MHz, e.g. AK
NIGHT
Short distances
Medium distances
Worldwide distances
Radio link
200
500
1000
2000
5000
10000
20000 km

Fig.3 Distance ranges that can be covered with the frequency ranges of 5 -10 MHz and 10–30 MHz (typical ranges of rotatable log.-periodic antennas

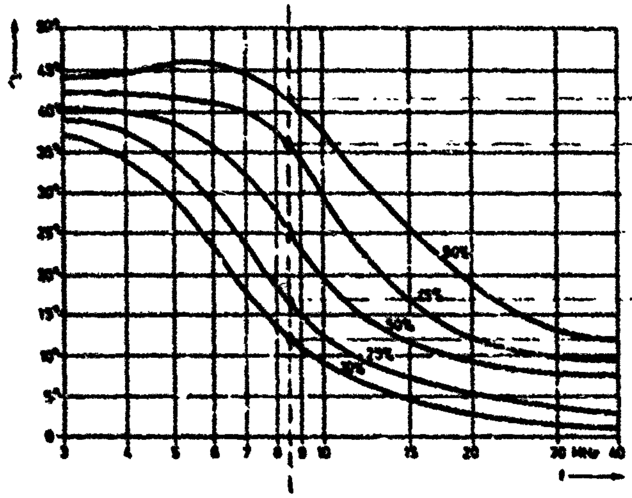


Fig.4(a) Elevation-angle statistics for medium- and long-haul communication

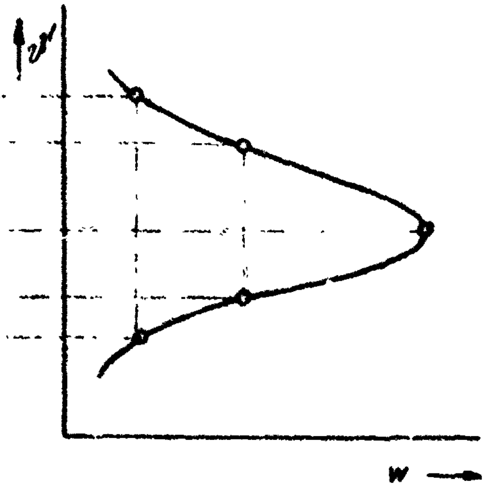


Fig.4(b) Probability distribution of elevation-angles for frequency f_1

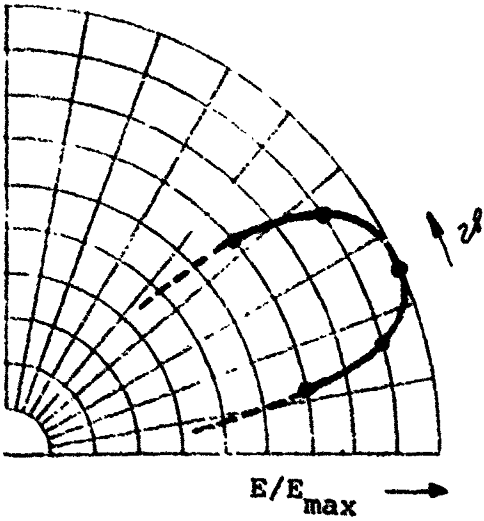


Fig.4(c) Shaping of the vertical pattern according to the distribution of elevation-angles

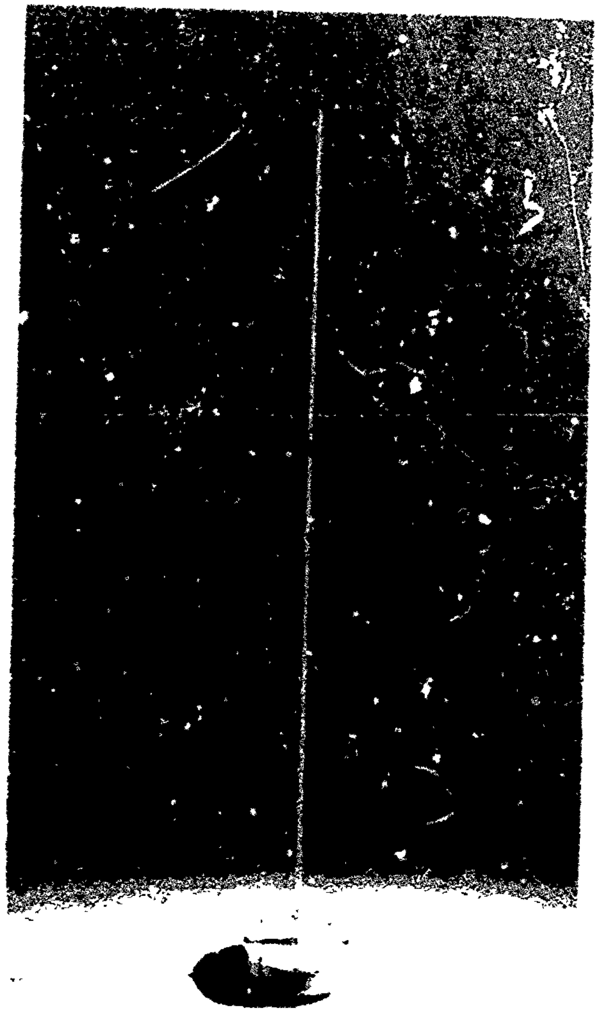


Fig 5(a) Active HF [illegible]

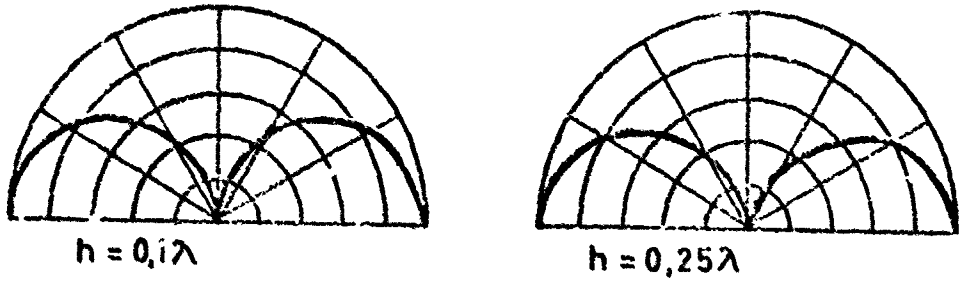


Fig.5(b) Vertical patterns of vertical antennas with a height h ≤ λ/4

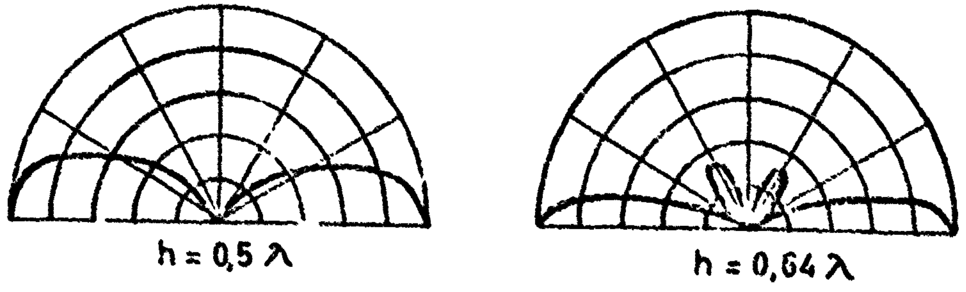


Fig.5(c) Vertical patterns of vertical antennas with a height h > λ/4

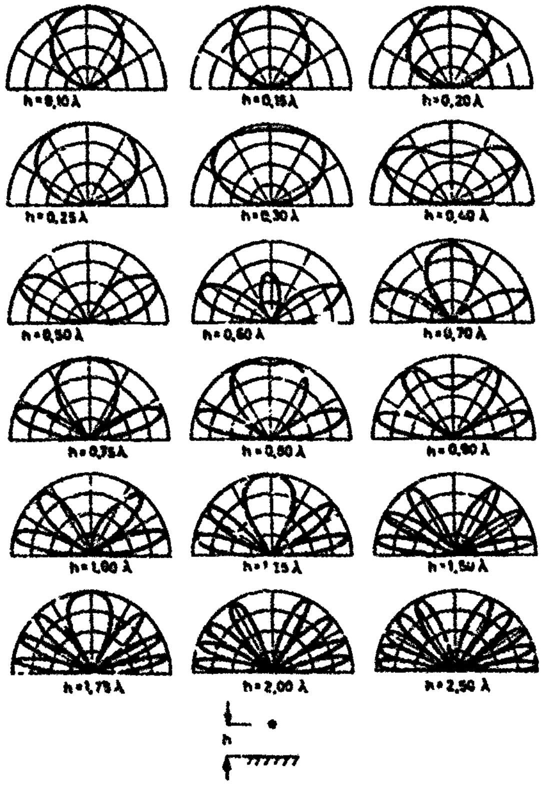


Fig.6(a) Vertical patterns of horizontal dipole over perfectly conducting ground in the plane perpendicular to the axis of the dipole

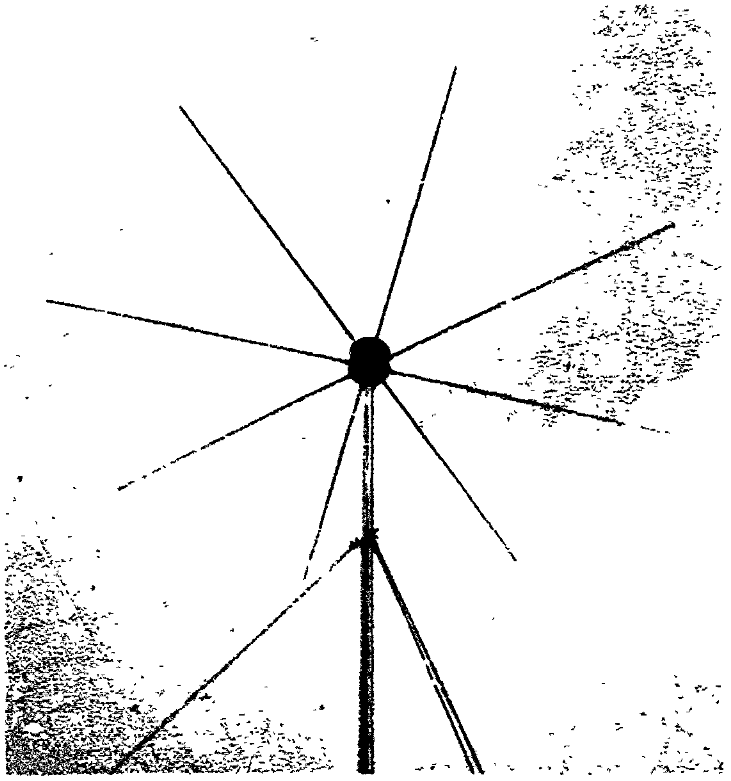


Fig.6(b) Active HF turnstile dipole HE 004

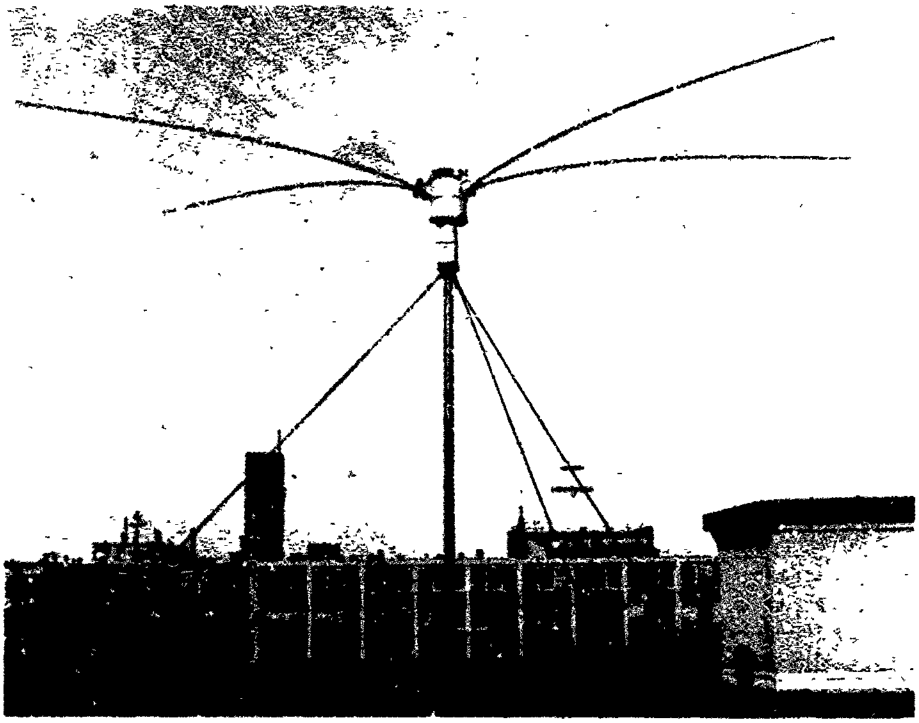


Fig 6(c) Single-mast mounted HF transmitting dipole HK 007 for 6 channels in the 2–30 MHz range

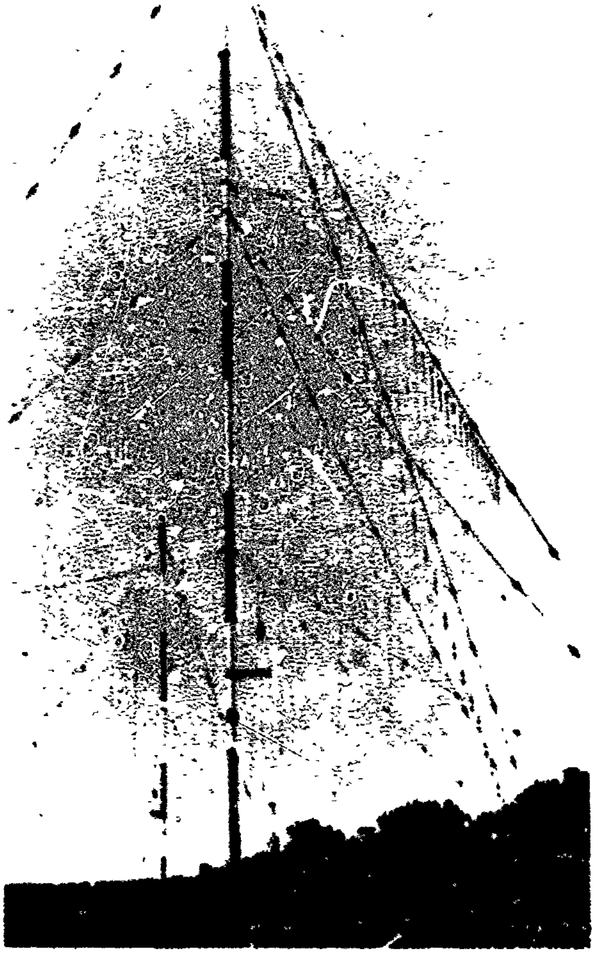
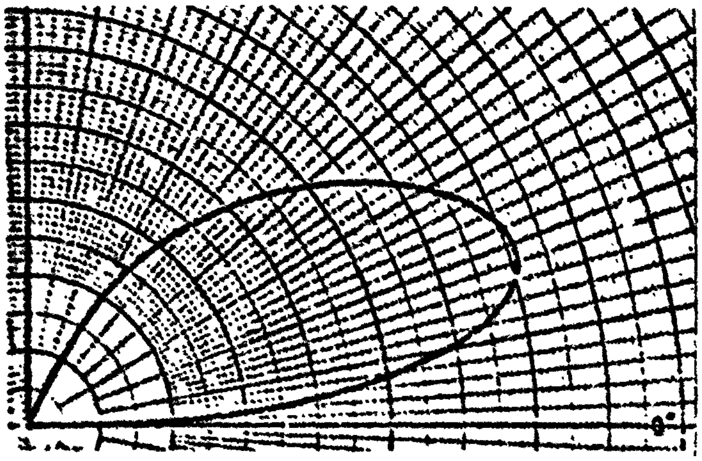
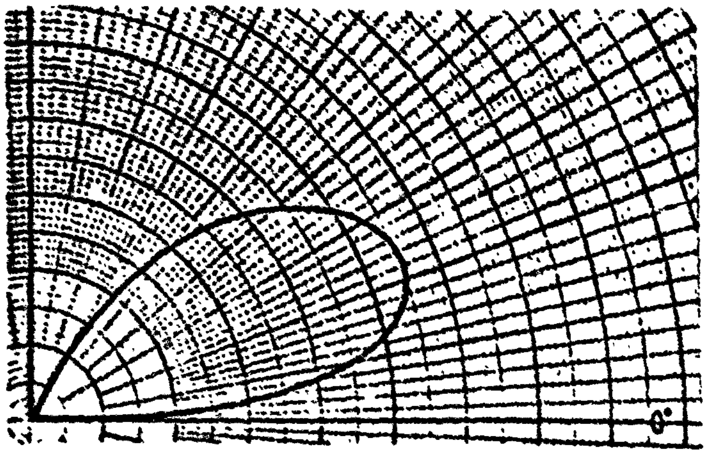


Fig.7(a) Vertical log.-periodic antenna AK 210 for 1.5–30 MHz.



f = 4 MHz

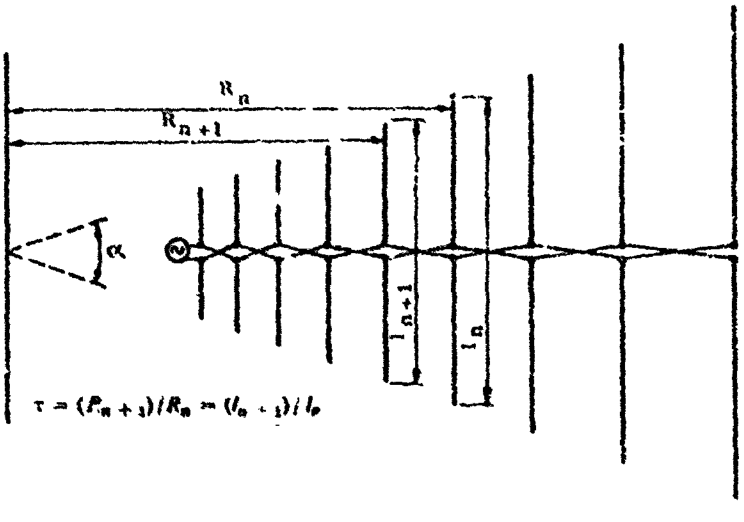


f = 30 MHz

Bodenkonstanten:

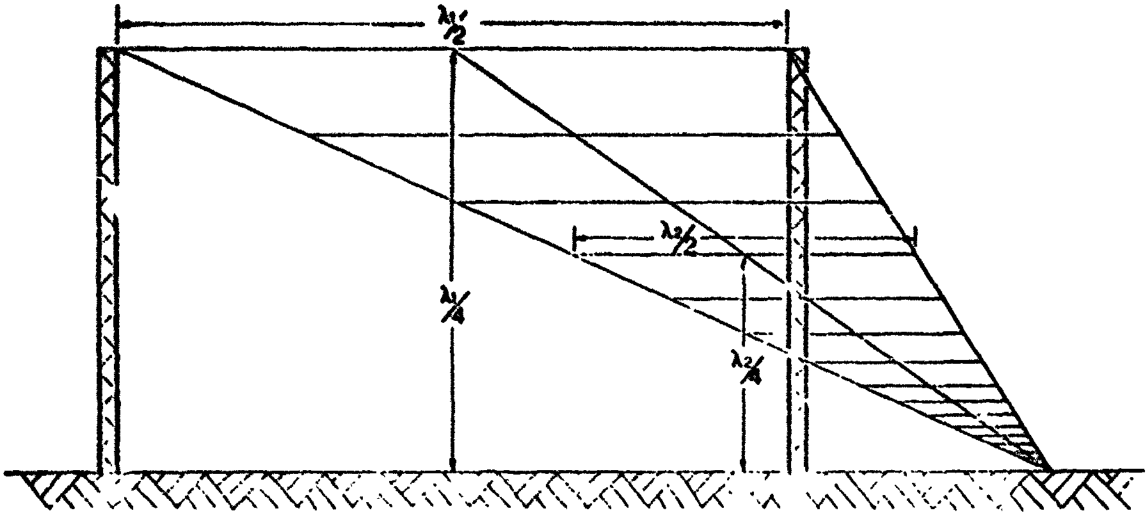
$$\kappa = 10^{-2}\ \frac{1}{\Omega \cdot m} \quad ; \quad \varepsilon_r = 10$$

Fig.7(b) Vertical patterns of vertical log.-periodic antenna AK 210



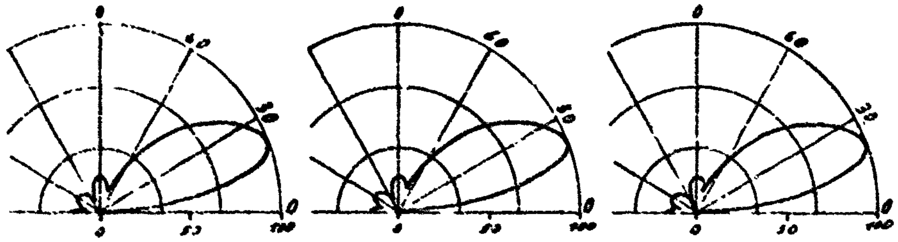
$\tau = (R_{n+1})/R_n = (l_{n+1})/l_n$

Fig.8 Log.-periodic dipole antenna (principle sketch)



$\lambda_1/2$
$\lambda_1/4$
$\lambda_2/2$
$\lambda_2/4$

Fig.9(a) LP antenna with tip at the level of the ground



3 MHz

10 MHz

18 MHz

Fig.9(b) Vertical patterns of LP antenna arranged according to Figure 9(a)

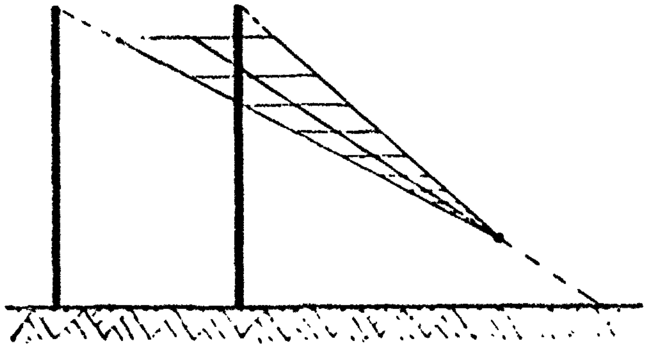
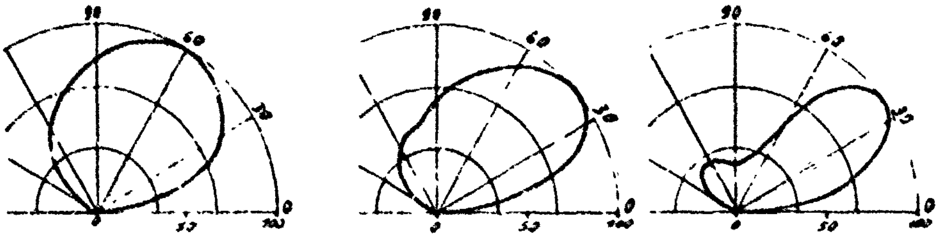


Fig 10(a) LP antenna with elevated tip



3 MHz 10 MHz 18 MHz

Fig 10(b) Vertical patterns of LP antenna arranged according to Figure 10(a)

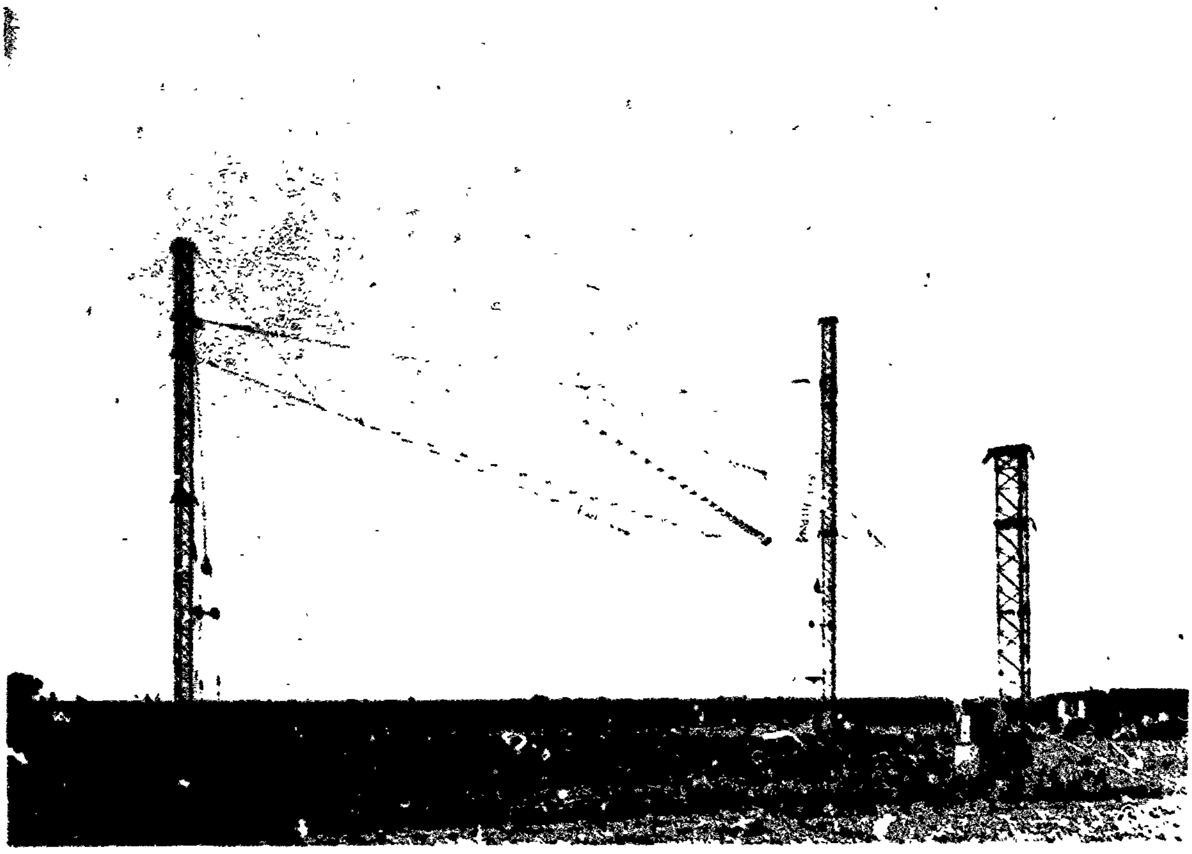


Fig.10(c) LP antenna AK 440 for 4 30 MHz

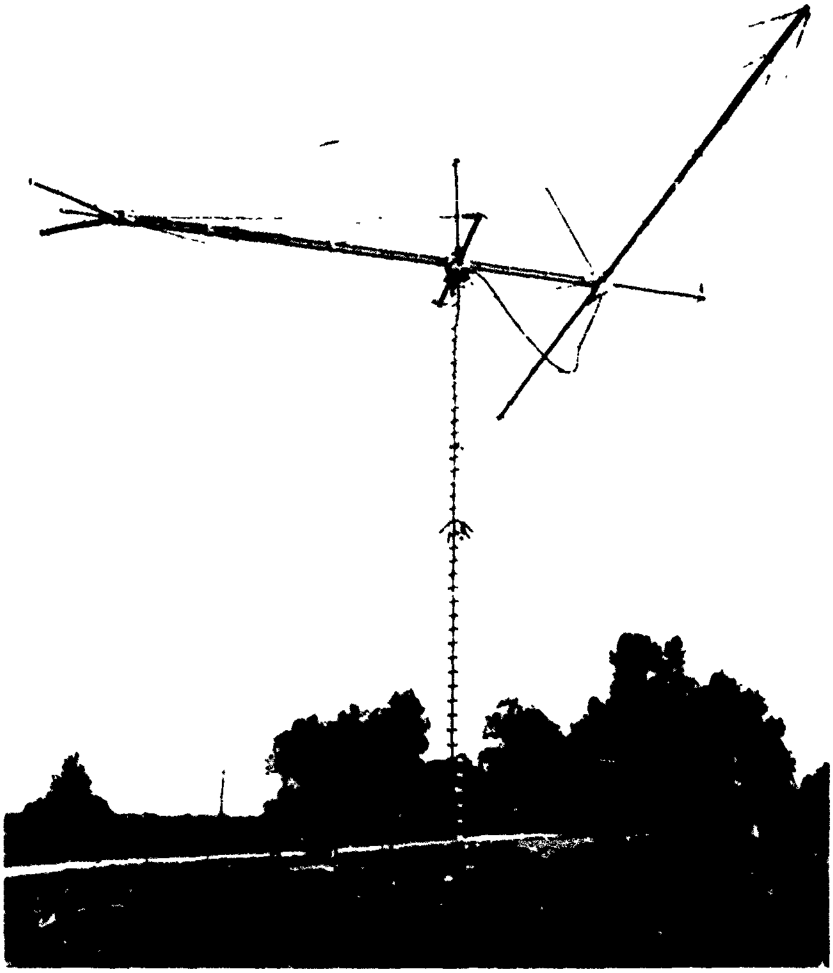


Fig 11(a) Rotatable LP antenna AK 853 for 5 - 30 MHz

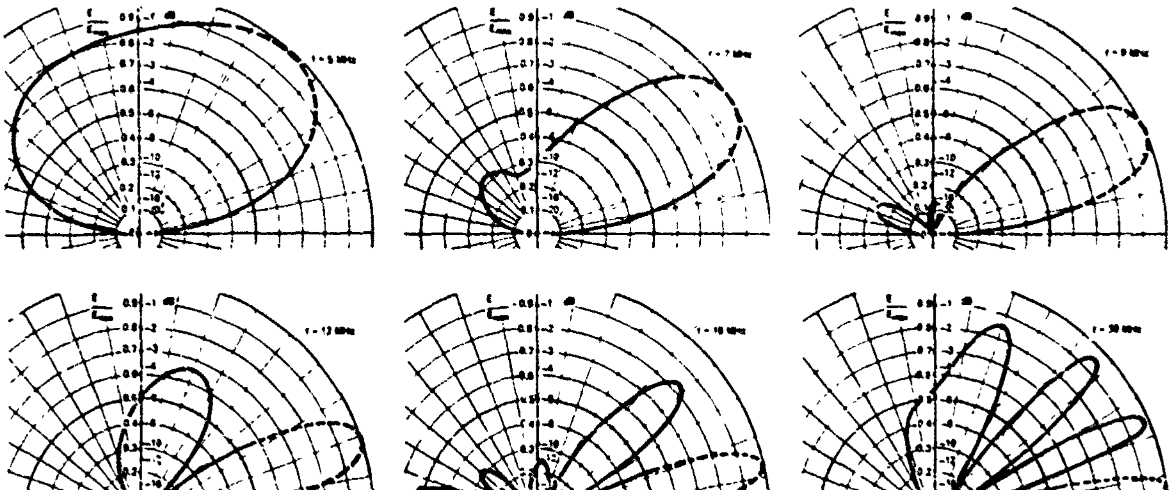


Fig 11(b) Vertical patterns of LP antenna AK 853

DISCUSSION

Comment by G. May, UK

Whilst agreeing on the importance of the elevation angle definition of the horizontal log periodic antenna that you have described I would like to emphasise the need to consider azimuthal directivity. In our practical tests we find that the horizontal log periodic antenna that we use for aircraft communications at ranges of 2000 km is quite inadequate because of its lack of directivity. Our Beverage antenna outperforms the horizontal periodic array.

Author's Comment

I agree on the importance of azimuthal directivity as signal-to-noise ratio is proportional to directivity under certain boundary conditions. So an antenna with 20dB directivity (e.g. a Beverage antenna) is better than a 13dB log. periodic – as long as the signal is arriving exactly from that direction where the pattern maximum points to. Great circle deviations, however, change this result considerably. Due to such propagation effects a signal will be picked up less often with maximum directivity the smaller the horizontal beamwidth is. For a certain azimuthal deviation the reduction in directivity is the greater, the smaller the beamwidth is. So we soon reach an angle of deviation where the reduced directivity of the Beverage antenna is smaller than the LP directivity that is less affected due to broader beamwidth.

The attached sketch gives some examples from the research work that we have carried out on this subject (the propagation data shown are taken from H.A. Whale, Radio Research Centre, University of Auckland). It can be seen that there are times where the narrow beam antenna outperforms the log periodic, but there are also times where the contrary is true. A compromise between directivity and availability of a certain signal-to-noise ratio is therefore needed. Taking into account these statistical considerations LP antennas have proven to offer increased availability and reliability of HF links also at several places for aircraft communication.

T. Elkins, US

The ITS-78 Propagation Prediction program produces significant elevation angle errors, as compared to accurate modelling and ray tracing techniques, as the frequency approaches the MUF. Does your experience support this and if so, have your found a reasonably simple solution?

Author's Reply

I think that these errors do not affect the results very much as we only consider the frequency with the highest reliability which often occurs considerably below the MUF. In addition a safety factor can be used by which the frequency has to be below the MUF (this safety or correction factor can depend on several parameters).

D. Scholz, Ge

The primary advantage of high elevation angle patterns lies in the preference given to maximum radiation for short and medium haul sky-wave communications. The question aims at an adverse effect: To what extent it can be expected that a certain amount of radiation from far-distant transmitters is suppressed?

This would render feasabie the manipulation of interference potential generated in undesired or irrelevant distance ranges -- an interesting aspect from frequency management and a military point of view.

Author's Reply

This is a very interesting aspect of correct pattern shaping for short and medium-haul links as well as for any other distance range. When pointing the vertical pattern to elevation angles needed for a certain distance range we reduce the antenna gain in unwanted directions. Therefore interfering signals arriving from other distance ranges are received with reduced sensitivity. To what extent a signal is suppressed depends, of course, on the deviation from the angle of maximum fire and on the shape of the pattern and therefore on the operating frequency, but 10dB or more can be achieved in many cases. Using the same pattern shaping at the transmitting station may reduce the interfering field strength in addition and can double this value.

Furthermore this effect is also very useful for increasing the security against unwanted radio reconnaissance.

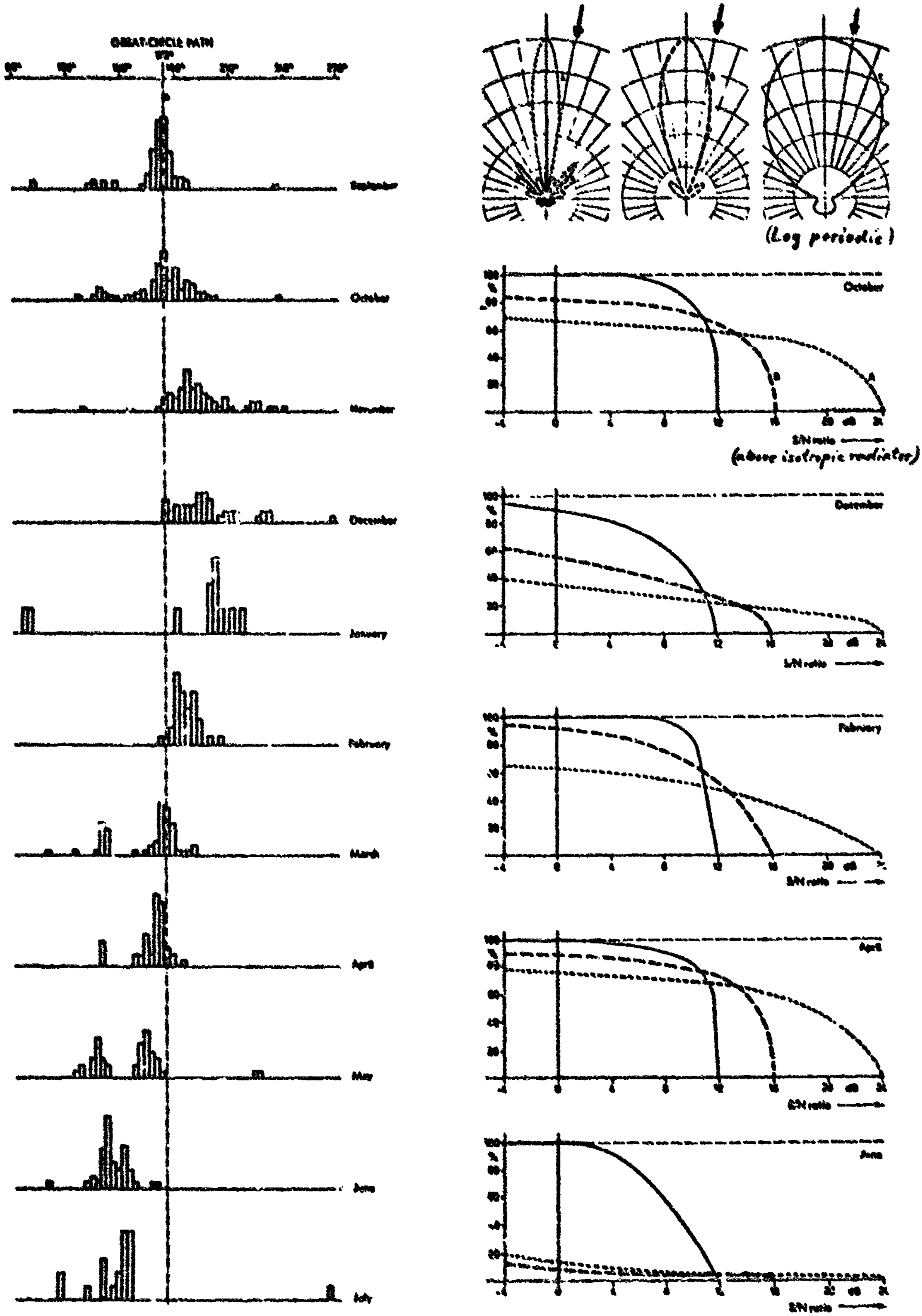


FIG. 3 Left: percentage of deviations from great-circle path of propagation path (test results in eleven successive months). Right: effect of deviations on signal-to-noise ratio with antennas of differing directivity (A rhombic antenna, B dipole curtain, C log periodic antenna). Patterns show relative field strengths, other curves show rate at which value on abscissa is obtained or exceeded.

APPENDIX A

LIST OF PARTICIPANTS

AARONS, J. Dr	Senior Scientist, Air Force Geophysics Lab. L.G. Hanscom Field, Bedford, Mass.01731, USA
ALBRECHT, H.J. Dr Panel Chairman	FGAN, 5307 Wachtberg-Werthhoven, Konigstrasse 2, Germany
ALBUQUERQUE, Ing.	Instituto Superiore Tecnico, Lisboa, Portugal
ANTE, E. Mr	ESG, Elektronik System GmbH, Postfach 800569, 8000 München 80, Germany
ANTHIMIO DE AZEVEDO, Dr	Instituto Nacional de Meteorologia e Geofisica, Rua C Aeroporto de Lisboa, 1700 Lisboa, Portugal
ARLINDO, Jose Claudino Arriaga, Mr	Institute Nacional de Meteorologia e Geofisica, Rua C Aeroporto de Lisboa, 1700 Lisboa, Portugal
BARCLAY, L.W. Mr	Home Office, Director of Radio Technology, Waterloo Bridge House, Waterloo Road, London SE1 8UA, UK
BARRETT, T.J. Mr	735 State St. Mission Research Corp. Santa Barbara, Cal.93102, USA
BELROSE, J.S. Dr EPP Panel Member	Communications Research Center, Dept. of Communications, P.O.Box 11490, Station H, Ottawa, K2H8S2, Canada
BIBL, K. Dr	University of Lowell, 450 Aiken Street, Lowell, Mass.01854, USA
BIRKNER, K.J. Dr	Messerschmitt Boelkow-Blohm, Postbox 801160, München, 80, Germany
BLYTHE, J.H. Dr EPP Panel Member	TEC Marconi Electronics Lab. Great Baddow Research Lab. Great Baddow, Chelmsford, (STD-245), UK
BOITHIAS, Mr EPP Panel Member	C.N.E.T., 38, rue du General Leclerc, 92131 Issy-les Moulineaux, France
BOSSY, L. Prof.	Institut Royal Meteorologique, 3 Avenue Circulaire, B 1180, Bruxelles, Belgium
BRADLEY, P.A. Mr	Appleton Laboratory, Ditton Park, Slough, SL3 9JX UK
BRAMLEY, E.N. Mr	Appleton Laboratory, Ditton Park, Slough, SL3 9JX, UK
BARRICK, D.E. Dr	NOAA, ERL, R45X5, Boulder, Col.80302, USA
BROCHE, Prof.	LSEET, La Giponne, Bld. des Armaris, 83100 Toulon, France
BRUSCAGLIONI, P. Dr	Istituto Ricerca Onde Elettromagnetiche, Via Panciatici 64, 50127 Firenze, Italy
BURGESS, B. Dr EPP Panel Member	R & N Dept. P(6) Building, RAE, Farnborough, UK
CALAMIA, M. Prof.	Facolta Engegneria, Istituto Elettronica, Via di S. Maria 3, 50139, Firenze, Italy
CAMPOS, Jose, Manual da Silva, Ing.	Posts & Telecommunications of Portugal, Rua do Conde de Redondo, 79, Lisboa, Portugal
CARATORI, Mr	LETTI, 9, Avenue de la Division Leclerc, 94230 Cachan, France
CARMONA, Jose Goncalves, Mr	Companhia Portuguese Radio Marconi, Praca Marques de Pombal, 15, 1200 Lisboa, Portugal
CARNEIRO, Rogerio, Manuel Ferreira Simoes, Ing.	Posts & Telecommunications of Portugal, Rua do Conde de Redondo, 79, Lisboa, Portugal
CARTER, J.R. Mr	British Aerospace Aircraft Group, Chester Road, Woodford, Stockport, Cheshire, SK7 1QR, UK
CHECCACCI, P.F. Prof. EPP Panel Member	Istituto Ricerca Onde Elettromagnetiche, Via Panciatici 64, 50127, Firenze Italy
COYNE, V.J. Mr EPP Panel Member	Chief Strategic Surveillance Branch, Surveillance Div. Rome Air Development Center, OCS, Griffiss AFB, N.Y.13441, USA

DAMBOLDT, Th. Dr	Forschungsinstitut der DBP, P.O.Box 5000, 6100 Darmstadt, Germany
DA SILVA MANIES, Eng. M.V.G.	Direccao do Servico de Electricidade e Telecomunicacoes, Rua Escola do Exercito, N.13, Lisboa, Portugal
DEGAUQUE, Prof.	Lille University, Electronics Dept. B.P.36, 59650, Villeneuve d'Asq, France
DELOGNE, P. Prof.	Laboratoire de Telecommunications, BT Maxwell, B 1348, Louvain la Neuve, Belgium
DERYCK, L. Dr	Universite de Liege, Institut de Physique, 4000 Sart Tilman, Belgium
DIAS, Manual Rosa, Mr	Instituto Nacional de Meteorologia e Geofisica, Rua C Aeroporto de Lisboa, 1700 Lisboa, Portugal
DOMINICI, P. Prof.	Istituto Nazionale di Geofisica, Via R.Borghi, 11B, 00184 Roma, Italy
ELKINS, T.J. Dr	RADC/EEP, Hanscom AFB, Ma 01731, USA
FELSEN, L.B. Prof	Polytechnic Institute of N.Y. Route 110, Farmingdale, N.Y.11735, USA
FENWICK, R.B. Dr	Barry Research Corporation, P.O.Box 61989, Sunnyvale, CA.94088, USA
FERNANDES, Antonio Carvalho, Prof.	Instituto Superiore Tecnico, Ava Rovisco Pais, 1000 Lisboa, Portugal
FERNANDES, Manual Carvalho, Mr	Instituto Nacional de Meteorologia e Geofisica, Rua C Aeroporto de Lisboa, 1700 Lisboa, Portugal
FIGANIER, Joao, Prof.	Instituto Superior Tecnico, Ava Rovisco Pais, 1000 Lisboa, Portugal
FRANCO, Domingos Antonio Piers, Ing.	Post and Telecommunications of Portugal, Rua do Conde de Redondo, 79, 1000 Lisboa, Portugal
GALE, D.J. Mr	Dept. of Electronic & Electrical Engineering, University of Surrey, Guildford, Surrey, UK
GEEM van, Th.A. Maj.	Royal Netherlands Air Force Aft. Verbindingen, C Kamp, Vliegbasis Ypenburg, Rotterdamseweg 35, 2289 AC Rijswijk, Netherlands
GERSON, N.C. Mr	Laboratory for Physical Sciences, 4928 College Avenue, College Park, MD 20740, USA
GIULI, D. Prof.	Facolta Engegneria Istituto Elettronica, Via S. Maria 3, 50139, Firenze, Italy
GOODMAN, J.M. Dr	Code 7560, 4555 Overlook Avenue, Naval Research Lab. Washington, D.C 20375, USA
GOURVEZ, P. Mr	CNET, B.P. N.40, 22301, Lannion, France
GOUTELARD, C.S. Prof.	LETTI, 9, Avenue de la Division Leclerc, 94230 Cachan, France
HANBABA, M.R. Dr	CNET, B.P. N.40, Lannion 22301, France
HARNISCHMACHER, E. Mr	Leopoldschanze 26, D 7814 Breisach, Germany
HENNEK, E.	Messerschmitt Bolkow Blohm, Post Box 801160, 8000 München 80, Germany
HILL, D.A. Mr	Dept. of Commerce, NTIA/ITS, Boulder, Colorado, 80302, USA
HUNT, R. Mr	University of Exeter, Physics Dept. Stocker Road, Exeter, Devon, UK
JENKINS, R.W. Dr	C.R.C. P.O.Box 11490, Station H, Ottawa, Ontario, K2H 8S2, Canada
JONES, T.B. Dr	Physics Dept., University of Leicester, University Road, Leicester, LE1 7RH, UK
JONG de G. Dr	Delft University of Technology, Department of Geodesy, Thijsseweg 11, Delft, Netherlands
KANE, J. Mr	Office of Naval Research (Code 465), 8000 North Quincy Street, Arlington, VA 22217, USA
KAUFMAN, I. Dr	ONR, 223, Old Marylebone Road, London, NW1 5TH, UK
KEROUB, [illegible] Dr	Radio Observatory Haifa, P.O.Box 4655, Haifa, Israel
KOOB, K. Dr	MBB Dept. AE 22, Postfach 801149, 8000 München 80, Germany
KNEBEL, U. Mr	Bundeswehramt für Wehrtechnik und Beschaffung 8072, Manching, Germany
KRAUSE, J. Mr	AEG Telefunken, Aussenstelle Bonn, Friedrich Ebert Allee 26, 5300 Bonn 1, Germany
KUHLEN, H.P. Mr	ESG Elektronik System GmbH, Postfach 800569, 8000 Munchen 80, Germany
LAMPERT, E. Dr Panel Member	Siemens AG Wv Fu, EF21, Hofmannstr.51, 8 Munchen, Germany
LANGE HESSE, G. Dr Panel Member	Max-Plank Institut für Aeronomie, Postfach 20, D-3411, Katlenburg, Lindau, Germany
LARCHER BRINCA, A. Prof.	Instituto Superior Tecnico, Ava Rovisco Pais, 1000 Lisbon, Portugal

LEMPKE, Jr C.T. Mr	Laboratory for Physical Sciences, 4928 College Avenue, College Park, MD 20740, USA
LIPA, B. Dr	Remote Measurement Lab. 320 A, SRI International, 333 Ravenswood Ave. Menlo Park, CA.94025, USA
LOURENCO, Hermenegildo Duarte, Cdr.	Direccao das Infraestructuras navais. M.Marine Praca do Comercio, 1100 Lisboa, Portugal
LYTLE, J.R. Dr	L-156, Lawrence Livermore Lab. P.O.Box 808, Livermore, CA.94550, USA
MACHADO, Jose Manuel, Albuquerque Braz. Ing.	Posts & Telecommunication of Portugal, Rua do Conde de Redondo 79, Lisbon, Portugal
MANILHA, Tito Manuel Escada, Mr	Instituto Nacional de Meteorologia e Geofisica, Rua C, Aeroporto de Lisboa, 1700 Lisboa, Portugal
MAY, G. Mr	Ministry of Defence (PE) Radio and Navigation Aircraft Establishment, Farnborough, Hants, UK
MATA, L.A Dr	Instituto Nacional de Meteorologia e Geofisica, Rua C, Aeroporto de Lisboa, 1700 Lisboa, Portugal
MCKINNON, D.J. Mr	NDHO/Ottawa Attn: DTA (CE), 101 Col. By Dr. Ottawa, Ontario, Canada
MENDES, A.S. Dr EPP Panel Member Host Coordinator	Instituto Nacional de Meteorologia e Geofisica, Rua C, Aeroporto de Lisboa, 1700 Lisboa, Portugal
MERKEL, M.A.	Dept. of the Army, HQ US Army Communications Electronics, Engineering Instalation Agency, Fort Huachuca, AZ, 85613, USA
MILLMAN, G.H. Dr	General Electric Company, Building 4, Room 57, Court Street Plant, Syracuse, N.Y.13221, USA
MUHMERT, W. Mr	ESG Elektronik Systems GmbH, Postfach 800569, 8000 München 80, Germany
NETO, Mario Alberto Martins, Ing.	Posts & Telecommunication of Portugal, Rua do Conde de Redondo 79, Lisboa, Portugal
NICOLLET, M. Prof.	Laboratoire de Physique de l'Exosphere, Universite de Paris VI, T 14, 4 Place de Jussieu, 75230 Paris CEDEX 05, France
PASCOAL, Alvaro Casimiro dos Reis, Mr	Instituto Nacional de Meteorologia e Geofisica, Rua C, Aeroporto de Lisboa, 1700 Lisboa, Portugal
PELLEGRINI, P.E. Dr	Istituto Ricerca Onde Elettromagnetiche, Via Panciatici 64, 50127 Firenze, Italy
POEVERLEIN, H. Prof. Dr	Tech. Hochschule, 6100 Darmstadt, Germany
PROLSS, G.W. Dr	Institut für Astrophysik, Auf dem Hügel 71, 53, Bonn, Germany
RABIN, H. Dr	Naval Research Lab. Code 7000, Washington, D.C.20375, USA
REINISCH, B.W. Prof.	University of Lowell, 450 Aiken Street, Lowell, Mass.01854, USA
RETO, Antonio Manuel dos Santos, Cdr.	Direccao das Infrastructuras navais M. Marine, Praca do Comercio, 1100 Lisboa, Portugal
REYNOLDS, P. Mr	Dept. of Physics, Stocker Road, Exeter University, Exeter, Devon, EX4 4QL, UK
RICCIARDI, B.V. Mr	USA AVRADA, Attn: DAVAA-E, Fort Monmouth, N.J.07703, USA
RILO, Francisco Velez, Prof.	Faculdade de Engenharia, Laboratorias de Electronica, Rua dos Bragas, 400 Porto, Portugal
ROCHA, Jose Pereira, Ing.	Posts & Telecommunications of Portugal, Rua do Conde de Redondo, 79, Lisboa, Portugal
ROMEIRAS, Filipe, Prof.	Instituto Superior Tecnico, Ava Rovisco Pais, 1000 Lisboa, Portugal
ROTTGER, J. Dr	Max-Planck Institut für Aeronomie, P.O.Box 20, D-3411 Katlenburg-Lindau 3, Germany
RUSH, C.M. Dr	NITA/ITS, 325 Broadway, Boulder, Co.80302, USA
SALEMA, Carlos, Prof.	Instituto Superior Tecnico, Ava Rovisco Pais, 1000 Lisboa, Portugal
SCHMERLING, E.R. Dr EPP Panel Member	Code ST5-NASA Hqs., Washington, D.C.20546, USA
SCHOLZ, D. Lt.Col.	NARFA, Streitkrafteamt Abt.V. Postfach [illegible]05 003, 5300 Bonn 2, Germany
SHEARMAN, E.D.R. Prof.	Dept. of Electronics & Electrical Engineering, University of Birmingham, B15 2TT, UK
SOICHER, H. Dr EPP Panel Member	Headquarters CORADCOM, Attn: DRDCO-COM-RF-5, Fort Monmouth N.J.07703, USA

SPRENKELS, C. Lt.Col. EPP Panel Member	Etat Major Force Aerienne (VDM/C) Quartier Reine Elisabeth, Rue d'Evere, B 1140, Brussels, Belgium
STARK, A. Mr	Rhode & Schwartz, Abt.4 PN, Muehldorfstr.15, D 8000 München, 80, Germany.
STEIN, V. Dr	Institut fur Hochfrequenztechnik, DFVLR, D-8031, Oberpfaffenhofen, Germany
TAAGHOLT, J. Mr EPP Panel Member	Ionosphere Laboratory, Technical University of Denmark, DK 2800, Lyngby, Denmark
TAVARES, Carlos Direitinho, Mr	Instituto Nacional de Meteorologia e Geofisica, Rua C, Aeroporto de Lisboa, Lisboa, Portugal
TANNER, R.L. Dr	Technology for Communications International, 1625, Stierlin Road, Mountain View, CA.94043, USA
THOMAS, E.C. Mr	Physics Department, University of Leicester, University Road, Leicester, LE1 7RH, UK
TISCHER, S. Mr	Bundesakademie fur Wehrwerwaztung und Wehrtechnik, Seckenhaimer Landstr. 8–10, Manheim, Germany
VAN UFFELEN, Mr	TRT, B.P.21, 92350 Le Plessis Robinson, France
VAUGHN AGY, Mr	NTIA/ITS 3 325 Broadway, Boulder, Co.80303, USA
VIDDELEER, R. Prof. EPP Panel Member	Akkerhoorn bloem 41, 2317 KS Leiden, Netherlands
VISSINGA, H. Mr	Physics Laboratory, TMO, P.O.Box 96864, The Hague, Netherlands
WAIT, J.R. Dr	Room 242, ERL, NOAA, US Department of Commerce, Boulder, Co.80302, USA
WAHLEN, J. Mr	AF Geophysics Laboratory, Hanscom AFB, MA 01731, USA
WASHBURN, T.W. Dr	Remote Measurements Lab. Building 320 A, SRI International, 333 Ravenswood Ave., Menlo Park, CA.94025, USA
YEH, K.C.	Dept. of Electrical Engineering, University of Illinois, Urbana, Illinois, 61801, USA

Interpreters

Miss M.C.MAZAUD, Mrs J.RADISSON, Mrs D WAUDBY

AGARD

Lt. Colonel J.B. CATILLER, EPP Executive
Mrs M.Tessier, EPP Secretary

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	**2. Originator's Reference**	**3. Further Reference**	**4. Security Classification of Document**
	AGARD-CP-263	ISBN 92-835-0253-1	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 TOPICS IN HF PROPAGATION

7. Presented at

a Symposium of the Electromagnetic Wave Propagation Panel held in Lisbon, Portugal 28 May–1 June 1979.

8. Author(s)/Editor(s)	**9. Date**
V.J.Coyne	November 1979

10. Author's/Editor's Address	**11. Pages**
Strategic Surveillance Branch Rome Air Development Center, Griffiss A.F. Base New York 13441, USA	628

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	Telecommunication
Ionospheric propagation	Ocean surveillance
High frequencies	NATO

14. Abstract

These proceedings include the papers and discussions presented at the AGARD Symposium on "Special Topics in HF Propagation" sponsored by the Electromagnetic Wave Propagation Panel in Lisbon, Portugal in May 1979. The purpose of the symposium was to transfer information on the state-of-knowledge of HF propagation as it applies to communication, surveillance, and other systems contemplated or presently in use by NATO. The 49 papers were divided as follows. Nine were on operational considerations, eight were on modeling of ionospheric parameters, four were on high latitude ionospheric effects, nine were on propagation through irregularities, four on remote sensing and ocean surveillance, nine on tunnel transmission and surface waves, and six on equipment considerations.

AGARD Conference Proceedings No.263
Advisory Group for Aerospace Research and Development, NATO
SPECIAL TOPICS IN HF PROPAGATION
Edited by V.J.Coyne
Published November 1979
628 pages

These proceedings include the papers and discussions presented at the AGARD Symposium on "Special Topics in HF Propagation" sponsored by the Electromagnetic Wave Propagation Panel in Lisbon, Portugal in May 1979. The purpose of the symposium was to transfer information on the state-of-knowledge of HF propagation as it applies to communication, surveillance, and other systems contemplated or presently in use by

P.T.O.

AGARD-CP-263

Electromagnetic wave transmission
Ionospheric propagation
High frequencies
Telecommunication
Ocean surveillance
NATO

AGARD Conference Proceedings No.263
Advisory Group for Aerospace Research and Development, NATO
SPECIAL TOPICS IN HF PROPAGATION
Edited by V.J.Coyne
Published November 1979
628 pages

These proceedings include the papers and discussions presented at the AGARD Symposium on "Special Topics in HF Propagation" sponsored by the Electromagnetic Wave Propagation Panel in Lisbon, Portugal in May 1979. The purpose of the symposium was to transfer information on the state-of-knowledge of HF propagation as it applies to communication, surveillance, and other systems contemplated or presently in use by

P.T.O.

AGARD-CP-263

Electromagnetic wave transmission
Ionospheric propagation
High frequencies
Telecommunication
Ocean surveillance
NATO

AGARD Conference Proceedings No.263
Advisory Group for Aerospace Research and Development, NATO
SPECIAL TOPICS IN HF PROPAGATION
Edited by V.J.Coyne
Published November 1979
628 pages

These proceedings include the papers and discussions presented at the AGARD Symposium on "Special Topics in HF Propagation" sponsored by the Electromagnetic Wave Propagation Panel in Lisbon, Portugal in May 1979. The purpose of the symposium was to transfer information on the state-of-knowledge of HF propagation as it applies to communication, surveillance, and other systems contemplated or presently in use by

P.T.O.

AGARD-CP-263

Electromagnetic wave transmission
Ionospheric propagation
High frequencies
Telecommunication
Ocean surveillance
NATO

AGARD Conference Proceedings No.263
Advisory Group for Aerospace Research and Development, NATO
SPECIAL TOPICS IN HF PROPAGATION
Edited by V.J.Coyne
Published November 1979
628 pages

These proceedings include the papers and discussions presented at the AGARD Symposium on "Special Topics in HF Propagation" sponsored by the Electromagnetic Wave Propagation Panel in Lisbon, Portugal in May 1979. The purpose of the symposium was to transfer information on the state-of-knowledge of HF propagation as it applies to communication, surveillance, and other systems contemplated or presently in use by

P.T.O.

AGARD-CP-263

Electromagnetic wave transmission
Ionospheric propagation
High frequencies
Telecommunication
Ocean surveillance
NATO

NATO. The 49 papers were divided as follows. Nine were on operational considerations, eight were on modeling of ionospheric parameters, four were on high latitude ionospheric effects, nine were on propagation through irregularities, four on remote sensing and ocean surveillance, nine on tunnel transmission and surface waves, and six on equipment considerations.

ISBN 92-835-0253-1

NATO. The 49 papers were divided as follows. Nine were on operational considerations, eight were on modeling of ionospheric parameters, four were on high latitude ionospheric effects, nine were on propagation through irregularities, four on remote sensing and ocean surveillance, nine on tunnel transmission and surface waves, and six on equipment considerations.

ISBN 92-835-0253-1

NATO. The 49 papers were divided as follows. Nine were on operational considerations, eight were on modeling of ionospheric parameters, four were on high latitude ionospheric effects, nine were on propagation through irregularities, four on remote sensing and ocean surveillance, nine on tunnel transmission and surface waves, and six on equipment considerations.

ISBN 92-835-0253-1

NATO. The 49 papers were divided as follows. Nine were on operational considerations, eight were on modeling of ionospheric parameters, four were on high latitude ionospheric effects, nine were on propagation through irregularities, four on remote sensing and ocean surveillance, nine on tunnel transmission and surface waves, and six on equipment considerations.

ISBN 92-835-0253-1

AGARD

NATO OTAN

7 RUE ANCELLE 92200 NEUILLY SUR SEINE

FRANCE

Telephone 745.08 10 · Telex 610176

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 Service
Department of National Defence
Ottawa, Ontario K1A 0Z2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92 Châtillon sous Bagneux

GERMANY
Zentralstelle für Luft- und Raumfahrt-
dokumentation und -information
c/o Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens, Greece

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direcção do Serviço de Material
da Forca Aerea
Rua da Escola Politecnica 42
Lisboa
Attn AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St Mary Cray
Orpington, Kent BR5 3RF

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn Report Distribution and Storage Unit

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 22151, USA

Microfiche
Space Documentation Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St Mary Cray
Orpington, Kent BR5 3RF
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 Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22151, USA

Printed Technical Editing and Reproduction Ltd
Harford House, 7 9 Charlotte St, London W1P 1HD

ISBN 92-835-0353